Title
The future is now : : effects of planning ahead in word production and comprehension

Author
Kleinman, Daniel Gregory

Publication Date
2013

UNIVERSITY OF CALIFORNIA, SAN DIEGO

The future is now: Effects of planning ahead in word production and comprehension

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Psychology & Cognitive Science

by

Daniel Gregory Kleinman

Committee in charge:

Professor Victor Ferreira, Chair
Professor Benjamin Bergen
Professor Jeffrey Elman
Professor Harold Pashler
Professor Keith Rayner

2013

The dissertation of Daniel Gregory Kleinman is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2013

DEDICATION

The final paragraph of this dissertation observes that no word is an island: Rather than being spoken in isolation, each one is produced in a rich, supportive context. The same is true of dissertations. Although this document bears my name, it would never have been written – or if it had, it would look very different – were it not for the many people who have helped to shape both my desire to be a scientist and the person I am today. I am honored to be able to take this opportunity to acknowledge their contributions, both direct and indirect, to my dissertation, my career, and my life.

To my family – my mom and my dad, my grandparents, Aaron, Marl, Steph, Pete, and my wonderful nieces – thank you for your endless love and support, for instilling in me a work ethic that I see in you every time I come home, and for encouraging me to pursue my passion. My interest in cognitive science was first sparked by philosophical questions we discussed on long hikes. Who knew then that I would decide to spend my life trying to answer them?

My desire to teach was shaped by the many great teachers I had in grade school, including Suzanne Williams, who encouraged my curiosity at a young age; Stacey Stebbins, Tom Avvakumovits, and Bob Jakovina, whose unfailing enthusiasm for teaching continues to be an inspiration; and Susan Stimson, who taught me how to write rigorously and about the importance of constructing a bibliography with appropriate citations (both of which came in handy for this dissertation).

I am deeply indebted to those students and professors from Pomona who helped set me on the path that eventually led me to graduate school: René Coppieters and Robert Thornton, who taught the enthralling introductory course that convinced me to major in cognitive science; Jay Atlas, who shaped my understanding of what it means to be a mentor and an academic; Shannon Otto, my fellow Cognitive Science major and constant companion throughout my Pomona career, who has provided much support and encouragement over the years and who remains, to this day, the only person who has ever succeeded in helping me to design, conduct, and write up an experiment in five days; and Debby Burke, under whose mentorship I joined my first lab, conducted my first real experiments, attended my first conference, and developed a passion for research. Debby, I have you to thank for being the first person to show me that it is possible to be a great teacher and produce world-class research at the same time, and for helping me realize that I want to devote my career to doing both.

At UCSD, I have been fortunate to be part of a community that has encouraged my intellectual development and nurtured me throughout my tenure as a graduate student. Rachel, thank you for supporting me through the ups and downs of the last several years, for reminding me of the things in life that really matter, and for giving me insightful feedback on every study in this dissertation (as well as many others) – you have my love and my gratitude. I am also grateful to Matt Hall, whose sunny disposition and boundless generosity (with both his time and his delicious baked goods) made coming to the lab each day a pleasure; to Liz Schotter, whose productivity inspires me to push myself harder; and to Nick Gruberg, Gary Oppenheim, Iva Ivanova and Rayna Brooks for intellectually stimulating conversations, for keeping me grounded, and for providing moral support. Tamar Gollan deserves special thanks for supporting me throughout the years, providing useful advice on life and research, and collaborating on a burgeoning number of projects. Most of all, this dissertation would not have been possible without the help of Vic Ferreira, whose guidance and endless patience allowed me to chart my own course. Vic, thank you for being such a great mentor and for always believing in me and in my potential even when data, deadlines, and federal grand jury service conspired against me. I chose to come to UCSD because I wanted to work with you; it ranks among the best decisions I have ever made.

Thank you all so, so much. I try every day to make you proud with the gifts you have given me. To you all I dedicate this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Chapter 2, in full, is a reprint of the material as it appears in Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1860-1877. The dissertation author was the sole investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Kleinman, D., & Ferreira, V. S. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Kleinman, D., Runnqvist, E., & Ferreira, V. S. The dissertation author was the primary investigator and author of this material.

CURRICULUM VITAE
Daniel Gregory Kleinman

EDUCATION

2008 - 2013	Ph.D. in Psychology & Cognitive Science Dissertation: *The Future is Now: Effects of Planning Ahead on Word Production and Comprehension* Chair: Victor Ferreira University of California, San Diego
2006 - 2008	M.A. in Cognitive Psychology Thesis: *The Effects of Homophones in the Picture-Word Interference Task* Chair: Victor Ferreira University of California, San Diego
2002 - 2006	B.A. in Linguistics & Cognitive Science, *cum laude* Thesis: *There Goes the Neighborhood: An Examination of Word Production and the Effects of Phonological Neighborhood Density* Chair: Deborah M. Burke Pomona College

ACADEMIC HONORS

2010 - 2011	Predoctoral Fellow; Center for Research in Language, UC San Diego
2006, 2008	National Science Foundation Graduate Fellowship Honorable Mention
2006	John Purvis Prize for best Cognitive Science thesis, Pomona College

RESEARCH INTERESTS

Psycholinguistics, word production, effects of attention on language processing, bilingual language selection

PUBLICATIONS

Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1860-1877.

Ferreira, V. S., **Kleinman, D.**, Siu, Y., & Kraljic, T. (2012). Do priming effects in dialogue reflect partner- or task-based expectations? *Psychonomic Bulletin & Review*, *19*, 309-316.

PAPERS IN PREPARATION

Gollan, T. H., **Kleinman, D.**, & Wierenga, C. E. (2013). *What's easier: Doing what you want, or being told what to do? Cued versus voluntary language switching and task switching in bilinguals*. Manuscript under revision.

Maxfield, N. D., **Kleinman, D.**, Morris, K., Morphew, K., Frisch, S. A., Ferreira, V. S., & L'Abbe, F. (2013). *Speech multitasking electrified*. Manuscript submitted for publication.

Kleinman, D., & Ferreira, V. S. (2013). *Duck, duck, ... mallard: Attentional requirements of word production affect which words people say*. Manuscript in preparation.

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2013). *Backing likely winners, ignoring likely losers: Comprehenders only predict best sentence completions*. Manuscript in preparation.

CONFERENCE PRESENTATIONS

Kleinman, D., & Ferreira, V. S. (2013, September). *Duck, duck, ... mallard: Advance word planning facilitates production of dispreferred alternatives*. Poster presented at the annual meeting of the Psychonomic Society, Toronto, ON, Canada.

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2012, November). *Cumulative semantic interference persists even in highly constraining sentences.* Poster presented at the annual meeting of the Psychonomic Society, Minneapolis, MN.

Kleinman, D. (2011, September). *Resolving semantic picture-word interference requires attention*. Paper presented at the Conference of Architectures and Mechanisms for Language Processing, Paris, France.

Kleinman, D., Siu, Y., Kraljic, T., & Ferreira, V. S. (2010, September). *Rapid effects of linguistic alignment on communication*. Poster presented at the International Workshop on Language Production, Edinburgh, Scotland.

Kleinman, D., & Ferreira, V. S. (2009, November). *Verb production hinders concurrent speech planning more than noun production*. Poster presented at the annual meeting of the Psychonomic Society, Boston, MA.

Kleinman, D., & Burke, D. M. (2005, November). *Do phonological processes contribute to aging effects on verbal short-term memory?* Poster presented at the annual meeting of the Psychonomic Society, Toronto, Ontario.

OTHER TALKS

Graduate Talk Series, UCSD (December 2013)
Center for Research in Language, UCSD (May 2011, February 2012, October 2013)
Cognitive Psychology Brown Bag, UCSD (November 2008, October 2011, April 2012)

TEACHING EXPERIENCE

Psychology of Language (Instructor; 2012)
Introduction to Statistics (TA; 2007, 2008, 2009, 2011)
Brain Damage and Mental Disorders (TA; 2006)

TEACHING INTERESTS

Cognitive Psychology
Statistics
Research Methods
Psycholinguistics
Bilingualism

PROFESSIONAL SERVICE

Ad hoc reviewer: *Cognition*, *Frontiers*, *Journal of Memory and Language*

REFERENCES

Victor S. Ferreira
vferreira@ucsd.edu
UC San Diego, Psychology

Tamar Gollan
tgollan@ucsd.edu
UC San Diego, Psychiatry

Deborah Burke
dmb04747@pomona.edu
Pomona College, Linguistics & Cognitive Science

ABSTRACT OF THE DISSERTATION

The future is now: Effects of planning ahead in word production and comprehension

by

Daniel Gregory Kleinman

Doctor of Philosophy in Psychology and Cognitive Science

University of California, San Diego, 2013

Professor Victor Ferreira, Chair

This dissertation consists of three studies that investigate the extent to which speakers and listeners can and do plan ahead during production and comprehension.

Study 1 investigates the attentional requirements of word selection. In two dual-task experiments, subjects categorized tones and then named pictures while word selection difficulty was manipulated using the picture-word interference and cumulative semantic interference paradigms. Results show that word selection requires domain-general attentional resources and that a difference in automaticity between word selection and another process (here, word reading) can affect performance in dual-task as compared with single-task settings.

Study 2 investigates whether a difference in automaticity between stages of word production can affect the words that speakers say when they plan their speech in advance. Specifically, if selecting a word from the lexicon requires more attention than activating it as a potential candidate, planning ahead should allow words more time to accrue activation prior to selection, asymmetrically facilitating the production of weakly active words. In two experiments, subjects named pictures that had multiple acceptable names under conditions that manipulated how soon subjects' attentional resources would be available to engage in word selection. Results show that speakers are more likely to use uncommon labels when word selection is delayed by another task, indicating that the attentional requirements of language production processes affect their outcome.

Study 3 investigates the predictions that comprehenders make about upcoming words, specifically focusing on whether words that are semantically related to a best sentence completion are pre-activated or inhibited (or neither). In three experiments that used the cumulative semantic interference paradigm, subjects named pictures that were presented either in isolation or after a strongly constraining sentence fragment. Equal interference effects across conditions indicate that words semantically related to the best completion of a sentence are unaffected by the processing of that sentence, and thus suggest that comprehenders only predict best sentence completions.

Together, these studies suggest that the manner in which speakers and comprehenders divide their attention between current and upcoming words affects the identity and processing of those upcoming words.

CHAPTER 1:

INTRODUCTION

When speakers speak, they unleash a veritable torrent of words that are produced at a rate of three to four per second. Each word is individually selected from an expansive mental dictionary that encompasses tens of thousands of words. This selection process is far from random: Every word must satisfy numerous constraints operating simultaneously at the levels of message, syntax, and semantics. Collectively, they must accurately communicate the speaker's intended message.

A listener's share of the linguistic processing burden is no less difficult. From the sounds that emerge from a speaker's lips, a listener must reconstruct the speaker's words and integrate them into the discourse context to recover the intended meaning. Furthermore, in order not to fall behind, the listener must make sense of each word at least as fast as the speaker produces a new one.

Although the output of language production is inherently sequential – words must be spoken one at a time – both speakers and listeners, like good investors, simultaneously deal with the present while preparing for the future. Speakers do not simply plan and then produce each word before beginning to plan the next; instead, they may plan several words at a time, identifying the roles that each one will play in the sentence and then filling those slots with words and retrieving each word's constituent sounds as it approaches the front of the output queue. Such advance planning can allow speakers to maintain a speech rate that would otherwise be quite difficult to keep up.

Listeners are no slouches, either: Rather than waiting for each word to emerge from the speaker's lips before beginning to process it, they may generate expectations about the meaning or the form of upcoming words. When their expectations are correct, listeners have an easier time processing those words.

In this dissertation, I investigate how speakers and listeners use the tools at their disposal to prepare and to predict upcoming words. Study 1 lays the groundwork for Study 2 by determining whether selecting a word for production requires domain-general attentional resources and whether potential differences in the automaticity between word selection and another process (word reading) can affect performance in dual-task as compared with single-task settings. Building on these results, Study 2 examines how differences in the automaticity between word selection and another process (activating potential words for production), combined with the tendency of speakers to plan words in advance, affects the words that speakers say. Study 3 investigates which words comprehenders predict on the basis of linguistic input by studying how those predictions affect the time course of subsequent word selection.

Stages of word production: Definitions and theoretical debates

In order to frame the theoretical questions of interest more precisely, it is necessary first to define the stages of word production. To use an example that will soon become familiar, a speaker who wants to retrieve the word “cat” must first select the semantic content to be expressed, which may be represented as decompositional semantic features (<IS A PET>, <MEOWS>; e.g., Dell, 1986) or as word-specific semantic representations (*lexical concepts*; Levelt, Roelofs, & Meyer, 1999). Next, she must identify which of the words, or lemmas (Levelt, 1989), in her lexicon best communicates that content (*cat*) in a process known as lemma selection. The phonological wordform, or lexeme, of the selected lemma (/kæt/) and its constituent phonemes (/k/, /æ/, /t/) are selected next, after which the speaker prepares her articulators to produce the phonemes and, finally, says “cat”.

As the questions considered in this dissertation are largely concerned with lemma selection, the stage at which a single word is plucked from a speaker’s expansive lexicon and readied for production, some theoretical background is in order. The word production literature is rife with debates over the *when* and *how* of lemma selection. As an example of the debate over the *when* of lemma selection, psycholinguists disagree over whether speakers must select a word before activating its phonemes (e.g., Garrett, 1980; Levelt et al., 1999; Roelofs, 1992) or whether phonemic representations are activated for multiple words – both the target and its competitors – before a word is selected (Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998); and, in the latter case, whether phonological activation can feed back to the lexical level to influence word selection (Dell, 1986; Ferreira & Griffin, 2003; MacKay, 1987; Rapp & Goldrick, 2000; Stemberger, 1985). As an example of the debate over the *how* of lemma selection, psycholinguists disagree over the mechanics of the selection process itself; i.e., whether words race independently toward an activation threshold that triggers selection (e.g., Dell, 1986; Oppenheim, Dell, & Schwartz, 2010); whether they compete with each other for selection, such that it takes longer to select a word for production when a competitor is more highly active (Levelt et al., 1999); or whether they engage in a special form of competition in which they inhibit the activation of competitors (Cutting & Ferreira, 1999; Dell & O’Seaghdha, 1994; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; for a review, see Goldrick, 2006; Rapp & Goldrick, 2000).

Despite their differences, one thing on which these accounts generally agree is the function of the selection process itself: to single out a target word from tens of thousands of available options in the lexicon. This is the *what* of lemma selection. If a speaker wants to retrieve the name of a furry house pet that meows cutely when it wants food, the goal of lemma selection is to ensure that the speaker ultimately selects the word *cat* and not, say, *dog*, *elephant* or *table*.

Semantic interference effects as a window into the chronometry of lemma selection

As speaking is often an easy and error-free process, psycholinguists have devised methods to make it more difficult in order to study how different stages of language production work. One such suite of methods, primarily used to study the time course of lemma selection, reliably generate semantic interference effects in which a target word is made more difficult to retrieve by the presentation of a same-category stimulus. For example, in the picture-word interference paradigm, subjects name a picture of a cat more slowly when it is accompanied by the (written or spoken) distractor word DOG than an unrelated control word such as MOON (e.g., Glaser & Düngelhoff, 1984; Lupker, 1979; Rayner & Springer, 1986; Schriefers, Meyer, & Levelt, 1990). More than a few researchers have observed that this effect is qualitatively similar to the Stroop effect (Stroop, 1935), in which subjects are slower to name the color of a written word when the word itself is incongruent with the ink color (e.g., BLUE in red ink) than when it is congruent (RED in red ink; e.g., Roelofs, 2003; see MacLeod, 1991 for a review). Both findings are easily explained by models in which lemmas compete for selection; under such an account, it takes longer to select *red* when *blue* is also active.

Not all semantic interference effects arise from bivalent stimuli. In the cumulative semantic interference paradigm (Howard et al., 2006), subjects are slower to name *cat* when other animal pictures (e.g., *dog*) have been named previously in a sequence of pictures from many different categories. That is, each act of word production slows the subsequent production of same-category words. All models of this effect agree that it results from incremental weight changes to the lexical-semantic network effected by previous acts of selection and that these weight changes affect the ease of lemma selection; i.e., the reason subjects are slower to begin naming "cat" is because naming *dog* makes the selection of the *cat* lemma more difficult. However, the models differ on the specific locus of the weight changes and whether it is necessary to posit competition during lemma selection (Belke, 2013; Howard et al., 2006) or not (Oppenheim et al., 2010) to account for the interference.

Although some of these tasks (especially picture-word interference and the Stroop task) may appear to lack external validity – in real life, one rarely needs to name incorrectly labeled pictures or read mismatching color words – it is undeniable that speakers regularly need to decide which of several active words to select for production. Semantic interference effects provide researchers with a useful way to study that decision process.

Research questions

The present studies use picture naming tasks, including both the picture-word interference and cumulative semantic interference paradigms, to address questions about what happens when speakers and comprehenders plan ahead. How does preparing to plan the next word, or expecting to hear a particular word next, affect the activation of words in the lexicon?

In the domain of word production, “planning ahead” refers here to the process of planning word *n*+1 before the planning of word *n* is complete; that is, planning multiple words simultaneously. Evidence suggests that speakers can activate, or pre-activate, the names of multiple objects simultaneously when they have enough attentional resources to spare (Mädebach, Jescheniak, Oppermann, & Schriefers, 2011; Malpass & Meyer, 2010; Meyer, Ouellet, & Häcker, 2008; Oppermann, Jescheniak, & Görges, 2013; Oppermann, Jescheniak, & Schriefers, 2008; Oppermann, Jescheniak, Schriefers, & Görges, 2010; see also Schotter, Ferreira, & Rayner, 2013). Can speakers select words *n* and *n*+1 for production at the same time as well, or is parallel processing limited to pre-selection stages of production? If word selection does require more attentional resources than pre-selection stages, that would create a bottleneck in advance planning, placing an important limitation on speakers’ ability to plan ahead. What effects would this limitation have on production?

In the domain of word comprehension, “planning ahead” refers here to comprehenders’ ability to generate expectations about the identities of upcoming words on the basis of linguistic input (see, e.g., DeLong, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004). For example, a comprehender who hears the beginning of a sentence that strongly suggests a particular completion (“After doing his laundry, Mark always seemed to be missing one…”) may predict that that completion – “sock” – will be the next word. This prediction leads to pre-activation of the *sock* lemma; that is, an increase in activation before the word itself is ever presented. What about the other lemmas in the comprehender's lexicon – how are they affected? Do comprehenders make multiple, graded predictions, in which case the *shirt* lemma might also be pre-activated (albeit to a lesser extent)? Do comprehenders actively inhibit other lemmas, which might facilitate the eventual recognition of *sock*? Or do comprehenders only make a single prediction – for *sock* – and leave the activation levels of other lemmas unaffected?

Study 1

As noted above, models of word production agree that the goal of lemma selection is to single out a target word from the lexicon for production. The act of determining which word should be produced would seem to be synonymous with determining which words should *not* be produced, which is why semantic interference is assumed to be resolved during lemma selection. However, a recent paper using a dual-task paradigm suggested that picture-word interference resolution occurs before lemma selection (Dell'Acqua, Job, Peressotti, & Pascali, 2007). In other words, when a speaker names a picture of a cat, the stage of processing that is slowed down when it is accompanied by the word DOG compared with the word MOON precedes the stage at which the speaker selects the lemma *cat* for production. Such an account, if true, would not only contradict models of lemma selection and picture-word interference (Levelt et al., 1999; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007); it would also necessitate a reevaluation of what exactly it means to select a word.

Study 1 (Kleinman, 2013) evaluates the hypothesis that Dell'Acqua et al.'s (2007) results could be attributable to the fact that picture-word stimuli are processed differently in dual-task and single-task contexts. This hypothesis was contingent on the assumption that selecting a lemma for production requires more attentional resources than word reading. As such, it uses the picture-word interference and cumulative semantic interference paradigms to determine whether lemma selection requires domain-general attentional resources (it does), whether a difference in automaticity between stages of a task can cause performance to differ between dual- and single-task contexts (it can), and whether models of word production need to be revised to account for Dell'Acqua et al.'s results (they don't).

Study 2

The results of Study 1 showed that speakers cannot select the lemma of a picture while they are simultaneously engaged in another attention-demanding task, making it unlikely that they can select the lemmas of words n and n+1 simultaneously. In contrast, they can simultaneously activate the names of words n and n+1. Taken together, this means that while speakers are planning word n, they should be able to activate potential lemmas for word n+1 without being able to select one (until after finishing attention-demanding processing for word n).

Study 2 aims to determine whether this difference in automaticity between lemma selection and pre-selection stages of production can affect the words that speakers say. Relative to producing words in isolation, planning words in advance should increase the amount of time that activation can accrue to lemmas prior to selection. When multiple words are equally appropriate but one (e.g., *couch*) is more accessible than the other (*sofa*), the bounded nature of activation should cause the extra activation to benefit weakly active names (*sofa*) more than strongly active names (*couch*), increasing the likelihood that speakers produce the weakly active (and generally dispreferred) name. This hypothesis was tested using a paradigm in which subjects named pictures on every trial under conditions that manipulated how soon their attentional resources would be available to engage in lemma selection. Subjects were predicted to be more likely to use dispreferred picture names when their attentional resources were otherwise occupied. If true, this would demonstrate that the attentional requirements of word production processes, combined with the tendency of speakers to initiate planning for word *n*+1 before completing attention-demanding processing for word *n*, can determine the words that speakers ultimately produce.

Study 3

Whereas Studies 1 and 2 collectively addressed the effects of planning ahead on production, Study 3 addressed the effects of planning ahead on comprehension. Expecting to hear “sock” leads to pre-activation of the *sock* lemma. How does it affect the activation levels of semantically related lemmas, like *shirt* and *pants*, which are often appropriate in the same kinds of contexts as *sock*?

To address these questions, Study 3 combined the presentation of strongly constraining sentences with the cumulative semantic interference paradigm. Subjects read sentence fragments (such as the one that leads to a strong expectation for *sock*) and named a picture presented at the end of the sentence that corresponded to the best completion (a picture of a sock). Depending on how much naming *sock* after a sentence slows the subsequent naming of *shirt* after a sentence relative to how much naming *sock* in isolation slows the subsequent naming of *shirt* in isolation, the paradigm can show whether (and if so, how) reading a strongly constraining sentence affects the activation of words other than the best sentence completion. In doing so, it can reveal the number of predictions comprehenders make and how strong those predictions are.

Summary

Although language is produced sequentially, both speakers and listeners prepare for upcoming words at the same time as they are producing and comprehending the current one. This forward-looking mindset ensures that the speaker can maintain a rapid, fluent speech rate while the listener can process the speaker's words as fast as they are produced. The studies in this dissertation will shed light on how both parties accomplish their respective linguistic feats.

References

Belke, E. (2013). 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*, *69*, 228-256. doi:10.1016/j.jml.2013.05.008

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318-344. doi:10.1037/0278-7393.25.2.318

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283-321. doi:10.1037/0033-295X.93.3.283

Dell, G. S., & O'Seaghdha, P. G. (1994). Inhibition in interactive activation models of linguistic selection and sequencing. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 409-453). San Diego: Academic Press.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722. doi:10.3758/BF03196827

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1121. doi:10.1038/nn1504

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, *14*, 86-90. doi:10.1111/1467-9280.01424

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (pp. 177-220). New York: Academic Press.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640-654. doi:10.1037/0096-1523.10.5.640

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*, 817-855. doi:10.1080/01690960600824112

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464-482. doi:10.1016/j.cognition.2005.02.006

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1256-1274. doi:10.1037/0278-7393.24.5.1256

Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1860-1877. doi:10.1037/a0033095

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-39. doi:10.3115/992628.992631

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*, 485-495. doi:10.3758/BF03198265

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. Springer-Verlag.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203. doi:10.1037/0033-2909.109.2.163

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech planning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 649–660. doi:10.1037/a0022330

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). 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*, *33*, 503-535. doi:10.1037/0278-7393.33.3.503

Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 523-537. doi:10.1037/a0018522

Meyer, A. S., Ouellet, M., & Häcker, C. (2008). Parallel processing of objects in a naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 982-987. doi:10.1037/0278-7393.34.4.982

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

Oppermann, F., Jescheniak, J. D., & Görges, F. (2013). Resolving competition when naming an object in a multiple-object display. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-013-0465-5

Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2008). Conceptual coherence affects phonological activation of context objects during object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 587–601. doi:10.1037/0278-7393.34.3.587

Oppermann, F., Jescheniak, J. D., Schriefers, H., & Görges, F. (2010). Semantic relatedness among objects promotes the activation of multiple phonological codes during object naming. *Quarterly Journal of Experimental Psychology*, *63*, 356–370. doi:10.1080/17470210902952480

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539-557. doi:10.1037/0278-7393.24.3.539

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460-499. doi:10.1037/0033-295X.107.3.460

Rayner, K., & Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture-word interference task. *British Journal of Psychology*, *77*, 207-222.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107-142. doi:10.1016/0010-0277(92)90041-F

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88-125. doi:10.1037/0033-295X.110.1.88

Schotter, E. R., Ferreira, V. S., & Rayner, K. (2013). Parallel object activation and attentional gating of information: Evidence from eye movements in the multiple object naming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 365-374. doi:10.1037/a0028646

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86-102. doi:10.1016/0749-596X(90)90011-N

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143-186). Hillsdale, NJ: Erlbaum.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of*

Experimental Psychology, *18*, 643-662. doi:10.1037/0096-3445.121.1.15

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443-467. doi:10.1037/0278-7393.31.3.443

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). 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*, *16*, 1272-1288. doi:10.1162/0898929041920487

CHAPTER 2:

RESOLVING SEMANTIC INTERFERENCE DURING WORD PRODUCTION REQUIRES CENTRAL ATTENTION

Abstract

The semantic picture-word interference task has been used to diagnose how speakers resolve competition while selecting words for production. The attentional demands of this resolution process were assessed in two dual-task experiments (tone classification followed by picture naming). In Experiment 1, when pictures and distractor words were presented simultaneously, semantic interference was not observed when tasks maximally overlapped. This replicates a key finding from the literature that suggested that semantic picture-word interference does not require capacity-limited central attentional resources and occurs prior to lexical selection, an interpretation that runs counter to the claims of all major theories of word production. In another Experiment 1 condition, when distractors were presented 250 ms after pictures, interference emerged when tasks maximally overlapped. Together, these findings support an account in which interference resolution and lexical selection both require central resources, but the activation of lexical representations from written words does not. Subsequent analysis revealed that discrepant results obtained in previous replication attempts may be attributable to differences in phonological (ir)regularity between languages. In Experiment 2, degree of semantic interference was manipulated using the cumulative semantic interference paradigm. Interference was observed regardless of task overlap, confirming that lexical selection requires central resources. Together, these findings indicate that a lexical selection locus of semantic picture-word interference – and models of word production that assume such a locus – may be retained.

A longstanding goal of cognitive psychology has been to determine how people are able to select a single response from among many alternatives. Nowhere is the set of possible responses larger than in the psycholinguistic domain. Estimates put the productive vocabulary size of a well-educated adult native speaker of English at around 30,000 words (Levelt, 1989). How is the language production system able to select the correct word from such a large set, and how does this selection process unfold over time?

Before addressing these questions, it is useful to situate them in the appropriate theoretical context: models of word production. Although different models disagree on specific processing assumptions regarding the spread of activation and the existence of competition between representations (e.g., Bloem & La Heij, 2003; Caramazza, 1997; Dell, 1986, 1988; Levelt, Roelofs, & Meyer, 1999), they largely agree on what the major stages of word production are and how they are ordered. Before producing a word (e.g., "cat"), speakers must first identify the semantic content they wish to express (<IS A PET>, <MEOWS>). Next, they must determine which word representation, often termed the lemma, best communicates that content (*cat*). This process is known as lemma selection, and it is the stage of production with which questions regarding word selection are concerned. The phonemes of the selected lemma (/k/, /æ/, /t/) are retrieved during phoneme selection. Finally, speakers prepare their articulators to produce the retrieved phonemes, after which the word is uttered.

Models of the language production system rely heavily on data from two sources: speech errors, which provide qualitative data about where production processes can go wrong; and reaction time (RT) tasks, which provide quantitative data about the time course of those processes. One such RT task is picture-word interference (PWI), one of the workhorses of psycholinguistic research over the last 40 years (e.g., Glaser & Düngelhoff, 1984; Lupker, 1979; Rayner & Springer, 1986; Roelofs, 1992; Schriefers, Meyer, & Levelt, 1990). On each trial, participants are presented with a picture to name accompanied by a distractor word (either spoken or written) to ignore. By varying the relation of the distractor word to the lemma for the picture name as well as the distractor stimulus onset asynchrony (SOA) – that is, the delay between the onset of the picture and the onset of the word – it is possible to trace the time course of lexical access. For example, Glaser and Düngelhoff (1984) showed in a classic study that picture naming latencies are slower when a picture (e.g., APPLE) is accompanied by a same-category written distractor word (*peach*) than an unrelated distractor (*nickel*), but only when the word is presented at a distractor SOA between -100 ms and +100 ms. Though studies differ on the exact timing conditions under which semantically related distractors generate interference, they generally agree on a narrow window that includes a distractor SOA of 0 ms; outside that window, the interference disappears (e.g., Damian & Bowers, 2003; Damian & Martin, 1999; Starreveld & La Heij, 1996). As this semantic PWI is often thought to arise from competition between lemmas during lemma selection (e.g., Levelt et al., 1999; Roelofs, 1992; but see Mahon, Costa, Peterson, Vargas, & Caramazza, 2007), these results point to a critical window for the effect in which semantically related distractors are activated early enough to affect lemma selection, but not so early that they can be discounted by the language system prior to lemma selection.

In parallel with their efforts to study how different stages of word production (e.g., lemma selection and phoneme selection) unfold over time through the use of PWI experiments, language researchers have used the psychological refractory period (PRP) paradigm to determine their attentional requirements. According to the central bottleneck model of attention (Pashler, 1984; Welford, 1952), a prominent account of performance in dual-task experiments, tasks can be decomposed into three discrete, serially ordered stages of processing: perceptual encoding, in which a stimulus is apprehended; response selection, in which a response is chosen on the basis of the apprehended stimulus and rules that govern stimulus-response mappings; and response execution, in which the chosen response is manually prepared. Crucially, response selection (but not perceptual encoding or response execution) requires the use of indivisible, central (i.e., domain-general) attentional resources. When participants must produce independent responses to stimuli that are presented in close temporal proximity, the limited nature of these resources – which can only be allocated to response selection processing for one task at a time – gives rise to a processing bottleneck.

The effect of the bottleneck on dual-task processing is shown in Figure 2.1 (similar to, e.g., Ferreira & Pashler, 2002, Figure 2). When the stimuli for the two tasks are presented with a long delay between their onsets (i.e., at a long task SOA), Task 1 response selection has already been completed prior to the completion of Task 2 perceptual encoding (Figure 2.1a). This means that Task 2 response selection can begin after perceptual encoding without delay. In contrast, when the stimuli for the two tasks are presented with only a short delay between their onsets (i.e., at a short task SOA), Task 2 perceptual encoding is completed before the completion of Task 1 response selection (Figure 2.1b). As Task 2 response selection requires the use of the same mental resources devoted to Task 1 response selection, it is delayed until Task 1 response selection finishes (represented by a dashed vertical line). This bottleneck creates cognitive “slack” during which no Task 2 processing occurs. Importantly for the present study, if the difficulty of Task 2 perceptual encoding were to increase by a moderate amount at a short task SOA, the added processing difficulty would be absorbed by the slack, and would not increase the Task 2 reaction time.

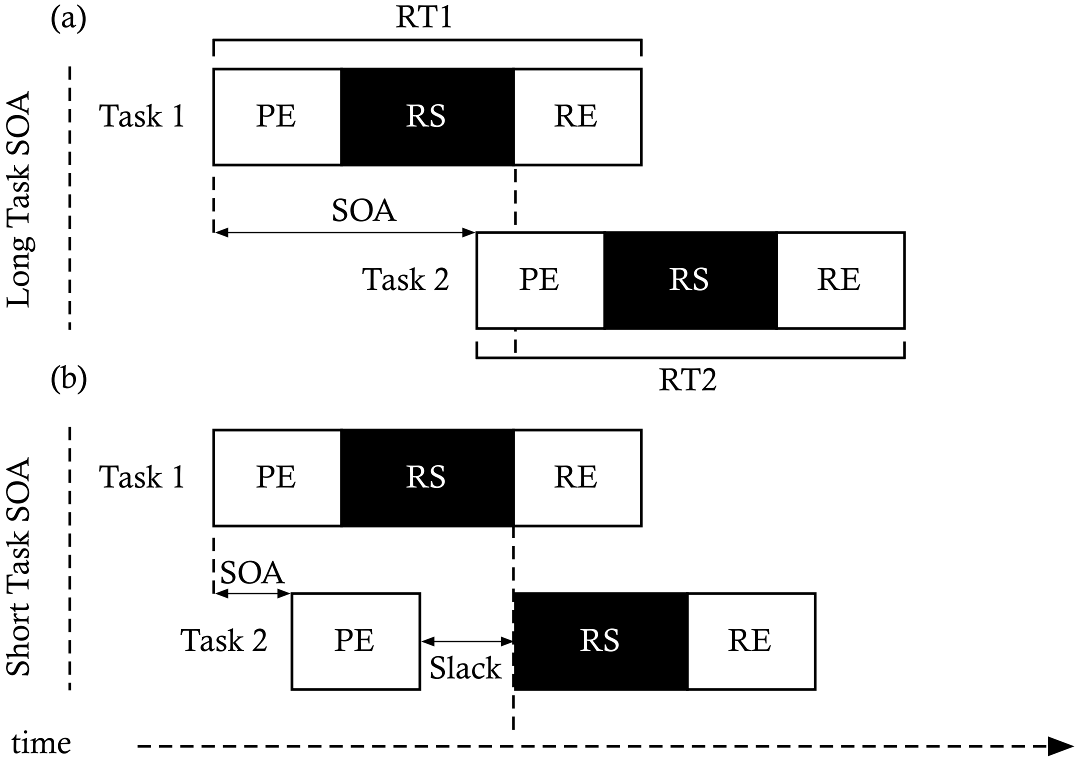


Figure 2.1. The central processing bottleneck in the dual-task paradigm. PE = perceptual encoding (beginning immediately after stimulus presentation), RS = response selection, and RE = response execution (after which a response is issued); SOA = stimulus onset asynchrony. RT1 and RT2 represent reaction times to each task. White boxes denote processing stages that do not require domain-general attentional resources; black boxes denote processing stages that do.

Semantic picture-word interference: perceptual or post-perceptual?

By applying the PRP paradigm to the study of word production, it is possible to determine the temporal locus and attentional requirements of each production stage (but see Roelofs & Piai, 2011). The same can be deduced for the semantic PWI effect. The first attempt to do so (Ferreira & Pashler, 2002, Experiment 2) exemplifies the methodology typical of such dual-task studies. Participants were presented with two stimuli on each trial (a picture-word stimulus followed by a tone) and instructed to respond to both in order as quickly as possible (by naming the picture and then pressing a button to identify the tone pitch as low or high). The relation of the word to the picture (APPLE-*peach* vs. APPLE-*nickel*) and the task SOA were varied to determine whether and when semantic interference affected reaction times (RTs) to each task.[1] Ferreira and Pashler found that pictures were named more slowly when accompanied by semantically related than unrelated distractors across all task SOAs, replicating the standard semantic PWI effect. More interestingly, this interference slowed tone discrimination latencies as well. According to the central bottleneck model, this indicates that semantic PWI must be resolved during either perceptual encoding or response selection, as an increase in duration of either stage would delay the completion of Task 1 response selection (the dashed vertical line in Figure 2.1b), which would in turn delay the commencement of Task 2 response selection and thus increase Task 2 RTs. In contrast, if semantic PWI

---

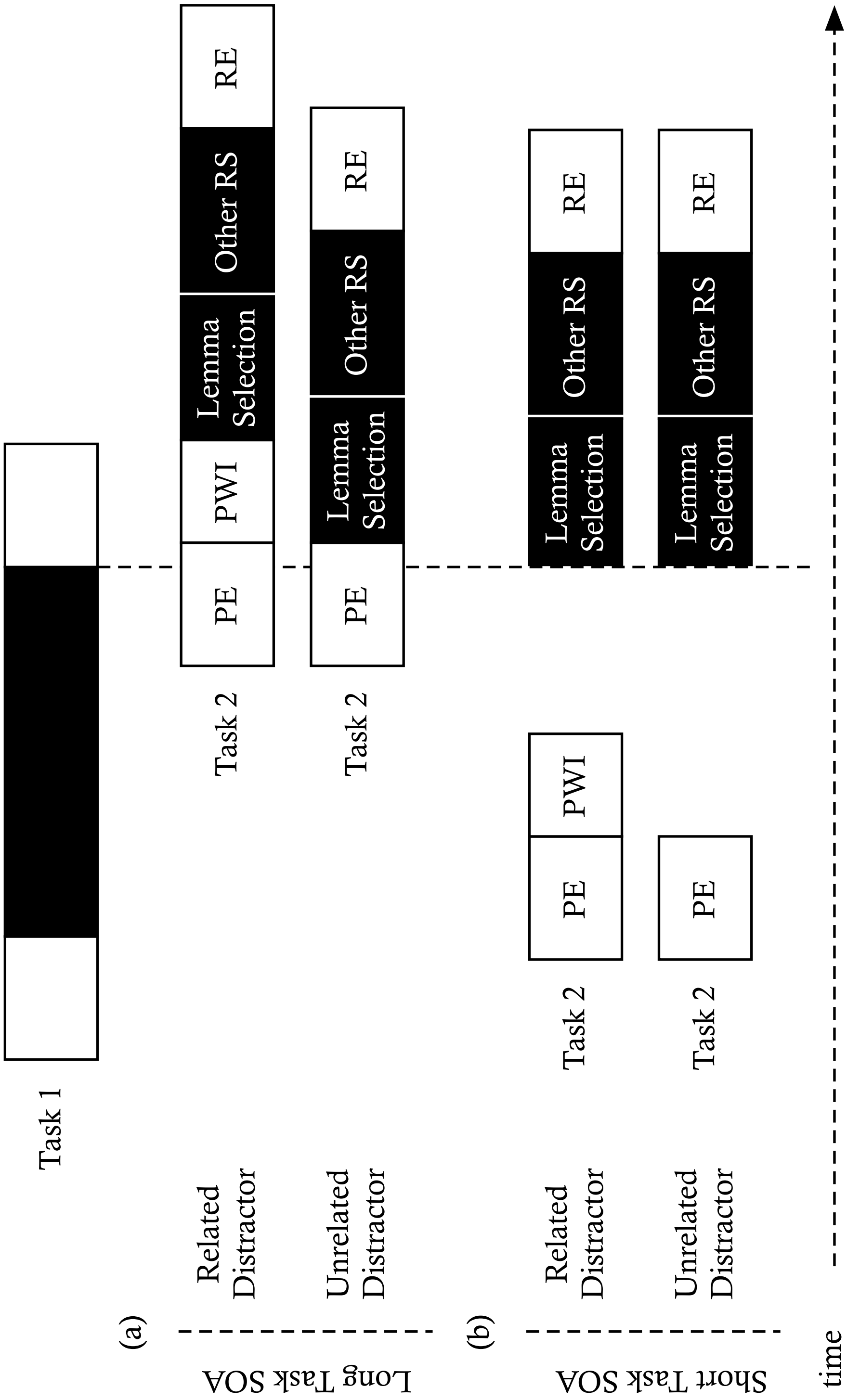
[1] Note that task SOA is distinct from distractor SOA. In a dual-task experiment, task SOA represents the delay between the presentation of the Task 1 and Task 2 stimuli (here, the picture and the tone). In picture-word interference, distractor SOA represents the delay between the presentation of the picture and the distractor word.

were to resolve during response execution, tone discrimination latencies would have been unaffected by distractor relatedness.

To further pinpoint the locus of semantic PWI resolution, Dell'Acqua, Job, Peressotti, and Pascali (2007) reversed the order of tasks that had previously been used in dual-task picture naming studies. Participants in their experiment were presented with a tone followed by a picture-word stimulus, and had to perform tone discrimination before picture naming. Unsurprisingly, they found that when the picture and word were simultaneously presented either 1000 ms or 350 ms after the tone, participants named the picture more slowly when it was accompanied by a semantically related word than a semantically unrelated word – the standard semantic PWI effect. When the task SOA was reduced to 100 ms, however, the PWI effect disappeared completely. According to the logic of the central bottleneck model, this indicates that semantic PWI resolution occurs during perceptual encoding, as the interference would have persisted at the short task SOA if it had either a response selection or response execution locus. With a perceptual locus, the extra processing generated by a semantically related distractor would be absorbed by the slack created by the central bottleneck at a short task SOA (see Figure 2.2).

The finding that semantic PWI affects perceptual processing is theoretically important for at least two reasons. First, it poses a major challenge to prominent accounts of the task. WEAVER++ (Levelt et al., 1999), a model of word production that has used PWI experiments to inform its assumptions about the time course of lexical access, posits that this interference is resolved during lemma selection. A different model of PWI, the

Figure 2.2. Dell'Acqua et al.'s (2007) account of picture-word interference (PWI) in a dual-task paradigm. Task 1 = tone discrimination; Task 2 = picture naming. PE = perceptual encoding; RS = response selection; RE = response execution; SOA = stimulus onset asynchrony; PWI = picture-word interference. Under Dell'Acqua et al.'s account, semantically related distractor words generate PWI at every task SOA. (a) When the picture-word stimulus is presented at a long delay after the tone, this PWI increases picture naming latencies. (b) When the picture-word stimulus is presented soon after the tone, PWI resolution is absorbed into the "cognitive slack" created by the bottleneck, yielding no effect of distractor relatedness.


Long Task SOA
(a)
Task 1
Related Distractor
Task 2
PE
PWI
Lemma Selection
Other RS
RE
Unrelated Distractor
Task 2
PE
Lemma Selection
Other RS
RE
Short Task SOA
(b)
Related Distractor
Task 2
PE
PWI
Lemma Selection
Other RS
RE
Unrelated Distractor
Task 2
PE
Lemma Selection
Other RS
RE
time

Response Exclusion Hypothesis (Mahon et al., 2007), argues that interference is resolved after lemma selection. These are not minor assumptions: If research were to conclusively demonstrate that semantic interference is resolved prior to lemma selection, both the models and our understanding of how words are produced would need to be fundamentally revised.

Each model can account for a perceptual locus of semantic PWI resolution only if lemma selection occurs during perceptual encoding as well. However, as the stage at which a speaker decides which word to say, lemma selection would seem to be the very definition of a response selection process. As such, it is likely to have a post-perceptual locus. Combined with the finding that semantic PWI is resolved during perceptual encoding, this calls into question the assumptions of both WEAVER++ and the Response Exclusion Hypothesis.

A second implication of such a finding is that it suggests semantic PWI and the Stroop effect reflect different underlying cognitive processes. In each trial of the Stroop task (Stroop, 1935), participants are presented with a color word and must name the color of the ink in which that word is written, which is either congruent or incongruent with the identity of the word. Participants are slower to name the color when it is incongruent with the word (e.g., the word BLUE in red ink) than when the two are congruent (RED in red ink).

A number of similarities exist between the tasks. Both Stroop and semantic PWI require participants to ignore a written word from the same semantic category as a to-be-produced target. Furthermore, when participants are instead asked to read the word aloud, the effects of picture-word relatedness (in PWI) and color-word congruency (in Stroop) are minimal. For these reasons, as well as others, the same cognitive processes are often thought to underlie the two tasks (e.g., Roelofs, 2003; see MacLeod, 1991 for a review).

The attentional demands of the Stroop task were first assessed in a dual-task experiment conducted by Fagot and Pashler (1992, Experiment 7). In that experiment, participants categorized the pitch of a tone and then named the color of a Stroop stimulus. The word was presented either 50 ms before, or 50, 150 or 450 ms after, the tone. At every task SOA, participants demonstrated a robust Stroop effect, naming the color of the word substantially slower for incongruent stimuli than congruent ones. Crucially, the size of this effect did not interact with task SOA, indicating that the Stroop manipulation slows response selection (or response execution; but see Magen & Cohen, 2002, 2010 for an alternate interpretation). This contrasts with the results of Dell'Acqua, Job et al. (2007), who found an underadditive interaction between semantic PWI and task SOA, thereby indicating that semantic PWI affects a perceptual stage of processing. If resolving Stroop interference and resolving semantic PWI have different attentional requirements, the claim that they arise from the same cognitive processes is no longer tenable.

Given the implications for models of word production, picture-word interference and the Stroop effect, much rides on the claim that semantic PWI affects a perceptual stage of processing. Thus, before accepting that these models have been seriously challenged, it is worth considering whether other interpretations of Dell'Acqua, Job et al.'s (2007) results are possible (an approach also adopted by van Maanen, van Rijn, & Borst, 2009). This paper will evaluate the possibility that the underadditive interaction they found between semantic relatedness and task SOA was due to the fact that picture-word stimuli are processed differently in dual-task and single-task settings.

Treating target pictures and distractor words as separate stimuli

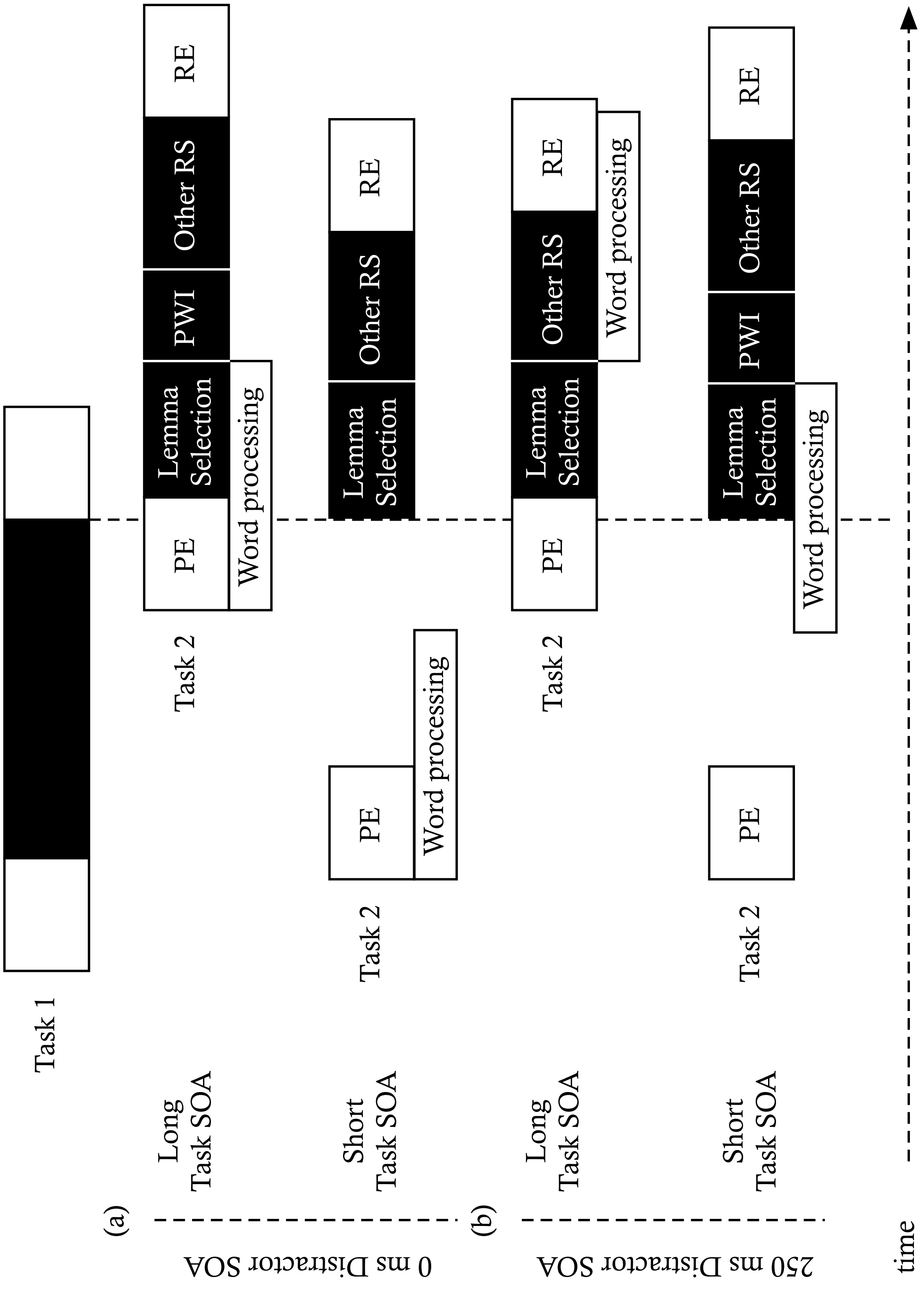
All of the papers that have used the central bottleneck model to describe picture-word interference in dual-task experiments have treated picture and word processing as part of the same task, jointly represented by single perceptual encoding, response selection and response execution stages (Ayora et al., 2011; Cook, 2007; Cook & Meyer, 2008; Dell'Acqua, Job et al., 2007; Ferreira & Pashler, 2002; Piai, Roelofs, & Schriefers, in press; Schnur & Martin, 2012; see also Piai & Roelofs, 2013; van Maanen et al., 2009; van Maanen, van Rijn, & Taatgen, 2012). Although it is true that participants must only select a single response for the task – the name of the picture – it may be an oversimplification to assume that the picture and word are processed as a single stimulus, and thus that they have interdependent attentional demands.

In fact, evidence suggests that they are not. For the reasons noted above, selecting a lemma for production in picture naming may require central attentional resources (although this hypothesis is first confirmed in Experiment 2). In contrast, word reading may not require central resources. Reynolds and Besner (2006, Experiment 1) conducted a dual-task experiment with Task 1 tone discrimination and Task 2 single word naming in which words were repeated after a large number of intervening trials. Participants were faster to read repeated words aloud on trials in which the word was presented 750 ms after the tone, but there was no effect of repetition at a short task SOA of 50 ms. This underadditive interaction between repetition and task SOA indicates that the orthographic-lexical processing underlying this repetition priming can proceed in parallel with central resource-demanding tone processing, and thus that written forms do not require such resources to activate their lexical representations (see also Cleland, Gaskell, Quinlan, & Tamminen, 2006; Dell'Acqua, Pesciarelli, Jolicœur, Eimer, & Peressotti, 2007; Ruthruff, Allen, Lien, & Grabbe, 2008; but see Lien, Ruthruff, Cornett, Goodin, & Allen, 2008; McCann, Remington, & Van Selst, 2000).

The different attentional demands of processing the picture and word stimuli may interact in an unexpected way when PWI is the second task in a dual-task setting. At short task SOAs, picture naming processes that require central resources – including lemma selection – should be postponed until after the completion of Task 1 response selection, creating cognitive slack. However, because orthographic-lexical processing can occur even when central resources are already engaged, distractor processing should not be similarly postponed, as it could proceed simultaneously with Task 1 response selection (see Figure 2.3a). As a result, whereas the temporal proximity of distractor processing and lemma selection normally gives rise to semantic interference in a single-task setting or at long task SOAs in a dual-task setting, the distractor would be processed substantially earlier relative to lemma selection at short task SOAs. This would be functionally equivalent to presenting the distractor before the picture in a single-task setting, another manipulation that would cause the distractor to be processed earlier relative to lemma selection. The possibility that reducing task SOA affects semantic PWI performance in the same way as reducing distractor SOA due to the different attentional demands of picture naming and word reading will be referred to as the *differential automaticity hypothesis*.

The logic of this hypothesis, adapted from Besner, Reynolds, and O'Malley (2009), could account for the underadditive interaction between semantic relatedness and task SOA in the semantic PWI task. As noted, Glaser and Düngelhoff (1984) found in a

Figure 2.3. Logic of Experiment 1 for trials with semantically related distractor words. Task 1 = tone discrimination; Task 2 = picture naming. PE = perceptual encoding; RS = response selection; RE = response execution; SOA = stimulus onset asynchrony; PWI = picture-word interference. (a) When the word and picture are presented simultaneously, PWI emerges at a long task SOA due to the temporal proximity of word processing to lemma selection. Reducing the delay between tone and picture temporally separates word processing (which does not require central attentional resources) and lemma selection (which does), causing PWI to diminish. (b) When the word is presented after the picture, no PWI is generated at a long task SOA because the word is processed too late to interfere with lemma selection. However, reducing the delay between the tone and picture causes the long distractor SOA and short task SOA to cancel out: Word processing re-synchronizes with lemma selection, leading to the emergence of PWI.


Task 1
(a)
0 ms Distractor SOA
Long
Task SOA
Task 2
PE
Word processing
Lemma
Selection
PWI
Other RS
RE
Short
Task SOA
Task 2
PE
Word processing
Lemma
Selection
Other RS
RE
(b)
250 ms Distractor SOA
Long
Task SOA
Task 2
PE
Lemma
Selection
Other RS
RE
Word processing
Short
Task SOA
Task 2
PE
Lemma
Selection
Word processing
PWI
Other RS
RE
time

single-task experiment that same-category semantic distractor words only slowed picture naming relative to unrelated distractors when the word was presented between 100 ms before and 100 ms after the picture. Importantly, when the distractor was presented several hundred milliseconds before the picture, no effect of semantic relatedness was observed, presumably because distractor activation had decayed prior to lemma selection. A short task SOA in a dual-task setting could lead to the elimination of semantic relatedness for the same reason, as central resource-demanding lemma selection would be postponed due to the central bottleneck and thus would occur too late to be affected by non-postponed distractor processing. Thus, it may not be the case that semantic PWI is absorbed into slack at a short task SOA (Dell'Acqua, Job et al., 2007); instead, the interference may never be generated in the first place.

Experiment 1

Under the differential automaticity hypothesis, when there is cognitive slack (i.e., at short task SOAs), decreasing either task SOA or distractor SOA by a particular length of time will reduce semantic PWI to the same extent. This is because both manipulations increase the delay between distractor word processing and lemma selection by the same amount. (In Figure 2.3a, decreasing task SOA would cause both Task 2 perceptual encoding and word processing to shift to the left, while decreasing distractor SOA would cause only word processing to shift to the left; either way, since lemma selection is subject to the central bottleneck, the delay between word processing and lemma selection would increase.) It follows that decreasing task SOA and increasing distractor SOA by the same length of time should have *no* effect on semantic PWI because the two manipulations should cancel each other out, leaving the timing of word processing unchanged. More generally, modifications to task SOA and distractor SOA are interchangeable as long as Task 2 processing is bottlenecked (i.e., there is cognitive slack). Thus, the chief determinant of whether or not semantic PWI emerges in such conditions should be the (signed) sum of the task SOA and the distractor SOA.

Experiment 1 tests this prediction by varying distractor SOA, presenting the word either simultaneously with the picture (i.e., at a 0 ms distractor SOA) or 250 ms after the picture. If semantic PWI disappears at short task SOAs because word processing occurs too early relative to lemma selection, delaying the presentation of the word relative to the picture should close the gap. This is best illustrated by comparing the short task SOA conditions in Figures 2.3a and 2.3b: At a short task SOA, increasing distractor SOA causes word processing and lemma selection to take place closer together in time, thereby re-introducing semantic interference.

Under this account, when the word and picture are presented simultaneously, semantic PWI should be evident only at long (1000 ms) and medial (350 ms) task SOAs, as Dell'Acqua, Job et al. (2007) showed. In contrast, when the word is presented 250 ms after the picture, semantic interference should not be evident at long and medial task SOAs because the distractor will be processed too late to affect lemma selection, but it should arise at a short task SOA (100 ms). This is due to the fact that pictures presented at a task SOA of 100 ms and a distractor SOA of 250 ms should show the same amount of semantic PWI as pictures presented at a task SOA of 350 ms and a distractor SOA of 0 ms, because the difference in time between distractor processing and lemma selection should be the same in both conditions (100 + 250 = 350 + 0).

Method

Participants. Forty-eight members of the University of California, San Diego community participated in Experiment 1. Participants received class credit for their participation. All reported English as their native language.

Apparatus. Stimuli were presented on an iMac computer running PsyScope X (Build 51; Cohen, MacWhinney, Flatt, & Provost, 1993; http://psy.ck.sissa.it). Three buttons on an ioLab button box were used to collect responses to the tone task. A Shure SM10A headworn microphone connected to the button box measured voice onset latencies.

Materials. The picture-naming materials were taken from Damian and Martin (1999, Experiments 1 and 2). Ferreira and Pashler (2002, Experiment 2) used the same set of pictures, excluding one (*ring*) at random for counterbalancing reasons; the same picture was excluded here. The 27 pictures were line drawings of common objects. Two written distractors were selected for each picture: one that was a member of the same semantic category as the picture, and one that was semantically unrelated to the picture. Related and unrelated distractor sets were matched with respect to length in terms of both letters and phonemes. All materials are reported in Damian and Martin (1999).

The acoustic materials were taken from Dell'Acqua, Job et al. (2007). Low, medium and high tones were pure tones at frequencies of 300, 600 and 1200 Hz, each lasting 50 ms in duration.

Design and analysis. Experiment 1 included three independent variables: (a) task SOA (the picture was presented either 100, 350 or 1000 ms after the tone, as in Dell'Acqua, Job et al., 2007), (b) distractor SOA (the distractor was presented either 0 or

250 ms after the picture), and (c) distractor relatedness (semantically related or unrelated). Distractor SOA was manipulated between blocks, with the order of blocks counterbalanced across participants; task SOA and distractor relatedness were manipulated within block.

As in Ferreira and Pashler (2002), participants were presented with 162 trials in an order determined by one of six stimulus lists. Pictures were presented in the same fixed order in every list; however, the conditions in which a picture was presented on a particular trial were different in each list. Pictures were not fully crossed with conditions within participant because doing so would have necessitated doubling the number of trials; as a result, no participant named the same picture presented with both its related and unrelated distractors at the same combination of task SOA and distractor SOA. Instead, every picture was presented to each participant once at each of the six SOA combinations. Within these six presentations, it was paired with each of its two distractors once at each task SOA. Each combination of picture and distractor was presented to each participant once at one distractor SOA and twice at the other. Across the 48 participants, every picture was presented in every condition 24 times.

Procedure. Before beginning the experiment, participants practiced the tone discrimination and naming tasks in four practice blocks, similar to Ferreira and Pashler (2002). In the first block, they viewed each picture accompanied by its correct name. In the second block, the three tones were presented and labeled so participants could distinguish them. Then, they had to discriminate a series of 45 tones by pressing one of three buttons to identify each one as low, medium or high. Each of the 27 pictures was presented once in isolation in the third block and once again in the fourth practice block, when the two tasks were combined on each trial. Participants were told that their primary task was to identify the pitch of the tone, and that they should always do so before naming the picture. Participants who used the wrong name for a picture in any practice block were corrected.

Trials in the actual experiment were structured the same as in the fourth practice block. Each trial began with a fixation point presented for 1000 ms. After its offset, a 500 ms delay was followed by a randomly presented tone. Either 100, 350 or 1000 ms after the onset of the tone, a picture was presented until the voice key registered a response. A centered, written distractor word was presented in 24-point font either simultaneously with the picture or 250 ms after its onset. As in Damian and Martin (1999, Experiment 2) and Ferreira and Pashler (2002, Experiment 2), the word was displayed for 200 ms, after which it was replaced by a visual mask (XXXXXXXX) for 500 ms. Because the word duration exceeded the threshold for conscious detection, the effect of the mask on semantic interference was likely modest (compare Damian & Martin, 1999, Experiments 1 and 2).

After the participant responded to both stimuli, the experimenter coded the accuracy of the vocal response and whether there was a voice key error, which arose when the microphone mistakenly recorded a response that was earlier or later than the actual onset of speech, or when the participant began an utterance with a filler word (e.g., “Um”). The next trial began 1500 ms after the experimenter coded the response. There were two blocks of 81 trials each (one for each distractor SOA), with a short break provided between blocks.

Results

In keeping with the participant exclusion procedures used by Schnur and Martin (2012), six participants were removed who made at least 20% errors in the tone discrimination task. In addition, another participant was removed who failed to follow instructions and responded to the picture before responding to the tone on at least 10% of trials.

The other 41 participants provided data for 6,642 trials, of which 87.5% (5,809) were analyzed. Trials were excluded when a participant responded to the stimuli out of order (38), or responded to the tone (343) or the picture (65) incorrectly; when the voice key failed to register the participant's response at the appropriate time (151); or due to experimenter error (2). Among remaining trials, those in which participants responded to the tone faster than 300 ms (6) or slower than 2000 ms (6), or responded to the picture faster than 300 ms (7) or slower than 3000 ms (15), were excluded. In addition, trials in which a participant had an RT for either the tone discrimination task (151) or picture naming task (152) that was at least 2.5 standard deviations greater than their mean RT for the same combination of task, task SOA and distractor SOA were excluded. (Note that some trials violated multiple criteria.)

The mean reaction times for the tone discrimination (RT1) and naming tasks (RT2) were analyzed separately, each using two 3 × 2 × 2 analyses of variance (ANOVAs) that treated participants (*F*1) and pictures (*F*2) as random variables. To assess the effects of distractor relatedness, planned comparisons were conducted within each combination of task SOA and distractor SOA. Variability is reported with 95% confidence interval halfwidths (Loftus & Masson, 1994).

Participants' mean RTs to both tasks as a function of task SOA, distractor SOA and distractor relatedness are shown in Figure 2.4.

Task 1 performance. Participants responded to tones 37 ms slower on trials with the least task overlap than trials with maximal or medial overlap, indicated by a main effect of task SOA, $F1(2, 80) = 5.40$, CI = ±26 ms, $p = .006$; $F2(2, 52) = 34.1$, CI = ±10 ms, $p < .001$). No other factors or interactions affected tone discrimination latencies (all *F*s < 2.5, all *p*s > .12).

Planned comparisons within each combination of task SOA and distractor SOA conditions revealed a potential effect of distractor relatedness only on trials in which the task SOA was 100 ms and the distractor SOA was 0 ms (henceforth referred to as the 100/0 condition), such that tones were responded to 20 ms slower when the picture and distractor were unrelated than when they were related. This effect was statistically significant only by participants, $F1(1, 80) = 4.07$, CI = ±19 ms, $p = .047$; $F2(1, 52) = 1.53$, CI = ±41 ms, $p = .222$; however, because this is the same combination of SOA conditions in which researchers have found conflicting results concerning the effects of distractor relatedness in Task 2 picture naming, the implications of this 20 ms difference will be addressed in the Discussion. The effect of distractor relatedness was not significant in any other combination of SOA conditions (all *F*s < 2.3, all *p*s > .13).

Task 2 performance. A robust PRP effect was observed, such that participants named pictures 219 ms slower when they were presented 350 ms after tones than when they were presented 1000 ms after tones, and an additional 196 ms slower when they were presented 100 ms after tones. These differences reflect the postponement of Task 2 central resource-demanding processes that occurs when stimuli are presented closer in

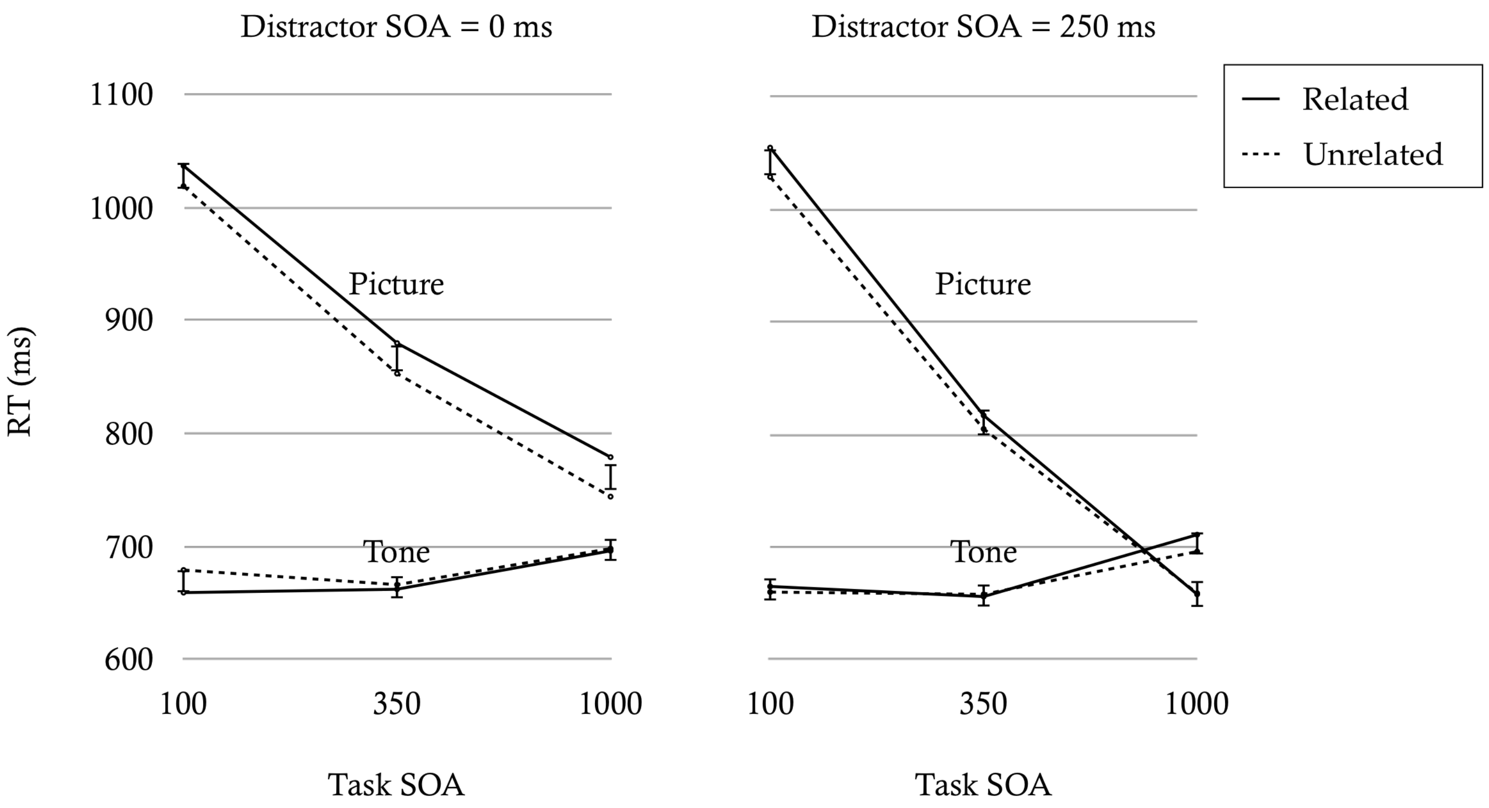


Figure 2.4. Experiment 1 tone discrimination (Task 1) and picture naming (Task 2) latencies shown as a function of task stimulus onset asynchrony (SOA), distractor SOA, and distractor relatedness. Error bars for each task represent 95% confidence interval halfwidths; pairwise differences between condition means that exceed error bar length are statistically significant.

time, as indicated by a significant main effect of task SOA, *F*1(2, 80) = 278, CI = ±28 ms, $p < .001$; *F*2(2, 52) = 1535, CI = ±12 ms, $p < .001$. Participants also named pictures 20 ms slower when distractors were related than when they were unrelated, and 49 ms slower when distractors were presented simultaneously than when they were delayed, as indicated by significant main effects of distractor relatedness, *F*1(1, 40) = 18.9, CI = ±9 ms, $p < .001$; *F*2(1, 26) = 21.2, CI = ±9 ms, $p < .001$, and distractor SOA, respectively, *F*1(1, 40) = 50.1, CI = ±14 ms, $p < .001$; *F*2(1, 26) = 100, CI = ±11 ms, $p < .001$.

However, the size of this latter effect was not constant across task SOAs: As the stimuli were presented closer in time, the effect of distractor delay (collapsing across relatedness) decreased, as indicated by an interaction between task SOA and distractor SOA, *F*1(2, 80) = 38.4, CI = ±19 ms, $p < .001$; *F*2(2, 52) = 63.9, CI = ±14 ms, $p < .001$. Relative to pictures with simultaneously presented distractors, pictures with delayed distractors were named 103 ms faster at a 1000 ms task SOA, 56 ms faster at a 350 ms task SOA, and 13 ms *slower* at a 100 ms task SOA, though post hoc tests revealed that only the first two of those pairwise comparisons were statistically significant, 1000 ms: *F*1(1, 80) = 119, CI = ±19 ms, $p < .001$; *F*2(1, 52) = 226, CI = ±14 ms, $p < .001$; 350 ms: *F*1(1, 80) = 34.9, CI = ±19 ms, $p < .001$; *F*2(1, 52) = 63.1, CI = ±14 ms, $p < .001$; 100 ms: *F*1(1, 80) = 1.99, CI = ±19 ms, $p = .162$; *F*2 < 1. In addition, there was a trend toward more semantic interference from related than unrelated distractors when the distractor was presented simultaneously (27 ms) than when the distractor was delayed (12 ms), as indicated by an interaction between distractor SOA and distractor relatedness that was marginally significant only by items, *F*1(1, 40) = 2.18, CI = ±14 ms, $p = .150$; *F*2(1,

26) = 4.13, CI = ±14 ms, *p* = .053. No other interactions affected picture naming latencies (all *F*s < 1.8, all *p*s > .18).

Planned comparisons were conducted to determine the timing conditions under which semantically related distractors interfered with picture naming. When distractors were presented simultaneously, semantic interference effects of 35 ms and 28 ms were observed when pictures were presented 1000 ms and 350 ms after tones, respectively, though this latter effect was marginally significant by items; 1000 ms: *F*1(1, 80) = 8.54, CI = ±24 ms, *p* = .005; *F*2(1, 52) = 5.63, CI = ±35 ms, *p* = .021; 350 ms: *F*1(1, 80) = 5.17, CI = ±24 ms, *p* = .026; *F*2(1, 52) = 3.44, CI = ±35 ms, *p* = .069. However, in the crucial condition in which pictures were presented only 100 ms after tones, the interference shrank to a non-significant 18 ms, *F*1(1, 80) = 2.16, CI = ±24 ms, *p* = .146; *F*2 < 1. Statistically, this finding agrees with Dell'Acqua, Job et al. (2007) and Ayora et al. (2011), who found no interference from semantically related distractors at short task SOAs.

When distractors were presented 250 ms after pictures, non-significant semantic interference effects of -1 ms and 12 ms were observed when pictures were presented 1000 ms and 350 ms after tones, respectively, all *F*s < 1. This was expected, because the distractor was presented too late to affect picture-naming processes. However, when pictures were presented only 100 ms after tones, related distractors slowed picture naming by 26 ms relative to unrelated distractors, though this effect was marginally significant by items, *F*1(1, 80) = 4.69, CI = ±24 ms, *p* = .033; *F*2(1, 52) = 2.85, CI = ±35 ms, *p* = .097.

In light of this predicted pattern of semantic interference – namely, that interference became non-significant at the short task SOA when words were presented simultaneously, but appeared at that same task SOA when words were delayed – it is surprising that a three-way interaction between task SOA, distractor SOA, and distractor relatedness was not observed. To increase statistical power, a post hoc contrast was conducted between the short and long task SOAs, comparing how task overlap affected the semantic interference effects differently for the two distractor SOAs. In essence, this is equivalent to computing the three-way interaction only across the short and long task SOAs. This contrast was marginally significant by both participants and by items, $F1(1, 80) = 3.49$, CI = ±12 ms, $p = .065$; $F2(1, 52) = 2.91$, CI = ±17 ms, $p = .094$, indicating that reducing task SOA from 1000 ms to 100 ms marginally affected semantic interference by different amounts for the two distractor SOAs – specifically, by reducing interference when distractors were presented simultaneously (Figure 2.3a) and by increasing interference when distractor presentation was delayed (Figure 2.3b).

Discussion

Experiment 1 showed that in a dual-task paradigm, simultaneously presented, semantically related distractor words interfered with picture naming except when they were presented 100 ms after tones. Conversely, semantically related distractor words presented after pictures interfered with picture naming *only* when they were presented 100 ms after tones. Furthermore, reducing task SOA from 1000 ms to 100 ms marginally affected semantic interference by different amounts at the two distractor SOAs. This pattern of data was predicted by an account that treats picture naming and word reading as two distinct tasks with dissociable attentional requirements, and – most importantly – it is compatible with accounts of semantic PWI that claim it occurs during (Levelt et al., 1999) or after (Mahon et al., 2007) lemma selection.

In addition to the expected main effects, there was a crossover interaction between task SOA and distractor SOA. Specifically, as task SOA decreased, increasing the overlap between tasks, the competition from delayed distractors (regardless of relatedness) increased more than the competition from simultaneously presented distractors, as evidenced by the steeper slopes for the RT2 curves in the right panel than in the left panel of Figure 2.4. This is important because it provides another source of evidence that, relative to simultaneously presented distractors, decreasing task SOA causes delayed distractors to be processed closer in time to central resource-demanding picture processing.

Consistent with this, it is worth noting that at every task SOA, RTs were slower in the distractor SOA condition that gave rise to more semantic interference, suggesting that both semantic interference and mean RTs were affected by the timing of distractor processing.[2] This relationship is concordant with the results of Glaser and Düngelhoff (1984), who found that both semantic interference and mean RTs peaked around a 0 ms distractor SOA. When they presented distractors simultaneously with the picture, compared with 300 ms after the picture, interference increased from 20 ms to 84 ms, and the mean RT for related and unrelated distractors increased from 614 ms to 752 ms.

---

[2] As noted, the 13-ms difference at the 100 ms task SOA was not significant. While this contradicts a prediction of the differential automaticity hypothesis, direct comparisons between different levels of distractor SOA may be confounded with other factors; for example, participants may simply have a harder time visually separating the picture and distractor when they are simultaneously presented.

The results of Experiment 1 may also be consistent with Dell'Acqua, Job et al.'s (2007) account. Under that account, semantic PWI is always generated when a related distractor is presented simultaneously with a picture, but this PWI is absorbed into slack at short task SOAs (see Figure 2.2). Although Dell'Acqua and colleagues did not manipulate distractor SOA or describe how such manipulations would be expected to affect picture naming latencies, it is reasonable to assume that delaying distractor presentation would similarly postpone semantic PWI resolution. Because PWI resolution is normally followed by cognitive slack, postponing that resolution would cause it to migrate across the slack. If the distractor SOA were long enough, the PWI resolution would begin to overlap with Task 1 response execution. At this point, because PWI resolution (under their account) must finish before lemma selection can begin, it would start to postpone Task 2 lemma selection, leading to an increase in picture naming latencies. Thus, the results of Experiment 1 do not necessarily rule out a locus of semantic PWI resolution that precedes lemma selection. However, as they also support interpretations that are consistent with existing models of the task (Levelt et al., 1999; Mahon et al., 2007), it is no longer necessary to fundamentally revise those models.

Nevertheless, several details slightly complicate this story. In particular, although the amount of semantic interference in the 100/0 condition did not reach statistical significance, the size of the effect was still 18 ms. This is closer in size to the effects observed in the 100/0 condition by Schnur and Martin (2012), who found significant semantic interference effects of 30 ms and 25 ms in two experiments, than it is to the effects observed by Dell'Acqua and colleagues, who found non-significant interference effects of -7 ms (Dell'Acqua, Job et al., 2007) and 2 ms (Ayora et al., 2011; Experiment

2). One possible reason for this, which hinges on the phonological regularity of the distractors, will be explored further in the General Discussion.

A further complication was the significant effect of semantic relatedness on the tone discrimination RTs in the 100/0 condition, which may have dampened the interference effect in picture naming. As tones were identified 20 ms faster when they were presented before pictures with related distractors than pictures with unrelated distractors, and RT1 and RT2 are tightly coupled at short task SOAs, that could have reduced the effect of semantic interference on picture naming latencies by 20 ms.

In most dual-task studies, RT1 effects would not necessarily change the interpretation of RT2 results. Some theoretical alternatives to the central bottleneck model claim that people can divide their central attentional resources between multiple tasks (e.g., Navon & Miller, 2002; Tombu & Jolicœur, 2003), allowing them to perform response selection for multiple tasks simultaneously. Counterintuitively, this capacity sharing causes RT1 to increase while leaving RT2 unchanged. This would normally mean that the difference in tone RTs in the 100/0 condition, which could be explained by positing that more capacity sharing occurred for unrelated distractors than for related distractors, could not be taken as evidence of a suppressed semantic interference effect in picture naming latencies.

However, that logic only applies when the stages of Task 2 are fixed in duration regardless of when they occur. If participants divided their attention between tasks in the present experiment, they would have (inefficiently) engaged in lemma selection at the same time as Task 1 response selection. On trials with simultaneously presented distractors, this would cause lemma selection to begin at the same time as in a single-task setting, bringing it closer in time to word processing. If this capacity sharing happened more often on trials with unrelated distractors, picture naming latencies on those trials might be relatively longer because the (unrelated) word would interfere more with lemma selection. Thus, even under a capacity sharing account, the effect of distractor relatedness on tone RTs would potentially indicate a suppressed semantic interference effect on picture naming RTs, a possibility that cannot, at present, be ruled out.

Nevertheless, the existence of interference suppression in the 100/0 condition would not account for the semantic interference effect from delayed distractors that emerges at the shortest task SOA or the interaction between task SOA and distractor SOA. For these reasons, under both the central bottleneck model and capacity sharing models, the picture and distractor word are best treated as separate stimuli with dissociable attentional requirements as described above.

Experiment 2

Under the account advanced in Experiment 1, semantic PWI with simultaneously presented distractors leads to the dissipation of semantic interference at the shortest task SOA because picture naming requires central attentional resources and word processing does not. If this is true, manipulations of lemma selection difficulty that do not involve word reading should be equally robust at short and long task SOAs. This would confirm the assumption, heretofore untested, that lemma selection requires central attentional resources. Experiment 2 tests the attentional requirements of lemma selection by using a more straightforward manipulation of its difficulty: cumulative semantic interference (CSI).

In the standard CSI paradigm, participants name a series of pictures presented one at a time. Unbeknownst to participants, pictures presented (non-consecutively) throughout the experiment constitute semantic categories (e.g., farm animals: *cow*, *horse*, *donkey*, *sheep*, *pig*). Picture naming latencies increase, by an approximately linear amount, as a function of how many previous category members have been named. For example, participants who previously took 610 ms to name *cow* take 635 ms to name *horse*; later, they will take 661 ms to name *donkey* (Howard, Nickels, Coltheart, & Cole-Virtue, 2006). Although there is disagreement over how exactly cumulative semantic interference arises (Howard et al., 2006; Oppenheim, Dell, & Schwartz, 2010), all accounts agree that naming *cow* will slow the subsequent naming of *horse* during *horse*'s lemma selection. Thus, this paradigm can be used as the second task in a dual-task experiment to probe the attentional requirements of lemma selection.

If, as predicted, lemma selection requires central attentional resources, equal cumulative semantic interference effects will be observed at short and long task SOAs. If lemma selection does not require such resources, however, the interference effect will be absorbed into slack at the short task SOA, reducing or completely flattening the slope of the interference curve relative to the long task SOA.

Method

Participants. Forty new members of the University of California, San Diego community participated in Experiment 2.

Apparatus. The same apparatus used in Experiment 1 was used in Experiment 2.

Materials. The picture-naming materials for Experiment 2 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). 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 only one category of animals). None of the 34 filler pictures belonged to any of the 12 categories.

The same acoustic materials used in Experiment 1 were used in Experiment 2.

Design and analysis. Experiment 2 included two independent variables: (a) task SOA (the picture was presented either 100 or 1000 ms after the tone), and (b) ordinal position (the picture was the first, second, third, fourth or fifth member of a semantic category presented within a given block). Both variables were manipulated within block.

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 within each block. 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 in a block was assigned to one of twelve unique lag sequences (out of a possible 24; 4!), and the order of categories within each block was counterbalanced across participants.

Every picture was presented once in each of two blocks, leading to 188 trials per participant. Within each block, all of the pictures in half of the categories, as well as half of the fillers, were presented at each task SOA. The task SOA of every picture switched between blocks. Across the 40 participants, every critical picture was supposed to be presented in every combination of block, ordinal position and task SOA four times; however, due to a counterbalancing error, 5% of combinations were presented two times and 5% of combinations were presented six times. Omitting these combinations did not affect either the pattern of data or any of the statistical analyses for picture naming latencies; thus, they are included in all analyses reported here.

Procedure. Before beginning the experiment, participants practiced the tone discrimination and naming tasks in two picture blocks. In the first block, participants practiced tone discrimination as in Experiment 1; however, in an attempt to reduce the frequency of tone errors, participants were given feedback when they pressed the wrong button. In the second block, participants practiced combining the tone task with the picture naming task in 54 trials (the same number of dual-task practice trials in Experiment 1). Half of the pictures were presented at each of the two task SOAs of 100 and 1000 ms. Practice pictures were not used in the experiment and did not belong to any of the 12 critical categories. As before, participants were told that their primary task was to identify the pitch of the tone.

Trials in the actual experiment were structured as in Experiment 1, except that only two task SOAs were used (100 and 1000 ms), and no distractors were presented. A short break was provided between the two blocks.

Results

Two participants were removed who made at least 20% errors in the tone discrimination task. In addition, another participant was removed who responded to the picture before responding to the tone on at least 10% of trials.

The other 37 participants provided data for 4,440 critical trials, of which 81.8% (3,631) were analyzed. Trials were excluded when a participant responded to the stimuli out of order (21), or responded to the tone (287) or the picture (235) incorrectly; when the voice key failed to register the participant's response at the appropriate time (101); or due to experimenter error (13). Among remaining trials, those in which participants responded to the tone faster than 300 ms (1) or slower than 2000 ms (28), or responded to the picture faster than 300 ms (2) or slower than 3000 ms (9), were excluded. In addition, trials in which a participant had an RT for either the tone discrimination task (97) or picture naming task (84) that was at least 2.5 standard deviations greater than their mean RT for the same combination of task and task SOA were excluded. (Note that some trials violated multiple criteria.)

As the effect of block number did not interact either with ordinal position (coded either nominally or linearly; all *F*s < 2.5, all *p*s > .12) or with task SOA (all *F*s < 1.3, all *p*s > .31), data were averaged first within and then between blocks. The mean reaction times for the tone discrimination (RT1) and naming tasks (RT2) were analyzed separately, each using two 2 × 5 ANOVAs that treated participants (*F*1) and categories (*F*2) as random variables. To assess the linear effects of ordinal position, planned linear contrasts were conducted within each level of task SOA.

Participants' mean RTs to both tasks as a function of task SOA and ordinal position are shown in Figure 2.5.

Task 1 performance. Participants responded to tones 69 ms slower on trials with more task overlap, indicated by a main effect of task SOA, *F*1(1, 36) = 29.4, CI = ±21 ms, $p < .001$; *F*2(1, 11) = 36.9, CI = ±21 ms, $p < .001$. Ordinal position also affected tone discrimination latencies, *F*1(4, 144) = 4.07, CI = ±15 ms, $p = .004$; *F*2(4, 44) = 4.43, CI = ±17 ms, $p = .004$, but this effect did not interact with task SOA (both *F*s < 1).

Planned linear contrasts within each level of task SOA assessed the linear effects of ordinal position, revealing that participants generally took longer to identify tones paired with pictures in higher ordinal positions. The average slowdown was 4.5 ms per position at a task SOA of 100 ms, which was significant only by categories, *F*1(1, 144) = 1.90, CI = ±6.4 ms, $p = .170$; *F*2(1, 44) = 4.79, CI = ±6.8 ms, $p = .034$, and 10.4 ms per position at a task SOA of 1000 ms, *F*1(1, 144) = 10.2, CI = ±6.4 ms, $p = .017$; *F*2(1, 44) = 9.79, CI = ±6.8 ms, $p = .003$. These slowdowns were not statistically different from each other, *F*1(1, 144) = 1.64, CI = ±4.6 ms, $p = .202$; *F*2 < 1.

Task 2 performance. Participants named pictures 389 ms slower when they were presented on trials with more task overlap, a robust PRP effect indicated by a main effect of task SOA, *F*1(1, 36) = 261, CI = ±41 ms, $p < .001$; *F*2(1, 11) = 1191, CI = ±21 ms, $p < .001$. Ordinal position also affected naming latencies, *F*1(4, 144) = 10.5, CI = ±19 ms, $p < .001$; *F*2(4, 44) = 6.17, CI = ±28 ms, $p < .001$, but did not do so differently across the two task SOAs (both *F*s < 1).

Planned linear contrasts within each level of task SOA assessed the linear effects of ordinal position, revealing that participants took longer to name pictures in higher

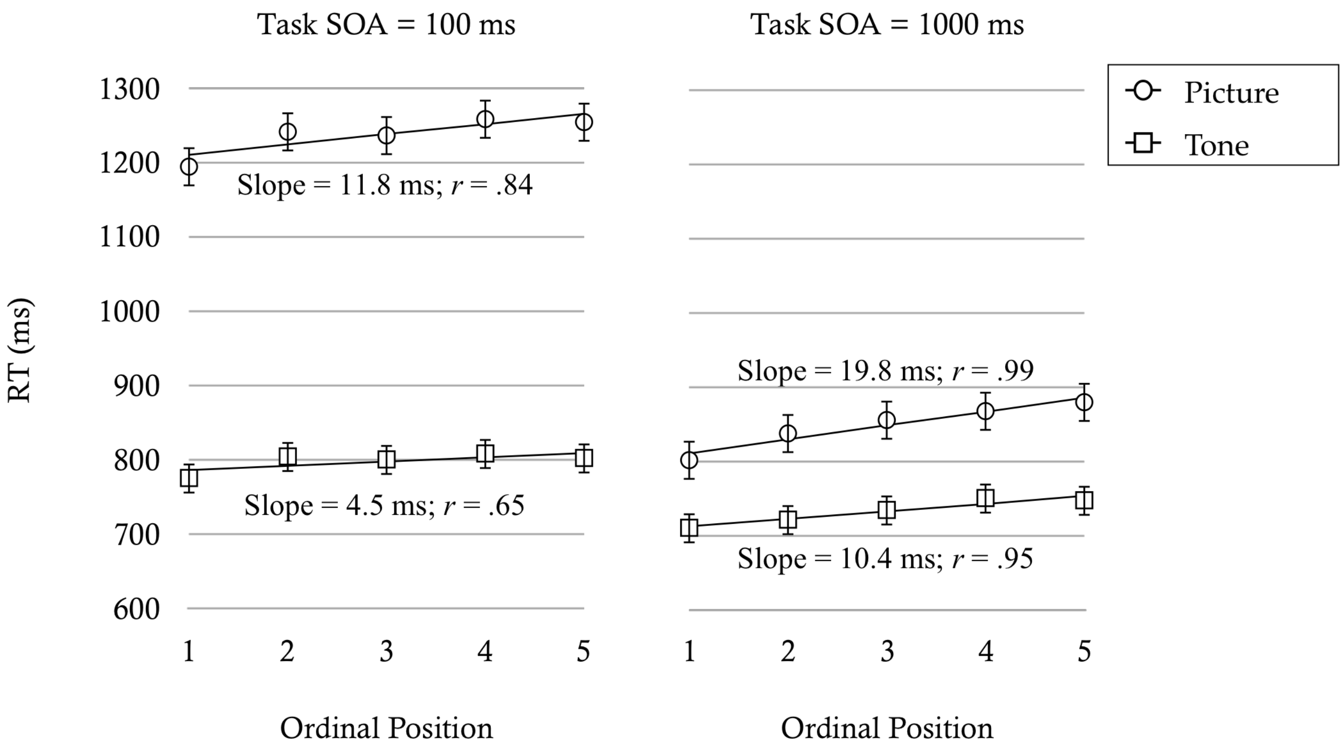


Figure 2.5. Experiment 2 tone discrimination (Task 1) and picture naming (Task 2) latencies shown as a function of task stimulus onset asynchrony (SOA) and ordinal position. Linear effects of ordinal position on reaction times are denoted by best-fit lines. Error bars for each task represent 95% confidence interval halfwidths; pairwise differences between condition means that exceed error bar length are statistically significant.

ordinal positions – the standard cumulative semantic interference effect. The average degree of interference was 11.8 ms per position at a task SOA of 100 ms, *F*1(1, 144) = 8.28, CI = ±8.1 ms, *p* = .005; *F*2(1, 44) = 15.4, CI = ±7.5 ms, *p* < .001, and 19.8 ms per position at a task SOA of 1000 ms, *F*1(1, 144) = 23.2, CI = ±8.1 ms, *p* < .001; *F*2(1, 44) = 29.1, CI = ±7.5 ms, *p* < .001. These interference effects were not statistically different from each other, *F*(1, 144) = 1.89, CI = ±8.1 ms, *p* = .172, *F*(1, 44) = 1.09, CI = ±7.5 ms, *p* = .303.

Given the unexpected slowdown in RT1 as ordinal position increased, the analogous slowdown in RT2 could potentially be attributable to extraneous factors, such as flagging attention, that would increase reaction times to both tasks. To demonstrate that a cumulative semantic interference effect above and beyond such factors exists, it would be sufficient to show that the linear effect of ordinal position on RT2 is greater than on RT1. To compare them, the data were submitted to 2 × 2 × 5 ANOVAs that additionally included task (tone discrimination vs. picture naming) as a factor, and the contrast weights that revealed linear effects of ordinal position on RT1 were subtracted from the contrast weights that revealed linear effects of ordinal position on RT2 within each level of task SOA. At a task SOA of 100 ms, the slowdown as ordinal position increased was greater for picture naming than for tone discrimination, though this difference was only marginally significant by participants, *F*1(1, 144) = 3.45, CI = ±3.9 ms, *p* = .065; *F*2(1, 44) = 4.19, CI = ±3.5 ms, *p* = .047. At a task SOA of 1000 ms, it was greater for picture naming both by participants and by categories, *F*1(1, 144) = 5.67, CI = ±3.9 ms, *p* = .019; *F*2(1, 44) = 7.31, CI = ±3.5 ms, *p* = .010. Thus, a cumulative semantic interference effect was evident at both task SOAs.

Discussion

Experiment 2 showed that in a dual-task paradigm in which pictures were presented after tones, a CSI effect emerged in naming latencies. Not surprisingly, this effect was present under conditions of minimal task overlap, in which the picture naming task was analogous to the standard CSI paradigm. Crucially, however, the effect was also present under conditions of maximal task overlap, when the stimuli were presented only 100 ms apart. Furthermore, the size of the interference effect was statistically equivalent across task SOAs. Because both accounts of the CSI effect agree that the interference slows lemma selection (Howard et al., 2006; Oppenheim et al., 2010), these findings indicate that lemma selection must be fully postponed due to the bottleneck, indicating a response selection or a response execution locus. As previous dual-task studies using picture naming as the first task have indicated a perceptual encoding or response selection locus of lemma selection (Cook, 2007, Experiment 1; Ferreira & Pashler, 2002), the only locus of lemma selection that agrees with all findings from dual-task studies is response selection. Thus, the results of Experiment 2 indicate that lemma selection requires central attentional resources.

Although the linear effect of ordinal position did not significantly interact with block number, the semantic interference effects did decrease between blocks, especially at the short task SOA. In the first block, the average slowdown per position was 18.0 ms at the 100 ms task SOA and 21.5 ms at the 1000 ms task SOA. In the second block, however, those slowdowns decreased to 5.6 ms and 19.8 ms, respectively (or 7.2 ms and 18.3 ms when category was treated as a random factor). This reduction in interference for pictures presented at the short task SOA in the second block (which were the same pictures presented at the long task SOA in the first block) appears to be the result of trial sequence, which becomes evident when the naming latencies are grouped based on the task SOA of the preceding trial. For trials in the second block that followed a short-SOA trial, the mean slowdown per position (with category treated as a random factor due to data sparsity when computing means for each participant) was 22.6 ms at the 100 ms task SOA and 20.6 ms at the 1000 ms task SOA. For trials that followed a long-SOA trial, those slowdowns decreased to 3.5 ms and 7.1 ms, respectively. Thus, the difference in CSI effects in the second block appears to result at least in part from the difference in frequency with which trials at each task SOA followed long-SOA trials: Among trials that contributed data to the analyses, trials at the 100 ms task SOA followed long-SOA trials 60% of the time, as opposed to 43% for trials at the 1000 ms task SOA. Although it is still not apparent why a preceding long-SOA trial should have diminished semantic interference so severely, it appears to have done so for both task SOAs.

One surprising aspect of the data was the existence of CSI-like effects not only on picture naming latencies, but on tone discrimination latencies as well. As noted above, this could have been due to participants slowing down over the course of the experiment. Indeed, in mixed-effects models that included trial number within block as a covariate, trial number significantly predicted RTs in the tone discrimination task (but not the picture naming task).[3]

---

[3] The data included in the ANOVAs were submitted to two mixed-effects models – one for each task – that included fixed effects of trial number within block (a continuous variable) and all main effects and interactions of block number, task SOA and ordinal position (also as a continuous variable). The models also included random factors for participants, categories and pictures, as well as interactions between every fixed effect (except trial number) and each random factor. In accordance with common practice for

An alternate explanation is that, rather than completely postponing lemma selection until after completing tone discrimination, participants were engaging in capacity sharing. If people share capacity more when a secondary task is difficult, that could lead to an increase in RT1 while leaving RT2 unchanged (Tombu & Jolicœur, 2003). (Note that, unlike in Experiment 1, this logic holds for Experiment 2 because the difficulty of lemma selection on a given trial should not change regardless of whether it is performed simultaneously with, or delayed until after, Task 1 response selection.) Importantly, this would not change the interpretation of Experiment 2 results, as they could only be explained by a capacity sharing account if lemma selection requires central resources. However, sharing could have caused tone RTs to increase when picture naming was more difficult, creating the false appearance of a CSI effect on tone discrimination latencies. Such an account could hold even for the long task SOA: Because 13.5% (247/1829) of tone discrimination RTs at that task SOA were longer than 1000 ms, participants occasionally saw the picture before responding to the tone. If those trials are removed, the average slowdown in the 1000 ms task SOA condition as ordinal position increased changes to 3.5 ms per position for the tone discrimination latencies (down from 10.4 ms) and 18.2 ms per position for the picture naming latencies (down

---

large data sets, absolute *t* values greater than or equal to 1.96 are taken to indicate statistical significance. All predictors were centered.

Tone discrimination latencies were significantly slower at higher trial numbers within each block, $\hat{\hat{\beta}} = 6.4$, $t = 4.4$, and at the shorter task SOA, $\hat{\hat{\beta}} = -63.1$, $t = -3.4$, but were not affected by any other main effects or interactions (all *t*s < 1.1). Replicating the results of the ANOVAs, pictures were named more slowly in the first block, $\hat{\hat{\beta}} = -70.3$, $t = -3.6$, at the shorter task SOA, $\hat{\hat{\beta}} = -385.4$, $t = -15.3$, and at higher ordinal positions, $\hat{\hat{\beta}} = 14.6$, $t = 2.4$, but were not affected by any other main effects or interactions, including the critical interaction between task SOA and ordinal position (all *t*s < 0.7).

slightly from 19.8 ms). This suggests that the appearance of a CSI effect in tone discrimination latencies at the long task SOA can be attributed to the difficulty of picture naming on trials with task overlap, rather than nuisance factors that increased reaction times to both stimuli as the experiment progressed. Although it is not possible to do an analogous analysis for the 100 ms task SOA condition, as participants saw the picture before responding to the tone on every trial, it is reasonable to assume that the tone discrimination latencies in that condition were similarly affected. Thus, the magnitude of the CSI effects in Experiment 2 should be construed not as the difference between the effects of ordinal position on the picture naming and tone discrimination RTs, but solely as the effects of ordinal position on picture naming RTs.

General Discussion

Since 2002 (Ferreira & Pashler, 2002), dual-task experiments have been used in combination with PWI to shed light on how stages of word production can be organized into serially ordered, higher-level processes. Dell'Acqua, Job et al. (2007) presented participants with tones followed by pictures that were accompanied by simultaneously presented visual distractors, and found that semantic interference from related distractors disappeared at a short task SOA. As they assumed that lemma selection required the use of central attentional resources, they interpreted these results to mean that, contrary to the predictions of a prominent model of lexical production (Levelt et al., 1999) and a task model of PWI (Mahon et al., 2007), the semantic PWI was resolved prior to lemma selection – that is, the interference was absorbed into slack and resolved concurrently with central resource-demanding tone processing.

Experiments 1 and 2 suggest another interpretation that is consistent not only with the experimental results of Dell'Acqua, Job et al. (2007), but also with existing accounts of the task. They demonstrate that the diminution of semantic interference at a short task SOA is consistent with an account in which semantic PWI is resolved during or after lemma selection. Under the differential automaticity hypothesis, word processing – but not lemma selection, as Experiment 2 showed – can be performed concurrently with central resource-demanding tone processing. When the word and picture are simultaneously displayed shortly after the onset of the tone, this dichotomy causes word processing to occur earlier relative to lemma selection than usual, allowing extra time for the distractor lemma activation to decay and leading to a reduction in semantic PWI. When distractor presentation is delayed relative to picture onset, this temporal separation is reduced at short task SOAs, causing word processing and lemma selection to occur closer together in time and – as Experiment 1 showed – leading to the re-emergence of PWI. As this finding can be accounted for by both the central bottleneck model and capacity sharing accounts of dual-task performance, and is consistent with any account of PWI that predicts greater semantic interference when word processing and lemma selection occur closer together in time (Levelt et al., 1999; Mahon et al., 2007), it is no longer necessary to take the results of Dell'Acqua, Job et al. (2007) as evidence that semantic PWI resolution precedes lemma selection.

Accounting for other PRP experiment data

Given the apparent similarities between semantic PWI and the Stroop task, the proposed account of semantic PWI in dual-task experiments raises an obvious question:

If competition between distractor processing and picture processing was reduced at a short task SOA due to differences in the attentional requirements of word reading and lemma selection, why was the same effect not observed in the Stroop task (Fagot & Pashler, 1992)? To answer this question, it is necessary to consider a number of interrelated differences between the standard Stroop and PWI experiment designs that may affect the magnitude and duration of semantic interference: differences in the size of the response set and the degree of stimulus repetition, differences in the degree of overlap between the distractor set and response set, and differences in the number of semantic categories represented in the response set. Although these differences will initially be considered independently, there is reason to believe that they interact in an important way, as will be demonstrated by an analysis of data collected by Piai et al. (in press).

First, there were differences in the size of the response set and in the frequency of stimulus repetition between the two tasks. Fagot and Pashler (1992) used three colors and three words, repeating each 224 times to each participant during the experiment. In contrast, Dell'Acqua, Job et al. (2007) used 48 pictures, each of which was repeated six times, and 96 words, each of which was repeated three times. La Heij and van den Hof (1995) explored the effects of these manipulations on the size of semantic PWI effects by repeatedly presenting pictures within blocks of trials. When blocks contained only four unique pictures, participants showed semantic interference effects of 3 ms, as compared with 12 ms when blocks contained twelve unique pictures. Furthermore, semantic interference was greater during the first half of the experiment than during the second half. Thus, all else being equal, these data suggest that interference should be larger when response sets are larger and when items are repeated fewer times.

Second, the distractor words and color names overlapped in the Stroop task – in fact, they were identical – whereas picture names and distractor words did not overlap in the PWI task. Either because of response priming or because it is difficult to discount task-relevant responses (Cohen, Dunbar, & McClelland, 1990), distractor words in the response set interfere more with performance in the Stroop (Proctor, 1978) and PWI tasks (Lupker & Katz, 1981) than distractor words that are not also targets. Although the effect of set overlap is contested (Caramazza & Costa, 2000), it seems to be especially large (or at least more reliable) when set size is small (Roelofs, 2001), as is commonly the case in Stroop experiments (e.g., Fagot & Pashler, 1992). Furthermore, the presence of congruent trials, in which the distractor word matches the target word, tends to increase interference effects (Lowe & Mitterer, 1982).

Third, the response set in the Stroop task consisted of words belonging to a single category – colors – whereas picture names in the PWI task belonged to a variety of categories. The contrast between these is akin to the difference between the semantically homogeneous and heterogeneous conditions in a semantic blocking experiment (e.g., Damian, Vigliocco, & Levelt, 2001), in which interference accumulates more rapidly in repeated groups of pictures when they share a semantic category.

Some of these methodological differences (overlap of response and distractor sets, presence of congruent trials, fewer semantic categories) may be able to explain why Stroop interference persisted in the short SOA condition (Fagot & Pashler, 1992), but semantic PWI did not (Dell'Acqua, Job et al., 2007). Other differences (size of response set, degree of repetition) may be inconsistent with such an account. It would be difficult to speculate whether, as a whole, these differences can account for the divergent results.

Fortunately, such speculation is unnecessary, as the question can be addressed empirically. Piai et al. (in press), in an unsuccessful attempt to reproduce the results of Dell'Acqua, Job et al. (2007) in Dutch, conducted six dual-task semantic PWI experiments. In every experiment, participants were presented with a tone and a picture-word stimulus, and had to perform tone discrimination before picture naming. Five of the experiments (1-5) used a short SOA condition of 0 ms, at which the tone, picture and word stimuli were presented simultaneously. Of these, four experiments (2-5) used 32 pictures that belonged to eight different semantic categories with four objects per category. Semantically related distractors were created by pairing each picture with the name of another same-category picture, and unrelated distractors were created by pairing pictures with distractors from another category; thus, the distractor set matched the response set. (Experiment 2 also contained congruent distractors.) Depending on the experiment, pictures were repeated between 5 and 12 times. Across these four experiments, semantic PWI effects at the 0 ms SOA ranged from 27 ms to 51 ms, with an average of 36 ms.

The other experiment that used a 0 ms SOA condition, Experiment 1, was designed to directly compare performance in the Stroop and semantic PWI tasks. As such, Piai et al. (in press) restricted the PWI block to three pictures, all from the same semantic category. Distractors were either congruent (the name of the presented picture), semantically related (the name of a different picture), or a string of Xs. Pictures were repeated 72 times each. At the 0 ms SOA, pictures were named 80 ms slower when paired with incongruent distractors relative to congruent distractors (and 92 ms slower than the string of Xs). Thus, a semantic PWI task with the design of a standard Stroop task showed more than twice as much interference as experiments with a larger response set, fewer repetitions, no congruent trials, and more semantic categories.

Interestingly, the Stroop-like semantic PWI experiment showed more interference than the PWI-like PWI experiments even though some of the methodological differences between them, when considered in isolation, indicated that the reverse should be true. Either the differences that favored more interference for the Stroop-like design outweighed the others, or the identified differences may have interacted in some way. For example, a small distractor set that matches the response set and consists solely of same-category words, combined with extreme repetition, may lead to activation saturation of the shared category node and target/distractor lemmas, thereby slowing the decay of distractor activation. At shorter task SOAs, when interference from written distractors in a standard PWI task would be decaying, this would prolong the interference from the words in the Stroop task (Fagot & Pashler, 1992) and the Stroop-like semantic PWI task (Piai et al., in press, Experiment 1). Regardless of which methodological differences are ultimately responsible, however, the data from Piai and colleagues suggest that the contrast between the reliability of Stroop and semantic PWI effects at a short task SOA cannot be taken as evidence that the distractor word is processed differently in the two tasks.

Because the differential automaticity hypothesis claims that all kinds of PWI should be affected – not just manipulations of semantic relatedness – it is important to consider whether it can be squared with the results of other PRP studies. Several experiments have been conducted to determine whether Task 1 manipulations of phoneme selection propagate to Task 2 RTs, with mixed results (Cook & Meyer, 2008;

Ferreira & Pashler, 2002, Experiment 2; Roelofs, 2008). However, only one paper has reversed the order of tasks to determine whether manipulations of Task 2 phoneme selection difficulty are absorbed into slack. In two experiments with Task 2 picture naming, Ayora et al. (2011) showed that pictures accompanied by simultaneously presented distractors were named faster at both short and long task SOAs when those distractors were phonologically related to picture names relative to when they were phonologically unrelated. If reducing task SOA caused the distractor words to be processed earlier relative to phoneme selection, that would be equivalent to reducing distractor SOA. However, single-task phonological PWI studies have found facilitation not just when words are presented simultaneously with or after pictures (e.g., Schriefers et al., 1990), but also at a -300 ms distractor SOA (Jescheniak & Schriefers, 2001, though distractor words were presented auditorily), and sometimes at a -150 ms distractor SOA (Jescheniak & Schriefers, 2001; but see Schriefers et al., 1990). This means that if a task SOA of 100 ms causes the distractor word to be processed (e.g.) 300 ms earlier relative to phoneme selection than at a 1000 ms task SOA, facilitation should still be observed. Thus, these data are consistent with the differential automaticity hypothesis.

Why do simultaneously presented distractors only sometimes show semantic interference at short task SOAs?

In attempting to figure out why they observed a semantic interference effect in the 100/0 condition when Dell'Acqua, Job et al. (2007) and Ayora et al. (2011) did not, Schnur and Martin (2012) suggested that individual differences concerning the difficulty of Task 1 might be at work. When they analyzed the data from nine participants whom they excluded for committing more than 20% errors in the tone task, they found a semantic interference effect of -45 ms in the 100/0 condition (as compared with 25 ms of interference among non-excluded participants). This is consistent with an account in which participants strategically postpone Task 2 processing after response selection when Task 1 is difficult in order to avoid completing Task 2 before Task 1, causing Task 2 response selection to be absorbed into slack (e.g., Meyer & Kieras, 1997). However, the data from Experiment 1 do not support this hypothesis. Across the six participants who were excluded for committing more than 20% errors in the tone task, the semantic interference effects when distractors were presented simultaneously were 18, 21 and 64 ms for the 100, 350 and 1000 ms task SOA conditions, respectively (compared with 18, 28 and 35 ms among participants who were included in the analysis). In particular, the interference effect in the 100/0 condition was exactly the same magnitude for the two groups.

As difficulty can be reflected in reaction times as well as error rates, such an account might also predict that participants who responded to tones more slowly at short task SOAs should show smaller semantic interference effects at those same SOAs. For each of the 41 participants included in the analysis, the average RT1 for the 100/0 condition (collapsing across distractor relatedness) was correlated with the size of the semantic interference effect on RT2 in the 100/0 condition. There was no relationship between these variables (p = .95), indicating a lack of support for the hypothesis that Task 1 difficulty determines the (dis)appearance of semantic interference.

If Task 1 difficulty cannot explain the unstable nature of semantic PWI in the 100/0 condition, what can? One possibility hinges on individual differences in word reading skill. Evidence from dual-task studies with Task 2 single word naming (Ruthruff et al., 2008) and lexical decision (Lien et al., 2006) indicates that reliance on central attentional resources during word reading may decrease as reading skill increases. Applied to the experiments at hand, skilled readers should show less semantic interference in the 100/0 condition than less skilled readers, since the ability to read words while central resources are engaged elsewhere is precisely what leads to the temporal separation of word and picture processing at short task SOAs. However, without any evidence that the participants tested by Dell'Acqua, Job et al. (2007) and Ayora et al. (2011) were more skilled readers than those tested by Schnur and Martin (2012) and Piai et al. (in press), this account is unable to explain the differences observed between experiments.

Another, more likely possibility is that the phonological regularity of written distractor words may play an important role in determining when the effect appears. In the Dual Route Cascaded model of visual word recognition and reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), when a printed word is to be read aloud, its phonological code – that is, the set of phonemes that constitute its pronunciation – is generated simultaneously via two routes. One, the lexical route, essentially performs a dictionary lookup, correctly retrieving the phonology of the written word and feeding activation in turn to the phoneme system. The other, the sub-lexical route, generates a phonological code through the use of language-specific rules governing the mapping of graphemes to phonemes. The sub-lexical route is the only one that will generate a code for nonwords and novel words because the language system cannot retrieve a stored code for a letter string it has never seen. However, it will generate an incorrect code for a phonologically irregular word that does not follow the normal pronunciation rules; e.g., it will generate a pronunciation for “pint” that rhymes with “mint”.

When they proposed the logic underlying the differential automaticity hypothesis, Besner et al. (2009) investigated the attentional demands of reading phonologically regular and irregular words aloud in a dual-task experiment. They concluded that phonological code retrieval via the lexical route could occur simultaneously with central resource-demanding Task 1 tone processing, but that phonological code generation via the sub-lexical route required central resources and thus was delayed by the central bottleneck.

This distinction may be relevant to the question of why semantic interference has sometimes been observed at short task SOAs, including the 100/0 condition, in English (Schnur & Martin, 2012) and Dutch (Piai & Roelofs, 2013; Piai et al., in press; but see van Maanen et al., 2012), but not in Italian (Ayora et al., 2011; Dell’Acqua, Job et al., 2007). Italian has a relatively shallow orthography, as the correct sequence of phonemes for any Italian word can be derived from its spelling (Lepschy & Lepschy, 1991). Thus, the phonological codes generated by the lexical route and the sub-lexical route will be identical. In contrast, English has a deep orthography and contains many phonologically irregular words with exceptional spelling-to-sound mappings, and Dutch falls between Italian and English on the orthographic depth spectrum (Seymour, Aro, & Erskine, 2003). Thus, in these two languages, the phonological codes generated by the two routes will sometimes conflict.

Although the picture-word interference task used in Experiment 1 did not require participants to read distractor words aloud, it is likely that participants still generated those words' phonological codes. Evidence for this comes from the facilitative effects of phonologically related distractors (e.g., Schriefers et al., 1990), as well as from a variety of other psycholinguistic paradigms (see Frost, 1998 for a review). Because the phonological codes generated by the lexical and sub-lexical routes for irregular English distractor words conflict, interference between the two codes should prolong distractor processing, reducing the effects of distractor decay in the 100/0 condition and leading to the emergence of semantic interference.[4] Alternatively, the code generated by the sub-lexical route might simply refresh the activation of the distractor lemma more when it conflicts with the code generated by the lexical route. Either way, because these codes do not conflict for regular distractor words – a set that includes every written distractor in Italian – distractor decay would proceed as normal in the 100/0 condition, leading to the disappearance of semantic interference.

If this is correct, any evidence for a potential effect of semantic relatedness in the 100/0 condition should have come exclusively from trials on which related distractors were phonologically irregular. In contrast, trials with phonologically regular related distractors should more closely replicate the results of Dell'Acqua, Job et al. (2007). To test this hypothesis, the regularity of the distractors used in Experiment 1 was computed using the Dual Route Cascaded model's GPC Strength Calculator (Coltheart et al., 2001). Grapheme-phoneme correspondences, a continuous measure of phonological regularity, were computed independent of rule position, using mean summed frequency. A median

---

[4] Though Besner et al. (2009) found that the effect of phonological regularity on reading aloud time at a short task SOA was eliminated, participants still made more errors at that SOA when reading irregular words than regular words. This increased difficulty is consistent with the present account.

split on these correspondences was performed within each level of distractor relatedness (related and unrelated), and participant means were computed separately for each level of regularity.[5]

The results are presented in Table 2.1. On trials with phonologically regular distractors, PWI decreased for simultaneously presented distractors by 15 ms and increased for delayed distractors by 26 ms as task overlap increased – the same overall pattern observed in Experiment 1. In addition, the effect of semantic relatedness was much smaller on trials with regular distractors: In the critical 100/0 condition, PWI shrank to 5 ms, as compared with a 51 ms effect on trials with irregular distractors. The same patterns, only more pronounced, were present in the item (distractor) means: PWI decreased for simultaneously presented, regular distractors by 23 ms and increased for delayed, regular distractors by 50 ms. Furthermore, the PWI effect in the 100/0 condition was -9 ms for regular distractors as compared with 56 ms for irregular distractors.

Two conclusions can be drawn from these analyses. First, there was a near-total absence of semantic interference in the 100/0 condition on trials with phonologically regular distractors, which contrasted with robust interference generated in the same condition by phonologically irregular distractors. Given that all Italian words, but not all English or Dutch words, are phonologically regular, this result can accommodate both the consistent lack of semantic interference in the 100/0 condition observed by Dell’Acqua

---

[5] The 27 distractors in each relatedness condition were split into two groups of 13. The median distractors, *plum* (related) and *needle* (unrelated), were excluded because they were associated with relatively extreme naming latencies and there was no principled method of determining whether they should group with more regular or less regular distractors. However, grouping them with more regular distractors disrupted the monotonic PWI trend in the delayed distractor condition reported here.

Table 2.1. Semantic picture-word interference effects in Experiment 1 by task SOA, distractor SOA, and distractor phonological regularity.

Distractor regularity and distractor SOA	Task SOA 100 ms	Task SOA 350 ms	Task SOA 1000 ms
Regular			
0 ms	5	16	20
250 ms	15	11	-11
Irregular			
0 ms	51	52	62
250 ms	52	20	14

Note. SOA = stimulus onset asynchrony; ms = milliseconds. Phonological regularity of distractors was determined by a median split on grapheme-phoneme correspondence within each level of distractor relatedness (see text for details).

and colleagues in Italian (Ayora et al., 2011; Dell'Acqua, Job et al., 2007) and the inconsistent effects of semantic interference in both English (Schnur & Martin, 2012; Experiment 1 of this paper) and Dutch (Piai & Roelofs, 2013; Piai et al., in press; van Maanen et al., 2012). Second, an increase in semantic interference for delayed distractors as task overlap increased was observed for phonologically regular distractors. This is important because it indicates that the differential automaticity hypothesis can explain the absence of interference in the 100/0 condition precisely under those conditions that fail to elicit it – namely, when distractors are phonologically regular.

In contrast to its effects on semantic PWI, the effect of phonological regularity on Stroop interference in dual-task settings is likely to be more limited. All else being equal, phonologically irregular color words should show more interference at short task SOAs than phonologically regular color words. However, with the methodological differences between Stroop and PWI experiments interacting to prolong Stroop interference, distractor activation would likely persist across Task 2 cognitive slack regardless of phonological regularity.

Conclusions

Picture-word interference, especially when used in a dual-task paradigm, is a complex task. In the context of the central bottleneck model and PRP logic, it is tempting to assume that each task consists of a single perceptual encoding stage, a single response selection stage, and a single response execution stage. However, such an assumption suffers from oversimplification (potentially like the central bottleneck model itself; cf. Magen & Cohen, 2010; Meyer & Kieras, 1997; Roelofs & Piai, 2011; Tombu & Jolicœur, 2003). As the task name suggests, picture-word interference results from the conflict between the processing of two stimuli – a picture and a word – that have different attentional demands. Picture naming processes at least as early as lemma selection require central attentional resources. In contrast, words can activate their corresponding lemmas without relying on central resources (Reynolds & Besner, 2006), though the generation of phonological codes itself consists of two subprocesses, one of which requires central resources (generation via the sub-lexical route) and one of which does not (generation via the lexical route; Besner et al., 2009). Furthermore, due to differences in orthographic depth, distractor processing may not behave the same way in every language. It is only by taking these complexities into account that accurate inferences may be drawn about the locus of interference generated by words in dual-task experiments.

Though theoretical disagreements remain over how speakers are able to select words from a large set of potential responses with ruthless efficiency, there is no longer any reason to think that semantic PWI is resolved prior to that selection process. Given this, the results of Dell'Acqua, Job et al. (2007) and Ayora et al. (2011) do not, in the end, present a challenge for major models of word production (Levelt et al., 1999), PWI (Mahon et al., 2007), or the Stroop task (Cohen et al., 1990; Roelofs, 2003).

Acknowledgments

This work was supported by the NIH under Grant Numbers R01 HD051030, NIDCD011492, and DC000041. I am very grateful to Victor Ferreira for his invaluable advice and to Mike Estrella, Angela Lee and Shivani Patel for their assistance with data collection. I would also like to thank Roberto Dell'Acqua and two anonymous reviewers for their helpful comments on an earlier version of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1860-1877. The dissertation author was the sole investigator and author of this paper.

Appendix

Critical Pictures Used in Experiment 2

Audio-visual: headphones, microphone, radio, speaker, television

Body parts: ear, eye, finger, hand, nose

Buildings: castle, church, house, lighthouse, windmill

Celestial phenomena: cloud, lightning, moon, rainbow, star

Clothes: glove, jacket, pants, skirt, sock

Farm animals: cow, donkey, horse, pig, sheep

Furniture: bed, chair, desk, stool, table

House parts: balcony, chimney, door, roof, window

Musical instruments: drum, guitar, piano, trumpet, violin

Tableware: cup, fork, glass, knife, spoon

Tools: ax, drill, hammer, saw, screwdriver

Transport: bus, car, helicopter, plane, truck

References

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: a dual-task study. *Frontiers in Psychology*, *2*, 57:1-10. doi:10.3389/fpsyg.2011.00057

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., …Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344-380. doi:10.1016/j.jml.2004.03.002

Besner, D., Reynolds, M., & O'Malley, S. (2009). When underadditivity of factor effects in the Psychological Refractory Period paradigm implies a bottleneck: Evidence from psycholinguistics. *Quarterly Journal of Experimental Psychology*, *62*, 2222-2234. doi:10.1080/17470210902747187

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. Journal of Memory and Language, 48, 468-488. doi:10.1016/S0749-596X(02)00503-X

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177-208. doi:10.1080/026432997381664

Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture-word interference paradigm: does the response set matter? *Cognition*, *75*, B51-B64. doi:10.1016/S0010-0277(99)00082-7

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 104-119. doi:10.1037/0096-1523.32.1.104

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332-361. doi:10.1037/0033-295X.97.3.332

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257-271. doi:10.3758/BF03204507

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256. doi:10.1037/0033-295X.108.1.204

Cook, A. E. (2007). *Capacity demands of word production.* (Unpublished doctoral dissertation). University of Birmingham, Birmingham, England.

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886-899. doi:10.1037/0278-7393.34.4.886

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, *10*, 111-117. doi: 10.3758/BF03196474

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345-361. doi:10.1037/0278-7393.25.2.345

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*, B77-B86. doi:10.1016/S0010-0277(01)00135-4

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283-321. doi:10.1037/0033-295X.93.3.283

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124-142. doi:10.1016/0749-596X(88)90070-8

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722. doi:10.3758/BF03196827

Dell'Acqua, R., Pesciarelli, F., Jolicœur, P., Eimer, M., & Peressotti, F. (2007). The interdependence of spatial attention and lexical access as revealed by early asymmetries in occipito-parietal ERP activity. *Psychophysiology*, *44*, 436-443. doi: 10.1111/j.1469-8986.2007.00514.x

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1058-1079. doi:10.1037/0096-1523.18.4.1058

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187-1199. doi:10.1037/0278-7393.28.6.1187

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*, 71-99. doi:10.1037/0033-

2909.123.1.71

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640-654. doi:10.1037/0096-1523.10.5.640

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464-482. doi:10.1016/j.cognition.2005.02.006

Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture-word interference. *The Quarterly Journal of Experimental Psychology*, *54A*, 371-382. doi:10.1080/713755981

La Heij, W., & van den Hof, E. (1995). Picture-word interference increases with target-set size. *Psychological Research*, *58*, 119-113. doi:10.1007/BF00571100

Lepschy, A. L., & Lepschy, G. C. (1991). *The Italian language today*. London: Routledge.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-39. doi:10.3115/992628.992631

Lien, M.-C., Allen, P. A., Ruthruff, E., Grabbe, J., McCann, R. S., & Remington, R. W. (2006). Visual word recognition without central attention: Evidence for greater automaticity with advancing age. *Psychology and Aging*, *21*, 431-447. doi:10.1037/0882-7974.21.3.431

Lien, M.-C., Ruthruff, E., Cornett, L., Goodin, Z., & Allen, P. A. (2008). On the nonautomaticity of visual word processing: Electrophysiological evidence that word processing requires central attention. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 751-773. doi:10.1037/0096-1523.34.3.751

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490. doi:10.3758/BF03210951

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology*, *36*, 684-700. doi:10.1037/h0080661

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*, 485-495. doi:10.3758/BF03198265

Lupker, S. J., & Katz, A. N. (1981). Input, decision, and response factors in picture-word interference. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 269-282. doi:10.1037/0278-7393.7.4.269

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203. doi:10.1037/0033-2909.109.2.163

Magen, H., & Cohen, A. (2002). Action-based and vision-based selection of input: Two sources of control. *Psychological Research*, *66*, 247-259. doi:10.1007/s00426-002-0099-0

Magen, H., & Cohen, A. (2010). Modularity beyond perception: Evidence from the PRP paradigm. *Joural of Experimental Psychology: Human Perception and Performance*, *36*, 395-414. doi:10.1037/a0017174

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). 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*, *33*, 503-535. doi:10.1037/0278-7393.33.3.503

McCann, R. S., Remington, R. W., & Van Selst, M. (2000). A dual-task investigation of automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1353-1370. doi:10.1037/0096-1523.26.4.1352

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, *104*, 749-791. doi:10.1037/0033-295X.104.4.749

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single bottleneck notion. *Cognitive Psychology*, *44*, 193-251. doi:10.1006/cogp.2001.0767

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

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 358-377. doi:10.1037/0096-1523.10.3.358

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica*, *142*, 332-342. doi:10.1016/j.actpsy.2013.01.006

Piai, V., Roelofs, A., & Schriefers, H. (in press). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Proctor, R. W. (1978). Sources of color-word interference in the Stroop color-naming task. *Perception & Psychophysics*, *23*, 413-419. doi:10.3758/BF03204145

Rayner, K., & Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture-word interference task. *British Journal of Psychology*, *77*, 207-222.

Reynolds, M. G., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1303-1323. doi:10.1037/0096-1523.32.6.1303

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107-142. doi:10.1016/0010-0277(92)90041-F

Roelofs, A. (2001). Set size and repetition matter: comment on Caramazza and Costa (2000). *Cognition*, *80*, 283-290. doi:10.1016/S0010-0277(01)00134-2

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88-125. doi:10.1037/0033-295X.110.1.88

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1580-1598. doi:10.1037/a0012476

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, doi: 10.3389/fpsyg.2011.00307. doi:10.3389/fpsyg.2011.00307

Ruthruff, E., Allen, P. A., Lien, M.-C., & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychonomic Bulletin & Review*, *15*, 337-343. doi:10.3758/PBR.15.2.337

Schnur, T. T., & Martin, R. (2012). Semantic picture-word interference is a post-perceptual effect. *Psychonomic Bulletin & Review*, *19*, 301-308. doi:10.3758/s13423-011-0190-x

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86-102. doi:10.1016/0749-596X(90)90011-N

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143-174.

doi:10.1348/000712603321661859

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 896-918. doi:10.1037/0278-7393.22.4.896

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662. doi:10.1037/0096-3445.121.1.15

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 3-18. doi: 10.1037/0096-1523.29.1.3

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, *16*, 987-999. doi:10.3758/PBR.16.6.987

van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*, 62-101. doi:10.1111/j.1551-6709.2011.01213.x

Welford, A. T. (1952). The "psychological refractory period" and the timing of high spead performance—A review and a theory. *British Journal of Psychology*, *43*, 2-19. doi:10.1111/j.2044-8295.1952.tb00322.x

CHAPTER 3:

DUCK, DUCK, … MALLARD: ATTENTIONAL REQUIREMENTS OF WORD PRODUCTION AFFECT WHICH WORDS PEOPLE SAY

Abstract

Different stages of word production have different attentional requirements. Such a difference in automaticity may mean that, relative to a word planned in isolation, a word produced in connected speech can be planned for longer prior to lexical selection, asymmetrically facilitating the production of dispreferred words ("sofa") relative to preferred alternatives ("couch"). This hypothesis was tested in two experiments in which subjects named pictures with multiple acceptable names under conditions that manipulated how soon subjects' attentional resources would be available to engage in lexical selection. In Experiment 1, pictures were named in single-task and dual-task contexts. In Experiment 2, pictures were named either at the beginning of a sentence or at the end of a sentence. Naming pictures in a dual-task context or at the end of a sentence delayed lexical selection relative to other conditions. The results of both experiments indicated that when subjects named pictures relatively quickly within each condition, they were more likely to use dispreferred names when lexical selection was delayed. In addition, dispreferred names were produced more often when responses were relatively slow, and the effect of response time was greater when selection was not delayed. This pattern of results is consistent with an account under which the duration of pre-lexical selection processing influences the words that speakers say. As such, they indicate that the attentional requirements of language production processes affect their outcome.

Although it feels effortless, speaking requires attention. As numerous studies have shown, talking on a cell phone while driving impairs driving performance, even if using a hands-free device (Briem & Hedman, 1995; Strayer & Johnston, 2001). The reverse is true as well, as driving impairs both linguistic production and comprehension (Becic, Dell, Bock, Garnsey, Kubose, & Kramer, 2010). It is not always the case that engaging in two tasks simultaneously impairs performance on either task, as demonstrated by (most) people's ability to walk and chew gum at the same time. Thus, the mutual interference between driving and talking must stem from their reliance on a shared pool of attentional resources, which, given the different task demands, are likely to be domain-general (i.e., central) in nature.

Like driving, language production is a complex task comprising many kinds and stages of processing. According to standard models, to produce a single word (e.g., *cat*), a speaker must first identify the semantic content to be expressed, which may be represented by conceptual features (<ANIMAL> and <FELINE>; Dell, 1986) or by a nondecompositional lexical concept (<CAT>; Levelt, Roelofs, & Meyer, 1999). Active conceptual representations spread activation to words, or lemmas, that can accurately express them, including the target lemma *cat* and semantically related words such as *dog*. Activation continues to spread to this set of words, which will be referred to henceforth as *lexical candidates*, until an active lemma is selected for production. The phonological wordform of the target lemma (/kæt/) is retrieved during lexeme selection, its constituent sounds (/k/, /æ/, /t/) are retrieved during phoneme selection, and these sounds must be ordered and sent to articulators before the word can finally be produced.

Given that speaking interferes with driving, at least some stages of language production must require central attention, but there is no reason to think that all of its stages do so. Striking evidence that multiple stimuli can simultaneously cause activation to spread throughout the language system comes from the picture-picture interference paradigm (e.g., Morsella & Miozzo, 2002; Navarrete & Costa, 2005). In this task, subjects are presented with two (often superimposed) drawings of objects and told to name the target picture (colored green) while ignoring the context picture (colored red). The basic finding is that subjects are faster to name a target picture (*bed*) when the name of the context picture is phonologically related (*bell*) than when it is unrelated (*hat*). This indicates that the name of the context picture must be activated at least to the phonological level before the subject engages in phoneme selection for the target picture, which must mean that both pictures were spreading activation from conceptual representations to lemmas, and from lemmas to phonemes, in parallel with each other. However, it does not necessarily mean that activation spreads automatically, as follow-up studies have shown that subjects are more likely to activate the phonology of a context picture name when processing resources are taxed relatively less (Mädebach, Jescheniak, Oppermann, & Schriefers, 2011) or when the context picture is visually or conceptually related to the target picture in some way (Oppermann, Jescheniak, & Görges, 2013; Oppermann, Jescheniak, & Schriefers, 2008; Oppermann, Jescheniak, Schriefers, & Görges, 2010). Similarly, results from eye-tracking studies indicate that subjects can initiate processing of to-be-named objects while fixating preceding objects (e.g., Morgan & Meyer, 2005; Schotter, Jia, Ferreira, & Rayner, 2013; see also Pollatsek, Rayner, & Collins, 1984; for a review, see Schotter, 2011), but this preview benefit appears to be attention-gated, as it may not occur when speakers do not intend to name the previewed object (Schotter, Ferreira & Rayner, 2013) and appears to be modulated by the difficulty of naming the preceding object (Malpass & Meyer, 2010; Meyer, Ouellet, & Häcker, 2008). Thus, some attentional resources may be required for activation to spread throughout the production system, but this spread can occur for multiple stimuli in parallel.

In contrast, other stages of language production seem to require enough attentional resources that they can only occur for one stimulus at a time. For example, dual-task experiments using the psychological refractory period paradigm have shown that lemma selection cannot be performed in parallel with attention-demanding processing for another task, as indicated by studies that manipulated the difficulty of lemma selection using the Stroop task (Fagot & Pashler, 1992), the cumulative semantic interference paradigm (Kleinman, 2013), and the picture-word interference paradigm (Ferreira & Pashler, 2002; Kleinman, 2013; Piai & Roelofs, 2013; Piai, Roelofs, & Schriefers, in press; Schnur & Martin, 2012; van Maanen, van Rijn, & Taatgen, 2012; but see Ayora et al., 2011; Dell'Acqua, Job, Peressotti, & Pascali, 2007 for an alternate account). This suggests that lemma selection requires central attentional resources.

If lemma selection is indeed a key attentional bottleneck in word production, it should be impossible to select multiple lemmas for production at the same time, as lemma selection for word *n* would tie up attentional resources needed to select the lemma of word *n*+1. Evidence suggests this may be true. For example, Alario, Costa, and Caramazza (2002) instructed subjects to name colored pictures using complex noun phrases (e.g., "The blue kite") while manipulating the lexical frequency of both the adjective (high-frequency: “blue”; low-frequency: “pink”) and the noun (high-frequency: “car”; low-frequency: “kite”). They found that subjects were slower to begin naming the picture when either word was low-frequency, indicating that subjects planned (at least part of) both words before starting to speak, but these frequency effects were strictly additive. Such a pattern of results could only be observed if subjects were retrieving the names of the pictures sequentially, as parallel retrieval would lead to an interaction between the frequency effects of the adjective and noun.[1] Follow-up experiments replicated this result and extended it to three-word noun phrases (e.g., “quatre voitures bleues”, or *four blue cars*): Although subjects took longer to begin speaking when either adjective was unpredictable or the noun was low-frequency, the difficulty manipulations did not interact with each other, suggesting that subjects retrieved all three names sequentially (Ayora & Alario, 2009).

The fact that selecting the lemma of a word requires central attentional resources means that it is not necessary for a speaker to get behind the wheel of a car to demonstrate the effects of divided attention on language production; simply planning multiple words is itself a dual-task process. A speaker can plan the semantic content of several words and begin to activate their lexical candidates in parallel but must select their lemmas sequentially, causing the selection of each word’s lemma to delay the selection of the next.

In keeping with previous research showing that differences in automaticity between stages of a task can alter performance in dual-task contexts compared with

---

[1] Although frequency effects are reliably attributed to the phonological level (e.g., Jescheniak & Levelt, 1994), they may also affect lemma selection; see Nozari, Kittredge, Dell, & Schwartz, 2010 for a review.

single-task contexts (Besner, Reynolds, & O’Malley, 2009; Kleinman, 2013), there is reason to think that such a difference for stages of word production may have hitherto unexpected effects on the words that speakers ultimately say. If activation accrues to lemmas until one is selected, delaying lemma selection should allow more activation to accrue prior to selection. This means that the lexical candidates for a word planned in advance (specifically, a word *n* for which planning begins before the lemma of word *n*-1 has been selected) should, on average, be more active at the time of selection than for a word planned in isolation. However, as the total amount of activation that any given lemma can accumulate is likely to be bounded, this extra activation should benefit weakly active lexical candidates more than strongly active lexical candidates.

To make this explicit, consider a scenario in which a speaker can retrieve any of several appropriate names for an object; say, *couch* or *sofa*. In a norming study (discussed below), 82% of subjects called a picture of this object “couch”, making it the dominant name, and 18% called it “sofa”, making it the non-dominant name. There is reason to believe that this disparity does not simply reflect idiosyncratic, speaker-specific preferences, but that individual subjects probabilistically choose between these acceptable names, and thus that this between-subjects measure of name agreement can be used to shed light on the selection process within each individual (Staub, Grant, Astheimer, & Cohen, 2012; this assumption will be further corroborated by data in the norming study). Given this assumption, even though the cause of subjects’ preference for *couch* is not clear, it can be represented in modeling terms as a difference in the activation levels of *couch* and *sofa* at the time of selection, with *couch* being more active.

According to the logic described above, planning the name of the couch/sofa picture while simultaneously planning other words that will be produced first should increase the amount of time that the lemmas *couch* and *sofa* receive activation prior to selection (because attentional resources are already occupied with performing lemma selection for the previous words). This prediction is shown in Figure 3.1. Because *couch* is more active, additional processing will cause it to approach its activation asymptote sooner than *sofa*, leading each extra fixed quantity of processing time to boost the activation of *sofa* more than *couch*. Thus, the difference in accessibility between *couch* and *sofa* should be reduced with additional processing. If the probability of selecting a word for production is proportional to its share of the total activation in the lexicon (Levelt et al., 1999), speakers should be more likely to select lexical candidates that were initially only weakly active (*sofa*) when planning words in advance than when planning words in isolation. (Consideration of how germane these and other assumptions are to the present account will be postponed until the General Discussion.)

The potential implications of this prediction for language production are substantial. If planning words in advance essentially buys the language production system more time to make otherwise dispreferred words accessible, the words that speaker produce may differ as a function of how far in advance they are planned. For example, if speakers can plan words at later sentence positions for longer prior to selecting their lemmas than words at earlier sentence positions, they might be more likely to use dispreferred words later in a sentence.

The logic of the present experiments is similar to that adopted by Griffin and Bock (1998) to explain why the frequency effect in picture naming (Oldfield &

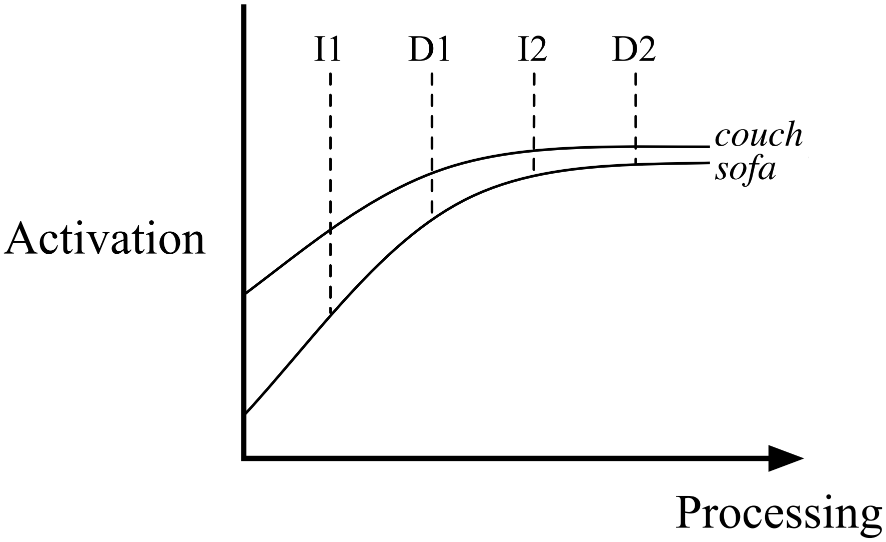


Figure 3.1. The relationship between the quantity of pre-lemma selection processing of a word (a function of both the duration of planning and the attention devoted to such planning) and the activation levels of that word's dominant (*couch*) and non-dominant (*sofa*) names. Labeled points denote when lemma selection occurs as a function of whether it is immediate (I) or delayed by previously planned speech (D), and whether responses are faster (1) or slower (2) within each condition. The probability of selecting the non-dominant name *sofa* at any given time is assumed to be its share of the total activation of all words in the lexicon.

Wingfield, 1965), in which naming latencies are faster to pictures with high-frequency names than pictures with low-frequency names, is smaller for pictures presented after strongly constraining sentence fragments (e.g., "On windy days, the boy went outside to fly his ___"). Under their account, which attributed word frequency effects to the ease of phonological wordform retrieval (Jescheniak & Levelt, 1994), the wordforms of pictures with high-frequency names have a higher resting (i.e., baseline) level of activation than the wordforms of pictures with low-frequency names, giving rise to the frequency effect. A constraining sentence fragment increases the activation level of the target lemma (*kite*). That input activation is combined with the resting activation level of the corresponding wordform (/kaɪt/) according to a logistic function, yielding an increase in the wordform activation level and thereby increasing the ease of wordform retrieval. Crucially, the nature of the function causes activation levels to approach an asymptote (see Griffin & Bock, 1998, Figure 1). As a result, the degree to which a fixed quantity of input activation affects the retrieval of a particular wordform is inversely related to that wordform's resting activation level. This explains why the frequency effect is smaller after sentence fragments: The fragments increase the input activation of both high- and low-frequency wordforms by a fixed amount, but this translates to a larger increase in the activation of low-frequency wordforms, reducing the frequency effect.

Our proposal, described above, is that an analogous process operates on lemma selection, with high- and low-frequency words replaced by dominant and non-dominant names, and with target activation increased not by sentential constraint but by a delay in lemma selection. This hypothesis was tested in two experiments in which subjects named pictures with multiple acceptable names (critical pictures) under conditions that manipulated how soon subjects' attentional resources would be available to engage in lemma selection. In Experiment 1, subjects named two pictures of objects on each trial (e.g., "tent, couch"). The critical picture, which was always named second, was presented either after subjects began naming the first picture (the no-preview condition) or at the same time as the first picture (the preview condition). In Experiment 2, two pictures were presented simultaneously, and subjects were instructed to produce a sentence that included the names of both pictures and a preposition describing their spatial relationship ("above" or "below"). A cue on each trial indicated the order in which subjects should name the pictures; the critical picture was either named first ("The couch is above the tent") or second ("The tent is above the couch").

In the preview condition in Experiment 1 and the named-second condition in Experiment 2 (henceforth referred to as the delayed selection conditions), the production system could begin planning the name of the critical picture while attentional resources were still occupied with processing the first picture. In the no-preview condition in Experiment 1 and the named-first condition in Experiment 2 (henceforth referred to as the immediate selection conditions), subjects were able to direct all (Experiment 1) or most (Experiment 2) of their attention to the critical picture as soon as it was presented. Thus, relative to the immediate selection conditions, the delayed selection conditions permitted the production system to plan the names of critical pictures for longer prior to lemma selection.

If delaying lemma selection for a picture increases the activation of all its names, but especially the non-dominant ones, three results should be observed. First, there should be a main (positive) effect of reaction time (RT; i.e., naming latency) on the probability of producing non-dominant names within each condition, as more pre-lemma selection planning should reduce the difference in activation between dominant and non-dominant names. In Figure 3.1, this is represented by greater relative activation for *sofa* at I2 (slower responses in immediate selection conditions) than at I1 (faster responses in immediate selection conditions), and at D2 (slower responses in delayed selection conditions) than at D1 (faster responses in delayed selection conditions). Second, this effect of RT on productions should be greater for the immediate selection conditions than for the delayed selection conditions, emerging as an interaction between RT and condition. This is because activation is closer to the asymptote when lemma selection is delayed than when it is not. That should reduce the effect of additional processing, as indicated by the larger difference in the relative activation of *sofa* at I2 compared to I1 than at D2 compared to D1. Third, and most crucially, there should be more productions of non-dominant names in delayed selection conditions than in immediate selection conditions (i.e., a main effect of condition), although this difference may only be apparent at relatively faster RTs in each condition. This is represented by greater activation at D1 and D2 than at I1 and I2 (although, consistent with the interaction described above, the difference is larger between D1 and I1 than between D2 and I2). Such a pattern of results, if obtained, would indicate that the attentional requirements of language production processes affect their outcome.

Norming Study

The present experiments hinge on the assumption that individual subjects probabilistically choose between acceptable picture names; e.g., that someone might call a particular picture either “couch” or “sofa”, and that the likelihood of selecting one name or the other can be manipulated by delaying lemma selection. As traditionally calculated, a picture’s name agreement is the proportion of subjects who call it by its most dominant name. In general, it may be that individual subjects are more likely to use multiple names for pictures with lower name agreement. However, that need not be the case: If, for example, 82% of subjects called a picture “couch” and 18% called it “sofa”, it may be that most subjects think the name they used is the only acceptable name (or, similarly, that they do not know the other name). If that were the case, subjects given a second opportunity to name the same picture would repeat the name they used the first time. In contrast, if name agreement indexes speaker-internal processes (as assumed), the response distribution across subjects should predict the likelihood that individual subjects repeat picture names. The norming study evaluated this hypothesis while providing data useful for the selection of materials for subsequent experiments.

Method

Subjects. Fifty members of the University of California, San Diego community participated in the norming study. Subjects received class credit for their participation. All reported being native English speakers.

Apparatus. Stimuli were presented on an iMac computer running PsyScope X (Build 51; Cohen, MacWhinney, Flatt, & Provost, 1993; http://psy.ck.sissa.it). A Shure SM10A headworn microphone connected to the button box measured voice onset latencies.

Materials. The 100 pictures were line drawings of common objects. Most pictures were selected from the International Naming Project database (IPNP; Bates et al., 2003); others were drawn in a similar style. Half (50; the high-codability group) were selected to be relatively high name-agreement so they could potentially be used as filler pictures in later experiments; the other half (the low-codability group) were selected to elicit multiple acceptable names so they could potentially be used as critical pictures in later experiments.

Procedure. Subjects were told that they would be shown pictures of everyday objects and that they should say the name that they thought best described each one, doing so as quickly as possible without making mistakes. Eight practice trials were followed by 100 experimental trials in an order determined by one of ten lists. On each trial, a fixation point (+) was presented in the center of the screen for 500 ms, after which the screen went blank for 500 ms. Then, the picture was presented in the center of the screen until the voice key registered a naming response. After the experimenter coded whether or not there was a voice key error (although naming latencies were not analyzed in the norming study), a variable inter-trial interval was sampled from a uniform distribution between 1000 and 2000 ms, inclusive.

This block of trials was followed by instructions indicating that subjects would name the same set of pictures for a second time. To reduce the likelihood that subjects would simply rely on the same name they used in the first block because it was easier to retrieve, they were instructed to use an alternate appropriate name for each picture if they could think of one, but to repeat the name they used in the first block if they could not. They were additionally instructed that if they used a different name, it should describe the same object with the same level of specificity as their first name. Eight practice trials were followed by 100 experimental trials with the same counterbalancing and trial structure as in the first block.

Results

The 100 subjects provided responses for 10,000 trials, of which 93.5% (9,351) were acceptable (4,739 from the first block and 4,612 from the second block). Trials were discarded when a subject did not provide a name for the picture or if the name provided was unintelligible (72), the name was not a noun (76), the name did not refer to the primary object in the picture (e.g., “dinner” for the picture “plate”; 276), the name was too general (e.g., “fruit” for the picture “pear” or “balance-y thing” for “scale”; 74), or the name was not an appropriate name for the picture (e.g., “tourniquet” for the picture “clamp”; 151).

Remaining responses to each picture were grouped according to the name of their central concepts. For example, the responses “bag”, “brown bag”, “grocery bag”, “lunch bag”, and “paper bag” were all classified as “bag”. When a response consisted of multiple nouns, neither of which was more central than the other and one of which represented the most dominant name (e.g., “bunny rabbit” for a picture typically called either “rabbit” or “bunny”), it was classified as a distinct response.

For each picture, the name with the highest proportion of responses in the first block was considered the dominant name, and that proportion was considered the picture’s between-subjects name agreement. For example, 49 subjects provided usable first-block responses to a picture that 40 of them called “couch”, so its between-subjects name agreement was 40/49 = 81.6%. The mean between-subjects name agreement was 99.7% for pictures in the high-codability group and 73.8% for pictures in the low-codability group. (In the second block, these numbers were 95.1% and 55.8% for the high- and low-codability picture groups, respectively, using the same dominant names as in the first block.)

To compute within-subject name agreement, 391 responses were additionally discarded where a subject had only one usable response for an individual picture. For example, if a subject called a picture “plate” in the first block and “dinner” in the second block, leading to the second response being discarded for being an inappropriate name, the first response was discarded here as well. This left 8,960 trials, or 4,480 response pairs, for which an individual subject used an appropriate name for a given picture in both blocks. The two names that each subject used to describe each picture were compared; for each picture, the proportion of subjects who used the same name twice was computed, and that proportion was considered the picture's within-subject name agreement. For example, of the 49 subjects who used two appropriate names for the "couch" picture, 25 of them used the same name in both blocks, so its within-subject name agreement was 25/49 = 51.0%. The mean within-subject name agreement was 94.9% for pictures in the high-codability group and 49.7% for pictures in the low-codability group.

The two measures of name agreement defined above are mathematically independent of each other. Given the between-subjects name agreement distribution for the couch/sofa picture, if every subject had chosen to repeat the same name in the second block, the within-subject name agreement would have been 100%. If every subject had chosen to use a different name in the second block, the within-subject name agreement would have been 0%. In either case, the between-subjects name agreement would have remained 81.6%. Thus, they represent distinct measures.

To determine whether between-subjects name agreement is predictive of within-subject name agreement, correlations were performed between the two measures for pictures in the low-codability group. (As there was 99.7% between-subjects name agreement for pictures in the high-codability group, they did not exhibit enough variation for such a comparison to be useful.) As shown in Figure 3.2, the measures were

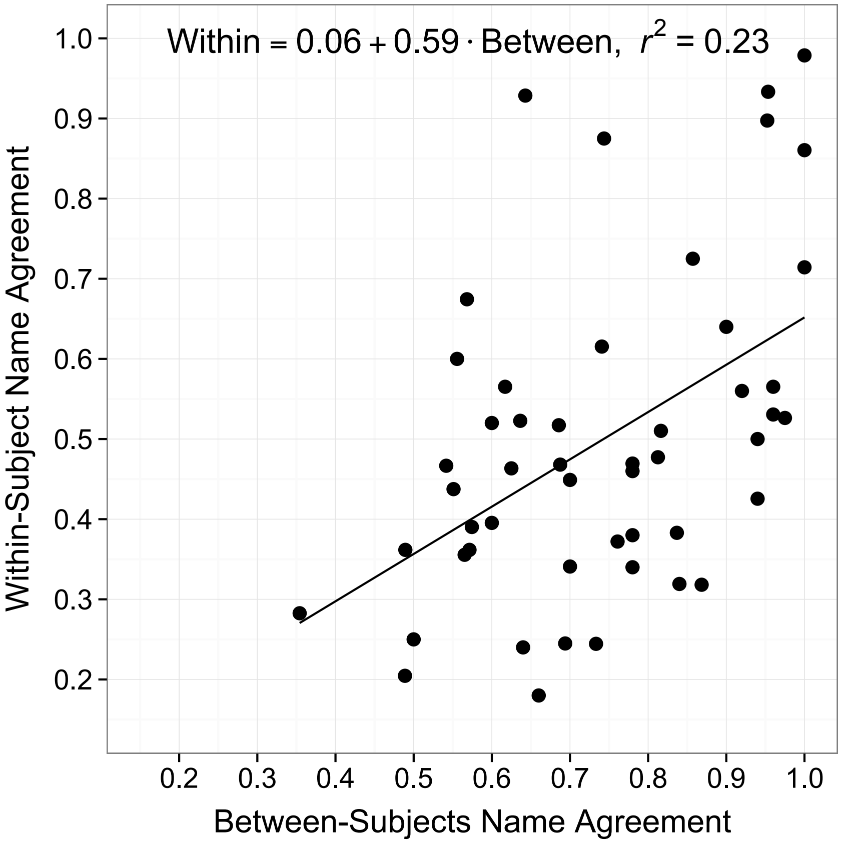


Figure 3.2. The relationship between different measures of name agreement (see text for definitions). Each point represents a picture in the low codability group in the norming study.

moderately (but significantly) correlated, $r(48) = 0.484$, $F(1, 48) = 14.6$, $p = .0004$, indicating that pictures for which different subjects use different names are also pictures for which individual subjects are willing to use multiple names.

Discussion

The purpose of the norming study was twofold. The primary goal was to identify pictures that elicited multiple names across subjects (to be used as critical pictures in subsequent experiments) and to identify other pictures that elicited only one name (to be used as filler pictures). Given the considerable variability in both measures of name agreement, this effort was successful. Second, given how important it is to the present research that individual subjects entertain multiple names for a given picture, it was necessary to determine whether subjects would be willing to do so. The fact that subjects reused names for pictures in the low-codability group only about half the time indicates that they are indeed capable of producing multiple appropriate names for a picture. Furthermore, the significant correlation between the two measures of name agreement suggests that between-subjects name agreement may index probabilistic processes that occur within individual subjects, as opposed to solely representing idiosyncratic, subject-specific preferences. In this regard, the results agree with those of Staub et al. (2012), who showed that it is possible to account for distributions of both responses and their reaction times in a speeded cloze task (Taylor, 1953) by assuming that all subjects have the same preferences for a given stimulus and that differences between subjects' responses arise from a stochastic selection process.

Of course, it is possible that explicitly instructing subjects to use a different name for the second block affected the relationship between the measures of name agreement. To assess this possibility, we turned to data from a previously published study (Ferreira, Kleinman, Kraljic, & Siu, 2012, Experiment 2) in which 48 subjects each participated in two experimental sessions, separated by one week. During the first session, subjects named a series of pictures, 24 of which depicted objects that were chosen to have multiple names. During the second session, subjects were presented with a subset of these pictures (12 each) on trials in which they had to describe them to an experimenter (priming trials; see Ferreira et al., 2012 for details). On both weeks, responses were scored according to which of two most-dominant names for the picture (as determined by a norming study) they matched; responses that matched neither name, and pictures that were only ever called by a single name, were discarded. Between-subjects and within-subject measures of name agreement were calculated as above. The measures were significantly correlated, $r(19) = 0.452$, $F(1, 19) = 4.87$, $p = .0398$, indicating that the results of the norming study presented above cannot be attributed to the instructions given to subjects.

Experiment 1

If the attentional requirements of word production facilitate the selection of non-dominant names when words are planned in advance, speakers should be more likely to produce non-dominant names like *sofa* when they plan those names at the same time as they are engaged in other attention-demanding processing (specifically, selecting the lemma of a preceding word) relative to when they plan those names in isolation. Experiment 1 tested this hypothesis by manipulating whether subjects could plan the names of critical pictures in advance or not. On each trial, subjects named two unrelated pictures, the second of which had multiple names; e.g., *tent* and *couch*. In the *no-preview* condition, *couch* was presented after subjects began to say “tent”, at which point all of their attention could be devoted to naming *couch*. In the *with-preview* condition, the two pictures were presented at the same time. Subjects’ attention was divided between the pictures until processing of the first picture was complete, at which point all of their attention could be devoted to naming *couch*.

The predictions, as laid out in the introduction and Figure 3.1, were relatively straightforward. First, longer RTs were expected to lead to a higher rate of non-dominant name production. Second, this effect of RT was expected to be modulated by preview condition, with the effect of RT on productions being greater for the no-preview condition. Finally, the with-preview condition was expected to elicit more non-dominant names than the no-preview condition, particularly at relatively faster RTs.

Method

Subjects. One hundred new members of the University of California, San Diego community participated in Experiment 1.

Apparatus. The same apparatus used in the norming study was used in Experiment 1.

Materials. Twenty-eight critical pictures with multiple names were taken from the norming study (see Appendix). They were selected largely on the basis of their within-subject name agreement, which ranged from 20.5% to 67.4% (mean: 40.9%); their between-subject name agreement ranged from 48.9% to 90.0% (mean: 68.2%). Many critical pictures had a secondary name (e.g., *sofa*) that accounted for the majority of non-dominant responses.

Each critical picture was paired with a filler picture that had relatively high name agreement, was semantically unrelated to the critical picture, and was phonologically unrelated to both the dominant and secondary names of the critical picture. Of the 28 filler pictures, 19 were taken from the norming study. Their within-subject name agreement ranged from 92.0% to 100% (mean: 97.2%) and their between-subject name agreement ranged from 98.0% to 100% (mean: 99.8%). The other nine filler pictures were taken from the IPNP database (Bates et al., 2003), according to which they had 100% name agreement across subjects. Additional filler pictures for practice trials were taken from the same source. All pictures were presented at a resolution of 200 x 200 pixels.

Design. Subjects were presented with two blocks of 14 critical trials each (28 trials total). The 28 pairs of pictures were divided into two lists of 14 pairs each. Each subject named one list in the no-preview condition and the other list in the with-preview condition; conditions were blocked within-subject. The order of conditions, the order of lists, and the assignment of lists to conditions were counterbalanced across subjects. Within each list, stimuli were presented in a random order.

Experiment 1 analyses included three variables: (a) preview condition (the critical picture was presented either after subjects began naming the filler picture or at the same time as the filler picture – the no-preview and with-preview conditions, respectively), (b) reaction time (RT; the total duration that a picture was on the screen until the subject named it), and (c) the interaction between preview condition and RT.

Procedure. Before beginning the experiment, subjects practiced picture naming in two blocks. In the first block, they named seven pairs of pictures in the no-preview condition; in the second block, they named seven pairs of pictures in the with-preview condition. No-preview trials began when a fixation point (+) appeared 150 pixels to the left of the center of the screen. After 500 ms, the fixation point was replaced by a filler picture (e.g., a picture of a tent) at the same time as another fixation point appeared 150 pixels to the right of the center of the screen. Once the voice key registered a naming response ("tent"), the filler picture disappeared and the remaining fixation point was replaced by the critical picture (couch/sofa). After a second naming response ("couch"), the critical picture disappeared. With-preview trials differed only in that the critical picture was presented at the same time as the filler picture (a second fixation point was not presented); it stayed on the screen after the first naming response until a second response was registered. After every trial, the experimenter coded whether or not there was a voice key error. A 1500 ms inter-trial interval was followed by the fixation point of the next trial.

Trials in the actual experiment were structured identically to the practice. Before each block, subjects were informed whether they would be naming pictures in the no-preview or with-preview condition. Each block was preceded by two practice trials containing only filler pictures.

Because it was difficult for the voice key to register two accurate naming latencies to pictures named in close succession, especially in the with-preview condition, a variable voice key delay was adapted to each subject's speaking rate in the practice and throughout the experiment. If the delay was too short, the subject would not have enough time to finish the first naming response in a trial before the microphone attempted to detect a second naming response (and falsely triggered to the first one for the second time). If the delay was too long, the microphone would not pick up the second naming response at its onset. After some trial and error, the default delay was set to 400 ms for no-preview trials and 200 ms for with-preview trials for all subjects. This meant that at the start of the practice trials, the microphone would not detect any microphone input in a no-preview trial for 400 ms, or in a with-preview trial for 200 ms, after registering the first naming response. When the experimenter deemed it necessary, these delays were adapted independently to each subject in increments of 25 ms to minimize voice key errors.

Results

The 100 subjects each named 28 critical pictures, providing responses for 2,800 critical trials. However, not all of these pictures were analyzed. Because it was important to use only those critical pictures for which the subject population likely had a single dominant name, the stability of each critical picture's dominant name was assessed. Name agreement data was combined from four sources, each of which contributed data from 50 subjects, to determine the names subjects use in relatively normal picture-naming contexts: the IPNP database (Bates et al., 2003), in which subjects named pictures in isolation; the first block of the norming study; the no-preview condition of Experiment 1; and the named-first condition of Experiment 2 (to be discussed). All of the critical pictures used in Experiment 1 were also presented in both the norming study and in Experiment 2, and all but four were named as objects in the IPNP. A critical picture was analyzed in Experiment 1 if, and only if, it met two criteria: (1) the dominant name was the same across at least 75% of the studies in which the picture was presented, and (2) the dominant name across experiments matched the dominant name for the no-preview condition in Experiment 1. Three pictures (*glass*, *ship*, *surgeon*) were discarded from both experiments for failing to meet the first criterion, and one additional picture (*coat*) was discarded from Experiment 1 for failing to meet the second criterion (see Appendix).

Of the 2,400 remaining critical trials, 1,927 (80.3%; 997 no-preview trials and 930 with-preview trials) were analyzed. Trials were discarded if there was a voice key error when subjects named the first picture (31 no-preview trials and 47 with-preview trials) or the second picture (103 no-preview trials and 168 with-preview trials), due to experimenter error (2), or if either the RT to the first picture or the inter-response interval (IRI; the time between responses) was less than 150 ms (1) or greater than 2000 ms (54). Remaining trials were discarded when a subject did not provide a name for the critical picture or if the name provided was unintelligible (10), the name was not a noun (2), the name did not refer to the primary object in the picture (45), the name was too general (6), or the name was not an appropriate name for the picture (32). Remaining responses to each picture were grouped in the same way as in the norming study and their RTs (measured from picture onset to naming response) were log-transformed to better approximate a normal distribution.

Two binary, mixed logit analyses (Jaeger, 2008) were conducted to determine whether the odds of using a non-dominant name were influenced by the factors of interest. The dependent variable for both analyses was the response given on each trial: Non-dominant responses were counted as "successes" and dominant responses were counted as "failures". The first analysis included the main effect of RT and its interaction with preview condition. The second analysis included only the main effect of preview condition (which was not included in the first analysis).[2] Subjects and pictures were

---

[2] All three factors were not combined into a single analysis because large differences in RTs between conditions led to substantial collinearity between the main effects of RT and condition, making credit assignment impossible. For example, when RTs were used (which were longer for the with-preview condition), a significantly

treated as random factors, and a maximal random effects structure was used (Barr, Levy, Scheepers, & Tily, 2013) in which all fixed effects were allowed to vary by both random factors. All predictors were centered and some were linearly scaled to facilitate model convergence; all reported data show unscaled estimates and standard errors.

Subject means are shown in Figure 3.3, and the results of the first analysis are reported in Table 3.1 (Analysis 1). Subjects produced non-dominant names significantly more often when RTs were longer, as indicated by a significant main effect of RT. However, the effect of RT was modulated by a significant interaction with preview condition. This indicates that the effect of RT was larger for the no-preview condition, or, equivalently, that the effect of preview condition (specifically, the number of non-dominant names produced in the with-preview condition as compared with the no-preview condition) was larger when RTs were relatively fast.

To determine whether the effect of RT was significant for both no-preview and with-preview conditions, separate models were fit to the data from each condition. Each model contained a main effect of RT, random intercepts for subjects and pictures, and a maximal random effects structure. The effect of RT was significant for both the no-preview condition, $\beta = 1.78$, $SE = 0.44$, $z = 4.07$, $p < .001$, and the with-preview condition, $\beta = 0.84$, $SE = 0.42$, $z = 2.01$, $p = .045$, confirming that slower responses increased the odds of producing non-dominant names in both conditions.

The results of the second analysis are reported in Table 3.1 (Analysis 2). Averaged by subject, 29.7% of subject responses in the with-preview condition were

---

different effect of preview condition was obtained than when IRIs were used (which were longer for the no-preview condition) even though the pattern of responses did not change. Conducting separate analyses eliminated this collinearity.

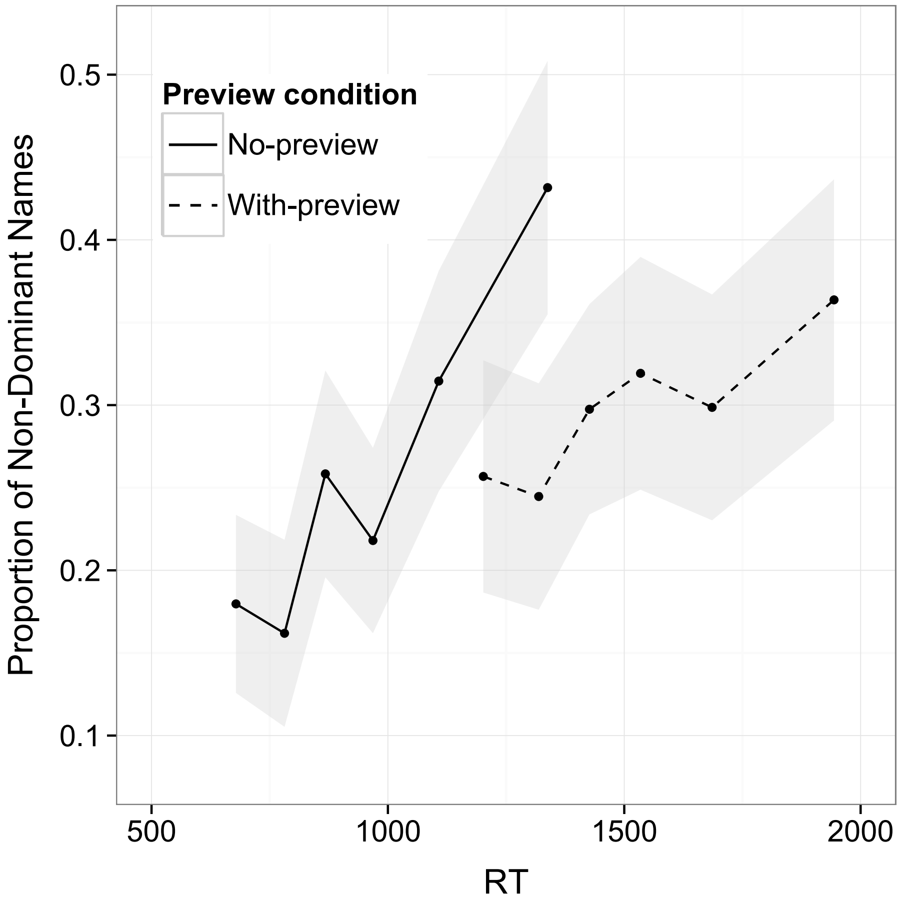


Figure 3.3. Experiment 1 probabilities of non-dominant name production shown as a function of preview condition and back-transformed reaction times (RTs). Each point represents a within-subject RT quantile. (Analyses were computed using continuous RTs; data were binned into quantiles for illustrative purposes only.) Ribbons denote within-subject 95% confidence intervals.

non-dominant names, compared with 26.1% of responses in the no-preview condition. Although this effect trends in the predicted direction, it was not significant. However, as the effect of preview condition was predicted to be largest for relatively faster RTs (see Figure 3.1) and given the interaction between RT and preview condition, this analysis was repeated for the fastest two-thirds of responses for each subject in each condition,

Table 3.1. Experiment 1 results derived from two mixed logit models.

Fixed effect	*β*	*SE*	*z*	*p*
Analysis 1				
Intercept	-0.94	0.12	-7.49	< .001
RT	0.83	0.19	4.47	< .001
RT * Preview condition	-1.11	0.52	-2.14	.032
Analysis 2				
Intercept	-1.02	0.14	-7.50	< .001
Preview condition	0.13	0.12	1.06	.290

Note. RT = log-transformed reaction times. "Success" = production of a non-dominant name. All fixed effects were allowed to vary by both random factors (subjects, pictures).

and again for the fastest third of responses in each condition. For the fastest two-thirds of responses, 28.4% of subject responses in the with-preview condition were non-dominant names, compared with 20.0% of responses in the no-preview condition; this main effect was significant, $\beta = 0.42$, $SE = 0.17$, $z = 2.43$, $p = .015$. For the fastest third of responses, the rates were 24.2% and 16.8% for the with- and no-preview conditions, respectively; this main effect was marginally significant, $\beta = 0.43$, $SE = 0.24$, $z = 1.79$, $p = .073$. As the effect size was the same for both subset analyses, the difference in significance is attributable to the larger error term in the analysis of the fastest third of responses. Thus, preview did significantly increase the production of non-dominant names when RTs were relatively fast.

Discussion

Experiment 1 showed that when subjects named pictures relatively quickly within each condition, they were more likely to use non-dominant names when they could preview the critical picture, but this preview effect disappeared for slower responses. This is consistent with the hypothesis that permitting speakers to begin planning a word while delaying lemma selection by occupying attentional resources with another task (here, naming the filler picture) boosts the activation of lexical candidates – especially less active candidates – prior to selection. Furthermore, subjects produced more non-dominant names in each condition when RTs were slower. This suggests that more pre-lemma selection planning time facilitates the production of non-dominant names regardless of whether the extra time is attributable to an attentional bottleneck or not.

At the same time, the relationship between RT and responses cannot, by itself, be interpreted as evidence for the present account. It could simply be the case that when subjects decide to produce non-dominant names, they take longer to retrieve; that is, the name determines the RT, not the other way around. However, an account under which RT does not affect responses would predict that preview condition should also have no effect on responses. As the results of Experiment 1 show, this prediction is incorrect. Thus, it is not possible to account for the present data without assuming that the duration of pre-lemma selection planning affects responses.[3]

---

[3] Nothing in the present account forbids the possibility that the selection of a non-dominant name takes longer than the selection of a dominant name, possibly because the duration of selection may be inversely related to the activation of the selected lemma (see, e.g., the booster mechanism described by Oppenheim, Dell, & Schwartz, 2010). Our central claim is simply that the causal arrow goes (not necessarily exclusively) in the

Subjects clearly processed the second picture in the with-preview condition before naming the first picture, as evidenced both by the effect of preview condition and by the fact that IRIs in the with-preview condition (603 ms) were much faster than RTs in the no-preview condition (977 ms). Was the quantity of this processing fixed, or did subjects process the second picture continuously throughout the trial? The answer to this question may shed light on the automaticity of picture processing. To address it, RTs to pictures in the with-preview condition were decomposed into two components: the RT to the first picture (RT1) and the IRI. If subjects processed the second picture relatively automatically, they should have produced more non-dominant names when RT1 was longer because the extra preview time would have continually increased the activation of lexical candidates. If processing the second picture required attentional resources, responses may be unrelated to RT1 if subjects chose not to continuously allocate attention to the second picture before naming the first. In either case, given that subjects could allocate their full attention to the second picture after naming the first, more non-dominant names should have been produced when IRI was longer.

An additional analysis was conducted on the data from the with-preview condition to determine the relative contributions of RT1 and IRI to the odds of producing a non-dominant name. The model contained main effects of RT1 and IRI (both log-transformed, though results were identical without the transformation), random intercepts for subjects and pictures, and a maximal random effects structure. The effect of RT1 was not significant, $\beta = -0.40$, $SE = 0.42$, $z = -0.96$, $p = 0.337$, but the effect of IRI was

---

opposite direction, with longer pre-lemma selection planning increasing the likelihood of producing non-dominant names.

significant, $\beta = 0.70$, $SE = 0.23$, $z = 3.00$, $p = 0.003$. Thus, subjects processed the second picture for a relatively fixed amount of time prior to naming the first picture. This pattern was not due to collinearity between the effects in the model, $r = -0.092$, indicating that there was not a trade-off between RT1 and IRI, and thus that subjects were not simply waiting until they prepared the name of the second picture before beginning to name the first. This suggests that activating lexical candidates requires attention, in keeping with the conclusions of Oppermann and colleagues (e.g. Oppermann et al., 2013). However, it must still require less attention than lemma selection; if it did not, subjects could have begun to select a lexical candidate for the second picture in parallel with its processing, and thus no effect of preview condition on responses would have been observed.

Experiment 2

Experiment 1 contrasted the outcome of picture naming in contexts that either did or did not permit advance planning. However, as isolated noun production is not characteristic of naturalistic speech, in which words are produced in sentences, a somewhat more ecologically valid contrast would be between noun production when less vs. more advance planning is permitted. If planning semantic content and activating lexical candidates requires fewer attentional resources than lemma selection, speakers should be able to plan a word produced at the end of a sentence for a longer time than a word produced at the beginning of a sentence. To test this hypothesis, subjects in Experiment 2 named two pictures (one critical), which were presented at the same time, within a sentence frame. The critical picture was either named first (“The couch is above the tent”) or second (“The tent is above the couch”) on each trial.

If attentional requirements facilitate non-dominant names for words planned farther in advance, the named-first and named-second conditions should mirror the no-preview and with-preview conditions from Experiment 1, respectively. The predictions for Experiment 2 are the same as those for Experiment 1.

Method

Subjects. One hundred new members of the University of California, San Diego community participated in Experiment 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.

Materials. The set of 40 critical pictures consisted of all 28 critical pictures from Experiment 1, 7 pictures from the low-codability picture group in the norming study that were not presented in Experiment 1, and 5 pictures from the IPNP database (Bates et al., 2003) that were not presented in either the norming study or Experiment 1 (see Appendix). The new pictures from the IPNP database had moderately high between-subjects name agreement (between 60.0% and 90.0%) and were selected primarily for the fact that they were generally called by two or three names that accounted for all subject responses.

As in Experiment 1, each critical picture was paired with a filler picture that had high name agreement, was semantically unrelated to the critical picture, and was phonologically unrelated to both the dominant and secondary names of the critical picture. The set of 40 filler pictures consisted of 25 pictures that were presented in Experiment 1, 7 pictures from the high-codability picture group in the norming study that were not presented in Experiment 1, and 8 pictures from the IPNP database (Bates et al., 2003) that were not presented in either the norming study or Experiment 1. The new pictures from the IPNP database all had 100% between-subjects name agreement. In addition to the 40 critical and 40 filler pictures selected for critical trials, an equal number of pictures (80) were selected for filler trials from the IPNP database (Bates et al., 2003), according to which all pictures had high between-subjects name agreement (between 86.7% and 100%). Additional filler pictures for practice trials were taken from the same source. All pictures were presented at a resolution of 200 x 200 pixels.

Design. Subjects were presented with two blocks of 40 trials each (80 trials total) in an order determined by one of four stimulus lists. On each trial, subjects were presented with two pictures to name; one picture was surrounded by a red square indicating that the subject should name that picture first. There was one independently manipulated factor of interest: order of mention; i.e., whether the critical picture on each trial was cued to be named first or named second.

On trials containing a critical picture, the cue always indicated that subjects should name the top picture first; on filler trials, the cue always indicated that subjects should name the bottom picture first. Across the four lists, each critical picture was presented equally often in each combination of order of mention (named-first vs. named-second) and block (first block vs. second block).

Each block contained an equal number of critical and filler trials. The sequence of cue locations was controlled so that the cue appeared in the same location on no more than three consecutive trials. The first trial in each block was a filler trial.

Experiment 2 analyses included three variables: (a) order of mention (the critical picture was named either first or second), (b) RT (the total duration from picture onset until the subject began speaking), and (c) the interaction between order of mention and RT.

Procedure. Subjects practiced 16 trials before proceeding to the 80 experimental trials. Practice trials and experimental trials had the same structure. Each trial began when a fixation point (+) appeared in the center of the screen for 1000 ms, after which the screen went blank for 500 ms. Then, two pictures appeared at the same time. One picture was centered 150 pixels above the center of the screen; the other picture was centered 150 pixels below the center of the screen. One of the pictures was bordered on all sides by a thin red square (10 pixels wide), which indicated to the subject that it should be named first. Because the square surrounded the top picture on all critical trials, subjects should always have produced a sentence of the form “The X is above the Y” on critical trials. For example, if the *tent* picture was on top (and thus surrounded by the red square), subjects should have said, “The tent is above the couch”; if the *couch* picture was on top, subjects should have said, “The couch is above the tent.”

Once the voice key registered a naming response, both pictures disappeared. At the same time, an empty rectangle with a black border (representing a progress bar) appeared at a point centered 258 pixels below the bottom picture. Over the next 1500 ms, it filled up from left to right in 10% increments every 150 ms. The filled bar remained on the screen for 750 ms, then disappeared. On each trial, the experimenter coded whether or not there was a voice key error; due to the inconsistency of the microphone in picking up the word “the”, trials in which the pictures disappeared between the onset of speech and the onset of the name of the first picture were considered acceptable. A 1500 ms inter-trial interval was followed by the fixation point of the next trial.

As in Ferreira (1996), subjects were instructed to prepare a complete, fluent utterance before beginning to speak. They were warned that because the pictures would disappear when the voice key registered a response, they would not have an opportunity to plan the name of the second picture after doing so. Furthermore, they were encouraged to complete their utterance by the time the progress bar filled up (although data were not excluded on the basis of whether or not they did so).

Subjects were given the opportunity to take a short break after the first block of 40 experimental trials.

Results

The 100 subjects each named 40 critical pictures, providing responses for 4,000 critical trials. However, as in Experiment 1, not all of these pictures were analyzed. Based on the criteria described in the Experiment 1 Results section, four pictures (*barbecue*, *glass*, *ship*, *surgeon*) were discarded for not having a consistent, dominant name across studies, and two additional pictures (*needle*, *picture*) were discarded because the dominant name in Experiment 2 did not match the name that was consistently dominant across studies (see Appendix).

Of the 3,400 remaining trials, 3,120 (91.8%; 1,563 named-first trials and 1,557 named-second trials) were analyzed. Trials were discarded if there was a voice key error (43 named-first trials and 47 named-second trials) or if the RT was greater than 10000 ms (8). This liberal RT cutoff was used because subjects were instructed to prepare the entire sentence before beginning to speak; however, the statistical significances of all reported results for Experiment 2 were identical when a 5000 ms cutoff was used instead, which would ultimately have resulted in another 71 responses being discarded. Remaining trials were discarded when a subject named the pictures in an incorrect order (3), used an incorrect preposition (30), corrected their utterance by changing it mid-sentence (20), did not provide a name for both pictures or if the names provided were unintelligible (10), the name for the critical picture did not refer to the primary object in the picture (43), the name was too general (28), or the name was not an appropriate name for the picture (53). (Some trials violated multiple criteria.) Remaining responses to each picture were grouped in the same way as in the norming study and Experiment 1, and their RTs were log-transformed to better approximate a normal distribution.

Analyses in Experiment 2 were structurally identical to those reported in Experiment 1, except that the factor of interest was order of mention instead of preview condition. Thus, the first analysis included the main effect of RT and its interaction with order of mention. The second analysis included only the main effect of order of mention.

Subject means are shown in Figure 3.4, and the results of the first analysis are reported in Table 3.2 (Analysis 1). Subjects produced non-dominant names significantly more often when RTs were longer, as indicated by a significant main effect of RT. However, the effect of RT was modulated by a significant interaction with preview

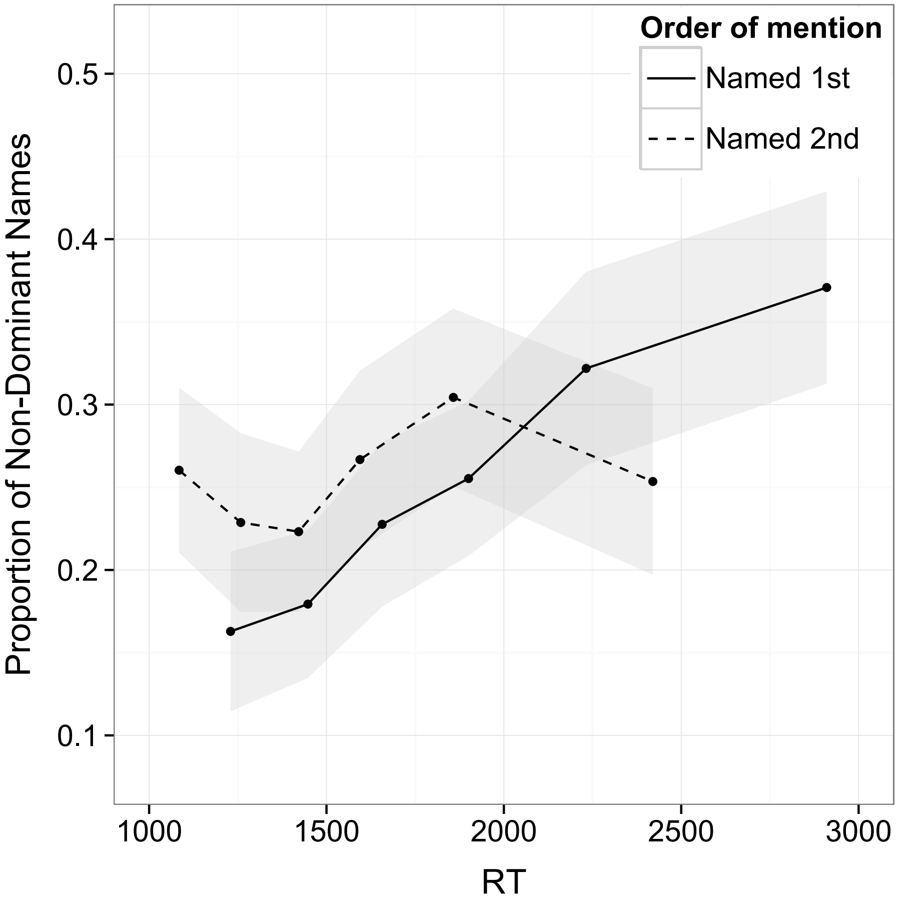


Figure 3.4. Experiment 2 probabilities of non-dominant name production shown as a function of order of mention and back-transformed reaction times (RTs). Each point represents a within-subject RT quantile. (Analyses were computed using continuous RTs; data were binned into quantiles for illustrative purposes only.) Ribbons denote within-subject 95% confidence intervals.

condition. This indicates that the effect of RT was larger for the named-first condition, or, equivalently, that the effect of order of mention (specifically, the number of non-dominant names produced in the named-second condition as compared with the named-first condition) was larger when RTs were relatively fast.

Table 3.2. Experiment 2 results derived from two mixed logit models.

Fixed effect	*β*	*SE*	*z*	*p*
Analysis 1				
Intercept	-1.27	0.16	-8.12	< .001
RT	0.41	0.15	2.72	< .001
RT * Order of mention	-0.66	0.22	-2.97	.003
Analysis 2				
Intercept	-1.24	0.15	-8.44	< .001
Order of mention	-0.07	0.10	-0.72	.471

Note. RT = log-transformed reaction times. "Success" = production of a non-dominant name. All fixed effects were allowed to vary by both random factors (subjects, pictures).

To determine whether the effect of RT was significant for both named-first and named-second conditions, separate models were fit to the data from each condition; the structure of these models was the same as in Experiment 1. The effect of RT was significant for the named-first condition, $\beta = 0.74$, $SE = 0.19$, $z = 3.78$, $p < .001$, but not for the named-second condition, $\beta = 0.13$, $SE = 0.17$, $z = 0.73$, $p = .468$. Thus, slower responses increased the odds of producing non-dominant names only in the named-first condition.

The results of the second analysis are reported in Table 3.2 (Analysis 2). Averaged by subject, 25.6% of subject responses in the named-second condition were non-dominant names, compared with 25.3% in the named-first condition, a non-significant difference. As in Experiment 1, this analysis was repeated for the fastest two-thirds of the responses for each subject in each condition, and again for the fastest third of responses in each condition. For the fastest two-thirds of responses, 24.4% of subject responses in the named-second condition were non-dominant names, compared with 20.2% of responses in the named-first condition; this main effect was not significant, $\beta = 0.13$, $SE = 0.15$, $z = 0.82$, $p = .410$. For the fastest third of responses, the rates were 24.6% and 17.1% for the named-second and named-first conditions, respectively; this main effect was significant, $\beta = 0.41$, $SE = 0.19$, $z = 2.11$, $p = .034$. Thus, order of mention did significantly increase the production of non-dominant names for the fastest RTs.

Discussion

Experiment 2 showed that when subjects named pictures relatively quickly within each condition, they were more likely to use non-dominant names when the critical picture was named second, but this effect disappeared for slower responses (and, as Figure 3.4 shows, even reversed; statistical support for this reversal is described below). In this respect, although the effect of order of mention was only significant for the fastest third of responses in each condition (as opposed to the fastest two-thirds for Experiment 1), the results of the two experiments agreed. They diverged, however, with respect to the relationship between RT and responses: Although subjects in Experiment 2 produced more non-dominant names when RTs were slower, this effect was driven entirely by responses in the named-first condition.

Why did RTs not affect productions (or vice-versa) in the named-second condition? The most likely possibility is that despite the instructions they were given to prepare a complete, fluent utterance before speaking, subjects only planned the name of the first picture before speaking. This agrees with Griffin (2001), who showed that although pictures with lower codability are named more slowly (Lachman, 1973; Lachman & Lachman 1980), the codability of the second and third objects named within a predictable sentence frame (e.g., B and C in the sentence "The A and the B are above the C") did not affect when subjects began speaking even when their sentences were fluent. Furthermore, in Experiment 2, (back-transformed) RTs were 270 ms faster in the named-second condition (1551 ms) than in the named-first condition (1821 ms). This difference in RTs is consistent with the possibility that the codability of the first-named picture (which was a low-codability critical picture in the named-first condition and a high-codability filler picture in the named-second condition) affected RTs more than the codability of the second-named picture because of differential subject attention to the two pictures prior to speech onset (see also Smith & Wheeldon, 1999).

If subjects did not retrieve the name of the second picture prior to speaking, it may be that a measure of RT that encompasses total planning time would be a better predictor of naming responses. To determine whether this was the case, RTs were measured by hand from picture presentation until the onset of the second picture name; for example, for the sentence "The tent is above the couch", from the moment the *couch* picture was presented until the subject said "couch." In the named-second condition, these RTs were decomposed into two components: the RT to speech onset (RT1) and the IRI (how much time elapsed between speech onset and when the subject said "couch"). An analysis that was structurally identical to the one reported in the Experiment 1 Discussion section was conducted to determine the relative contributions of each RT component (both log-transformed, though results were identical without the transformation) to the odds of producing a non-dominant name. Neither the effect of RT1, $\beta = 0.10$, $SE = 0.18$, $z = 0.55$, $p = 0.580$, nor the effect of IRI was significant, $\beta = 0.29$, $SE = 0.27$, $z = 1.05$, $p = 0.293$.

The lack of an effect of RT1 on responses is consistent with the results of Experiment 1. Combined with the fact that more non-dominant names were produced among the fastest responses in the named-second condition than in the named-first condition, this supports the idea that subjects processed the second picture for a relatively fixed amount of time prior to speech onset. Furthermore, this explains why, for the slowest third of responses in each condition, subjects produced significantly more non-dominant names in the named-first condition (34.5%) than in the named-second condition (28.3%), $\beta = -0.44$, $SE = 0.18$, $z = -2.46$, $p = 0.014$. As RTs solely reflected the planning of the first picture name, a long RT indicated that a subject processed the first picture (but not the second picture) for an extended period of time, shifting the time of selection to the right on the curves shown in Figure 3.1.

The lack of an effect of IRI on responses is inconsistent with the results of Experiment 1. The difference may be attributed to the fact that more words intervened between the two picture names in Experiment 2 ("is above the") than in Experiment 1 (no intervening words), providing a buffer that simultaneously masked any effect of planning time on word selection while absorbing any potential costs associated with selecting a non-dominant word (Griffin, 2001, 2003).

General Discussion

In two experiments, subjects named pictures with multiple acceptable names under conditions that manipulated how soon subjects' attentional resources were available to engage in lemma selection. The experiments yielded similar results: When lemma selection was delayed relative to the initiation of planning – either because pictures were previewed during the planning of upcoming speech in Experiment 1 or because they were named later in a sentence in Experiment 2 – subjects produced more non-dominant names than when they were free to engage in lemma selection immediately, provided that their responses were relatively fast. This difference between conditions was reduced as responses slowed.

The similarity of the experimental results is graphically depicted in Figure 3.5, which shows how the log odds of producing a non-dominant name in delayed selection conditions as compared with immediate selection conditions changed as responses got slower for each experiment. For Experiment 1, the proportion of non-dominant names produced for each with-preview RT quantile (represented by a point in Figure 3.3) was divided by the proportion of non-dominant names produced for each corresponding no-preview quantile. For Experiment 2, the proportion of non-dominant names produced for each named-second RT quantile (represented by a point in Figure 3.4) was divided by the proportion of non-dominant names produced for each corresponding named-first RT quantile. As Y-values represent the natural logarithms of these quotients, positive Y-values represent greater production of non-dominant names in delayed selection conditions than in immediate selection conditions. X-values represent equally-spaced RT percentiles corresponding to each quantile. Effects of RT percentile on differential

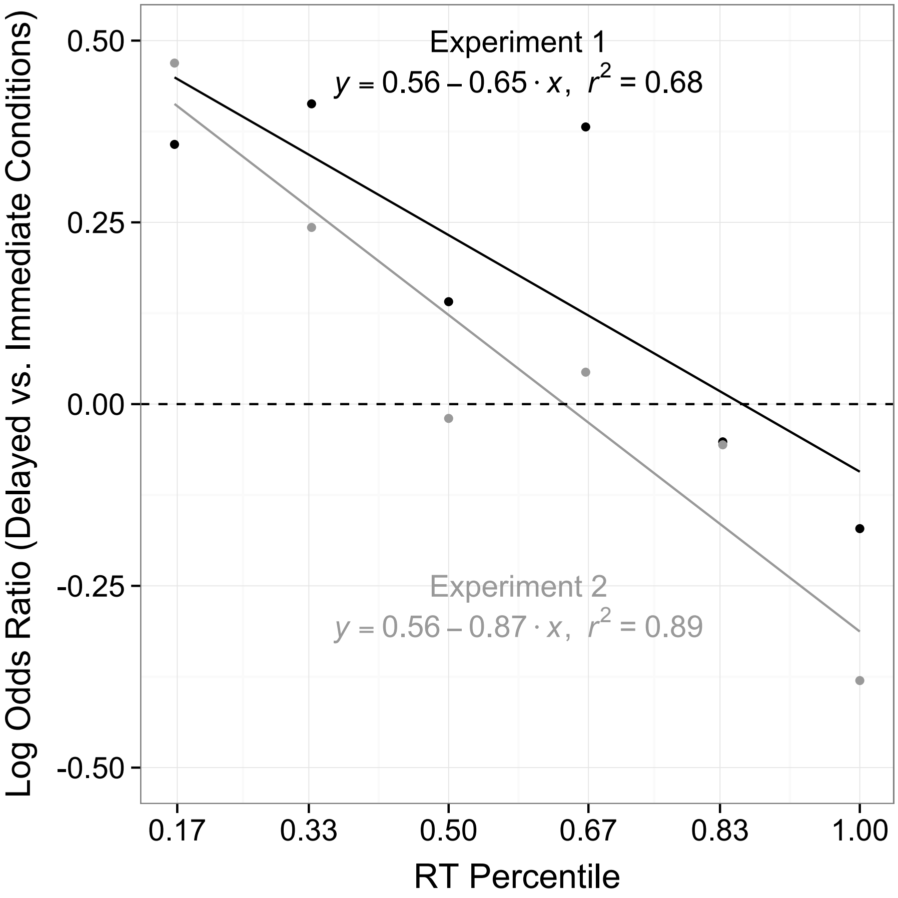


Figure 3.5. Relationships between reaction time (RT) percentile and the log odds of producing a non-dominant name in the delayed selection condition relative to the immediate selection condition for each experiment (see text for details).

production of non-dominant names between conditions are denoted by best-fit lines and their equations.

The positive intercepts of the best-fit lines for each experiment indicate that at faster RTs within each condition, subjects produced more non-dominant names for delayed selection conditions than immediate selection conditions. The negative slopes represent the interaction between condition and RT, such that the effect of condition was reduced (and, in Experiment 2, significantly reversed) as RT increased. The best-fit line for Experiment 2 has a steeper slope than for Experiment 1, reflecting the fact that increasing RT had a larger effect on the relative likelihood of producing non-dominant names between conditions in Experiment 2. However, the two lines have the same intercept, 0.56, which corresponds to 75% more non-dominant names produced in the delayed selection condition than in the immediate selection condition ($e^{0.56} = 1.75$). This suggests that the effect of delaying lemma selection was the same for the two experiments, and that discrepancies between experiments are attributable to differences in the relative amount of attention devoted to the two pictures.

The results are consistent with an account under which advance planning during word production facilitates the selection of contextually appropriate but weakly active lexical candidates. Because the production system can activate lexical candidates for multiple words in parallel but cannot engage in lemma selection while central attentional resources are otherwise occupied, lexical candidates for words planned in advance can continue to accrue activation while lemma selection is delayed. Due to the bounded nature of activation, this accrual process benefits weakly active candidates more than strongly active candidates. When attentional resources are finally freed up, allowing the production system to engage in lemma selection, the probability of selecting initially weakly active candidates is greater. As speakers often plan elements of their speech in advance, it is likely that the attentional requirements of language production processes affect their outcome not only in picture-naming tasks, but in naturalistic speech as well.

Necessary (and unnecessary) assumptions

As described above, the present account makes certain assumptions about the shape of activation curves, the nature of lemma selection, and the identity of the representations affected by the manipulations in Experiments 1 and 2. However, while these assumptions are sufficient to explain the observed pattern of results, they are not all necessary. Which assumptions are actually germane to the present account, and which are more flexible?

As shown in Figure 3.1, the activation curves for dominant and non-dominant names were assumed to have different starting points; i.e., differences in appropriateness and accessibility were represented as differences in baseline activation. In recognition of the fact that subjects will most likely never produce dominant and non-dominant names equally often, no matter how long they plan a picture name prior to lemma selection, these words were assumed to have different asymptotes, with the maximum possible activation level of the dominant name exceeding the maximum possible activation level of the non-dominant name. However, given the assumptions that each word is selected with a probability proportional to its share of the total activation (Levelt et al., 1999) and that activation levels approach an asymptote, the present results would be equally well accommodated if the two names had the same resting level of activation as long as the dominant name accumulated activation faster than the non-dominant name or if they had different baseline levels of activation along a single curve (e.g., Griffin & Bock, 1998). Even the asymptote assumption could be relaxed if the words had different baseline levels of activation and accumulated activation at the same rate. In all of these scenarios, the relative activation of the non-dominant name, and thus its probability of being selected, increases with additional processing time.

The different scenarios described above assume that lemma selection is a competitive process – i.e., that it takes longer to select a target word when a competitor is more active than when it is less active – but this assumption is itself not universally shared (see, e.g., Dell, 1986; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Oppenheim et al., 2010). Neither is it essential to explain the results of Experiments 1 and 2. If a noise parameter were introduced to make the relationship between quantity of processing and the activation of a lexical candidate non-deterministic, the data could be explained equally well by a horse-race model (cf. Staub et al., 2012). Under one possible implementation of such an account, the dominant name would have a higher resting level of activation and all names would noisily accumulate activation at the same rate. The process that checks whether any words have surpassed the activation threshold would be suspended until enough attentional resources were available, at which point it would choose randomly from all lexical candidates that had crossed it. As non-dominant names would generally take longer to cross the threshold than dominant names, delaying the checking process would increase the likelihood of selecting a non-dominant name.

It may not even be necessary to assume that the representations affected by the manipulations in Experiments 1 and 2 are word lemmas. Instead, it may be that attention affects production by delaying the selection of lexical concepts; i.e., word-specific semantic representations. We have used the term “appropriate picture names” (after Levelt et al., 1999) rather than “synonyms” to describe dominant and non-dominant names because the most common responses to each picture did not always share a common meaning. For example, a picture that most subjects called *road* was called *street* by other subjects and *highway* by still others. As these terms have distinct semantic representations, the key decision regarding what to call the picture may have occurred at the semantic level, where subjects chose between the lexical concepts <ROAD>, <STREET>, and <HIGHWAY> (cf. Staub et al., 2012). For the rest of the experimental logic to hold, it would have to be the selection of a lexical concept (instead of a word lemma) that was delayed by a simultaneously planned word, and activation would have to spread relatively more automatically through the production system down to at least the level of lexical concepts rather than necessarily spreading all the way to word lemmas.

Indeed, if activation could spread to the level of lexical concepts relatively automatically but *not* to word lemmas, the present studies would perhaps be more easily reconciled with the results of dual-task experiments showing that semantic interference is unaffected when a concurrently presented stimulus must be responded to first (Fagot & Pashler, 1992; Kleinman, 2013; Piai & Roelofs, 2013; Piai et al., in press; Schnur & Martin, 2012; van Maanen et al., 2012). For example, if a subject in a dual-task Stroop experiment presented with the word "red" in blue ink could activate the target word lemma *blue* from the lexical concept <BLUE> while performing a concurrent tone task, interference could potentially be reduced (or increased, depending on how much activation the interfering word lemma *red* received from the lexical concept <RED>). That such modulation did not occur (Fagot & Pashler, 1992) means it may be simpler (albeit not strictly necessary) to assume that neither the *blue* nor *red* lemmas received activation from their respective lexical concepts during the tone task, which would suggest that activation spread no farther than lexical concepts.

Taken together, the flexibility of these assumptions means that the results of Experiments 1 and 2 do not depend on individual properties of activation curves (starting points, accumulation rates, the existence of an asymptote), on the nature of lemma selection (competitive vs. non-competitive), or even on a lemma selection locus for the reported effects. Ultimately, there are only two necessary assumptions. First, the shape of the activation curves and the selection rule must combine in such a way that a difference in the probability of selection between more- and less-accessible representations is reduced with additional processing time. Second, regardless of the level of representation at which the selection process in question occurs, that selection process must require more attentional resources than the spread of activation within the production system (at least to that level).

Implications for language production and beyond

If the attentional demands of language production affect the relative ease of retrieving dominant vs. non-dominant names for a particular object, they might also affect the relative ease with which bilinguals can retrieve words in their dominant vs. non-dominant languages. For example, an English-Spanish bilingual can select either *dog* or *perro* to describe the same animal. Lab tasks typically show that bilinguals are slower to name pictures in their non-dominant language than in their dominant language (e.g., Gollan & Ferreira, 2009; Ivanova & Costa, 2008; Mägiste, 1979). However, when a word is planned in the context of a sentence, the attentional requirements of preceding words should increase the activation of lexical candidates (especially weakly active ones) prior to their selection. All else being equal, the name in the non-dominant language (*perro*) should benefit more than the name in the dominant language (*dog*), reducing the effect of differential accessibility on selection times. This might also partly explain why bilinguals code-switch (i.e., switch languages) during naturalistic speech (cf. Kootstra, Van Hell, & Dijkstra, 2009) even though lab tasks suggest that such switches are costly (Costa & Santesteban, 2004; Gollan & Ferreira, 2009; Meuter & Allport, 1999): If words in the currently spoken language are more active than words in the other language (and this difference is not solely due to active, language-wide inhibition; cf. Green, 1998), advance planning of a word should reduce the difference in accessibility between its names, increasing the likelihood that a bilingual will select the word in the other language and thereby code-switch.

There is also evidence that lexical accessibility drives linguistic choices at levels of representation other than lemma selection. For example, when speakers have the flexibility to choose between several syntactic structures – an active sentence like “The lightning is striking the church” or a passive sentence like “The church is being struck by lightning” – the relative accessibility of the two words that represent the point at which the structures diverge (“lightning”, “church”) affects their choice of structures (Bock, 1986, 1987). So, the more accessible the word *church* is relative to *lightning*, the more likely speakers will be to produce a passive sentence, as easily accessible words tend to be produced sooner (see Bock, 1982). The present studies suggest that the effect of lexical accessibility on syntactic structure may be modulated by the attentional demands of language production. For example, the results of Experiment 2 indicate that differences in accessibility should affect the choice of syntactic structure less when the choice point between structures occurs later in a sentence, as the attentional demands associated with planning preceding words would allow the less accessible word to ‘catch up’.

It may not even be necessary for lexical accessibility effects to underlie linguistic choices for analogous effects to be observed. If, say, an active sentence structure is simply easier to retrieve than a passive sentence structure, then if the selection of a syntactic structure requires more attentional resources than the activation of such structures, speakers’ preference for active sentence structures may be reduced in naturalistic speech. More generally, a difference in accessibility between a pair of response candidates for *any* task – language-related or otherwise – could potentially be reduced with advance planning, provided that the response selection process meets the requirements described above.

Conclusions

Stages of word production differ in their degree of automaticity. Speakers can plan semantic content and activate lexical candidates for several words at once, but they can select only one word lemma at a time. The present experiments showed that relative to planning a single word in isolation, planning a word in advance increased the amount of activation accrued prior to selection by all lexical candidates. This increase especially benefited candidates that were initially weakly active, making them more likely to be produced.

Provided that a small number of assumptions are met, the same logic may apply to any selection process in which a difference in initial accessibility between two responses is reduced with additional processing. Differences in attentional requirements between evidence accumulation and response selection may have far-reaching ramifications for decision-making in both linguistic and non-linguistic domains.

Acknowledgments

This work was supported by the NIH under Grant Numbers R01 HD051030, NIDCD011492, and DC000041. We are grateful to Liz Schotter for assistance with the norming study; to Rachel Ostrand, Xavier Alario, and Adrian Staub for helpful discussions; and to Sarah Bae, Bridget Borrison, Tiffany Chen, Kristi Cheng, Cameron Hays, Elizabeth Jimenez, Jessica Ma, Sheena Moore, Claire Perelman, and Maria Prieto for their assistance with data collection.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Kleinman, D., & Ferreira, V. S. The dissertation author was the primary investigator and author of this material.

Appendix

Critical Pictures Used in Experiments 1 and 2, Dominant Names, Name Agreement, and Exclusions

Dominant picture names and binned (between-subjects) name agreement values were computed for the base conditions in each study: IPNP (Bates et al., 2003); norming study (first block); Experiment 1 (no-preview condition); Experiment 2 (named-first condition). For the IPNP, name agreement was computed as (*elex1* + *elex2*). Pictures denoted with [a] were excluded from both experiments for having inconsistent dominant names across studies. Pictures denoted with [b] were excluded from individual experiments when the dominant name within the experiment did not match the dominant name across studies.

Picture	IPNP	Norming study	Experiment 1	Experiment 2
ax	ax (0.86)	ax (0.95)		ax (0.94)
baby		baby (0.84)		baby (0.80)
barbecue	barbecue (0.90)			grill (0.57) [a]
bow	bow (0.90)	bow (0.78)	bow (0.84)	bow (0.62)
bucket	bucket (0.66)	bucket (0.70)	bucket (0.86)	bucket (0.80)
carousel	carousel (0.60)			carousel (0.71)
chest	chest (0.63)	chest (0.55)	chest (0.67)	chest (0.73)
chicken	chicken (0.72)	chicken (0.78)	chicken (0.80)	chicken (0.76)
coat	coat (0.58)	coat (0.60)	jacket (0.58) [b]	coat (0.70)
couch	couch (0.74)	couch (0.82)	couch (0.84)	couch (0.87)
curtains	curtains (0.76)	curtains (0.86)		curtains (0.87)
frog	frog (1.00)	frog (0.96)		frog (0.98)
glass	glass (0.71)	cup (0.64)	glass (0.54) [a]	glass/cup (0.50) [a]
gorilla	gorilla (0.70)			gorilla (0.76)
gun	gun (0.92)	gun (0.96)		gun (0.94)
hat	hat (0.69)	hat (0.81)	hat (0.90)	hat (0.74)
jail		jail (0.87)	jail (0.82)	jail (0.59)
needle	needle (0.65)	needle (0.49)	needle (0.61)	syringe (0.53) [b]
painter		painter (0.76)	painter (0.93)	painter (0.77)
picture	picture (0.85)	picture (0.63)	picture (0.68)	painting (0.52) [b]
pillar	pillar (0.47)	pillar (0.62)	pillar (0.48)	pillar (0.48)
present	present (0.69)	present (0.78)	present (0.86)	present (0.66)
priest	priest (0.43)	priest (0.64)	priest (0.66)	priest (0.59)
rabbit	rabbit (0.84)	rabbit (0.69)	rabbit (0.60)	rabbit (0.82)
road	road (0.92)	road (0.94)		road (0.94)
rocket	rocket (0.98)	rocket (0.90)	rocket (0.80)	rocket (0.88)
rug	rug (0.68)	rug (0.78)	rug (0.71)	rug (0.59)
seesaw	seesaw (0.75)			seesaw (0.88)

Appendix: Continued

Picture	IPNP	Norming study	Experiment 1	Experiment 2
ship	boat (0.53)	ship (0.57)	ship (0.60) [a]	boat (0.51) [a]
stove	stove (0.72)	stove (0.69)	stove (0.63)	stove (0.84)
stroller	stroller (0.49)	stroller (0.57)	stroller (0.67)	stroller (0.53)
surgeon		doctor (0.49)	surgeon (0.58) [a]	surgeon (0.68) [a]
tape	tape (0.84)	tape (0.60)	tape (0.82)	tape (0.59)
teeth	teeth (0.79)	teeth (0.84)	teeth (0.72)	teeth (0.58)
tire	tire (0.90)	tire (0.92)		tire (0.86)
towel	towel (0.80)			towel (0.81)
trash	trash (0.43)	trash (0.50)	trash (0.46)	trash (0.69)
waiter		waiter (0.57)	waiter (0.76)	waiter (0.68)
woman	woman (0.69)	woman (0.70)	woman (0.72)	woman (0.69)
wood	wood (0.59)	wood (0.73)	wood (0.67)	wood (0.54)

References

Alario, F.-X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299-319. doi:10.3758/BF03196357

Ayora, P., & Alario, F.-X. (2009, November). *Sequential retrieval of lexical items during noun phrase production*. Poster presented at the annual meeting of the Psychonomic Society, Boston, MA.

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: a dual-task study. *Frontiers in Psychology*, *2*, 57:1-10. doi:10.3389/fpsyg.2011.00057

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

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., …Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344-380. doi:10.1016/j.jml.2004.03.002

Becic, E., Dell, G. S., & Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin & Review*, *17*, 15-21. doi:10.3758/PBR.17.1.15

Besner, D., Reynolds, M., & O'Malley, S. (2009). When underadditivity of factor effects in the Psychological Refractory Period paradigm implies a bottleneck: Evidence from psycholinguistics. *Quarterly Journal of Experimental Psychology*, *62*, 2222-2234. doi:10.1080/17470210902747187

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1-47. doi:10.1037/0033-295X.89.1.1

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575-586. doi:10.1037/0278-7393.12.4.575

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, *26*, 119-137. doi:10.1016/0749-596X(87)90120-3

Briem, V., & Hedman, L. R. (1995). Behavioural effects of mobile telephone use during simulated driving. *Ergonomics*, *38*, 2536-2562. doi:10.1080/00140139508925285

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257-271. doi:10.3758/BF03204507

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*, 491-511. doi:10.1016/j.jml.2004.02.002

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283-321. doi:10.1037/0033-295X.93.3.283

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722. doi:10.3758/BF03196827

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1058-1079. doi:10.1037/0096-1523.18.4.1058

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, *35*, 724-755. doi:10.1006/jmla.1996.0038

Ferreira, V. S., Kleinman, D., Kraljic, T., & Siu, Y. (2012). Do priming effects in dialogue reflect partner- or task-based expectations? *Psychonomic Bulletin & Review*, *19*, 309-316. doi:10.3758/s13423-011-0191-9

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187-1199. doi:10.1037/0278-7393.28.6.1187

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. doi:10.1037/a0014981

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67-81. doi:10.1017/S1366728998000133

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1-B14. doi:10.1016/S0010-0277(01)00138-X

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, *10*, 603-609. doi:10.3758/BF03196521

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313-338. doi:10.1006/jmla.1997.2547

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*, 277-288. doi:10.1016/j.actpsy.2007.06.003

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

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *20*, 824-843. doi:10.1037/0278-7393.20.4.824

Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1860-1877. doi:10.1037/a0033095

Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2009). Two speakers, one dialogue: An interactive alignment perspective on code-switching in bilingual speakers. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Interdisciplinary Approaches to Code Switching* (pp. 129-160). Amsterdam / Philadelphia: John Benjamins.

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*, 199-208. doi:10.1037/h0034633

Lachman, R., & Lachman, J. L. (1980). Picture naming: retrieval and activation of long-term memory. In L. W. Poon, J. K. Fozard, K. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 313-343). Hillsdale, NJ: Erlbaum.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-39. doi:10.3115/992628.992631

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech planning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 649–660. doi:10.1037/a0022330

Mägiste, E. (1979). The competing language systems of the multilingual: A developmental study of decoding and encoding processes. *Journal of Verbal Learning and Verbal Behavior*, *18*, 79-89. doi:10.1016/S0022-5371(79)90584-X

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 503-535. doi:10.1037/0278-7393.33.3.503

Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 523-537. doi:10.1037/a0018522

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. doi:10.1006/jmla.1998.2602

Meyer, A. S., Ouellet, M., & Häcker, C. (2008). Parallel processing of objects in a naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 982-987. doi:10.1037/0278-7393.34.4.982

Morgan, J. K., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 428-442. doi:10.1037/0278-7393.31.3.428

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 555-563. doi:10.1037/0278-7393.28.3.555

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*, 359-377. doi:10.1016/j.jml.2005.05.001

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, *63*, 541-559. doi:10.1016/j.jml.2010.08.001

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273-281. doi:10.1080/17470216508416445

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

Oppermann, F., Jescheniak, J. D., & Görges, F. (2013). Resolving competition when naming an object in a multiple-object display. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-013-0465-5

Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2008). Conceptual coherence affects phonological activation of context objects during object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 587–601. doi:10.1037/0278-7393.34.3.587

Oppermann, F., Jescheniak, J. D., Schriefers, H., & Görges, F. (2010). Semantic relatedness among objects promotes the activation of multiple phonological codes during object naming. *Quarterly Journal of Experimental Psychology*, *63*, 356–370. doi:10.1080/17470210902952480

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica*, *142*, 332-342. doi:10.1016/j.actpsy.2013.01.006

Piai, V., Roelofs, A., & Schriefers, H. (in press). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Pollatsek, A., Rayner, K., & Collins, W. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, *113*, 426-442. doi:10.1037/0096-3445.113.3.426

Schnur, T. T., & Martin, R. (2012). Semantic picture-word interference is a post-perceptual effect. *Psychonomic Bulletin & Review*, *19*, 301-308. doi:10.3758/s13423-011-0190-x

Schotter, E. R. (2011). Eye movements as an index of linguistic processing in language production. *Studies of Psychology and Behavior*, *9*, 16-23.

Schotter, E. R., Jia, A., Ferreira, V. S., & Rayner, K. (in press). Preview benefit in speaking occurs regardless of preview timing. *Psychonomic Bulletin & Review*.

Schotter, E. R., Ferreira, V. S., & Rayner, K. (2013). Parallel object activation and attentional gating of information: Evidence from eye movements in the multiple object naming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 365-374. doi:10.1037/a0028646

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*, 205-246. doi:10.1016/S0010-0277(99)00053-0

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2012, September). *Predicting the next word: Data and model from a speeded cloze task*. Paper presented at the

Conference of Architectures and Mechanisms for Language Processing, Riva del Garda, Italy.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*, 462-466. doi:10.1111/1467-9280.00386

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415-433.

van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*, 62-101. doi:10.1111/j.1551-6709.2011.01213.x

CHAPTER 4:

BACKING LIKELY WINNERS, IGNORING LIKELY LOSERS: COMPREHENDERS ONLY PREDICT BEST SENTENCE COMPLETIONS

Abstract

Comprehenders predict upcoming speech and text on the basis of linguistic input. How broad are these predictions, and 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 fragments ("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, 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 sentences pre-activate best sentence completions (*sock*) but do not affect the activation of less likely completions (*shirt*). Thus, comprehenders only predict the most probable completion for each sentence.

Language comprehenders and horror movie victims have something in common: Both would benefit from knowing what’s going to happen to 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 to 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 categories (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). 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 it matches the observed completion; that is, when the prediction is correct.[1] However, prediction may also affect the activation of words other than the most likely completion. One class of words likely to be so affected are those that are semantically related to that completion, as they share overlapping conceptual representations and thus are likely to appear in the same kinds of contexts. The present research focuses on how prediction affects these semantically related words. In other words, how does the "sock" sentence above affect the activation of the *shirt* lemma?

Three possibilities exist. The first is that semantically related lemmas are also pre-activated by strongly constraining sentences (though probably to a lesser extent than the most likely completion). If predictions are at least partly feature-based in nature, as opposed to being entirely lexically specific, this would be the natural result: Expecting to hear a word ("sock") that represents an article of clothing might lead to an increase in activation in the names of other articles of clothing, such as *shirt*, *jacket*, and *pants.* However, one consequence of predicting multiple words in a given word slot is that (at a minimum) all but one of them must be incorrect. Depending on the costs associated with incorrect predictions, making multiple predictions might be inefficient.

The second possibility is that semantically related lemmas are inhibited by strongly constraining sentences. In other words, generating an expectation of hearing "sock" might reduce the activation of the *shirt* lemma. The costs and benefits for the

---

[1] Alternative 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) or that pre-activating words in the lexicon is not equivalent to prediction (Van Berkum, 2009). The current data will not be able to adjudicate this long-standing debate, but results will be framed in terms of prediction.

language system are essentially inverted relative to those of the previous possibility. If a listener is highly certain of what an upcoming word will be, reducing the activation of competitors will, upon presentation of target, facilitate processing of the target word even more. However, if the prediction is incorrect and the sentence is instead completed by 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 strongly constraining sentences only affect the activation of their most likely completions. Under this account, predicting that a sentence will end in the word “sock” increases the activation of the *sock* lemma while leaving the *shirt* lemma unaffected. In a sense, this compromise represents a hedged bet: The comprehender’s language system is sufficiently confident in its prediction so as not to pre-activate other potential completions, but not so confident as to double down on that prediction by actively suppressing them.

Effects of sentential constraint on non-target word comprehension

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. So, 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 via either reaction times or electrophysiological measures. Schwanenflugel and colleagues’ (Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985) investigation into the scope of sentential constraint is representative of such behavioral studies. 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. In fact, responses to low-cloze completions were sometimes inhibited 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 inhibited such processing (Fischler & Bloom, 1979, 1985; Forster, 1981; Schuberth & Eimas, 1977).

In addition to reaction time 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 – 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., Jaeger & Snider, 2013; Oppenheim, Dell, & Schwartz, 2010; Pickering & Garrod, 2013), the effects of a prediction may be realized both when the success of a prediction is evaluated *and after* it 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 (“downstream”) benefits and costs that derive from the accuracy (or inaccuracy) of that prediction. Of course, under an account that does not posit prediction in the first place, neither *sock* nor *shirt* would be pre-activated, meaning there should be no effect on future processing as a result of incremental learning.

Testing 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. The present study does both by combining highly constraining sentences with a paradigm from language production research, the cumulative semantic interference task, to explore the effects of linguistic prediction. First reported in 2006 (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 three 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 a previously defined activation threshold. The presence of inhibition means that the duration of lemma selection is shorter when target lemma activation is higher, non-target (i.e., competitor) lemma activation is lower, or both. Importantly, the model does not fundamentally distinguish between these scenarios: Increasing the activation of the target lemma will cause it to inhibit its competitors more, while decreasing the activation of non-target lemmas will cause them to inhibit the target lemma less. Thus, the chief determinant of lemma selection duration is the difference between the activation levels of the target lemma and its competitors.

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 in bare picture naming, but the structure of the model – specifically, the fact that increasing target lemma activation is functionally equivalent to decreasing non-target lemma activation – establishes a clear relationship between the duration of lemma selection and the degree of semantic interference. Because the size of the cumulative semantic interference effect is determined by how active non-target lemmas are compared to the target lemma, interference should be smaller when lemma selection is faster, either because target lemma activation is increased, non-target lemma activation is decreased, or both.[2] Conversely, interference should be greater when lemma selection is slower, either because target lemma activation is decreased, non-target lemma activation is increased, or both. Thus, if a strongly constraining sentence pre-activates its high-cloze word lemma, thereby facilitating lemma selection, it should also reduce semantic interference. These effects could be further magnified or attenuated depending on whether the sentence inhibits or facilitates semantically related non-target word lemmas, respectively.

A second model of cumulative semantic interference was recently proposed by Belke (2013), who posited that the effect is attributable to learning that takes place at the interface between shared semantic features (e.g., <CLOTHING>, <WEARABLE>) and unitary, lexically specific semantic representations (<SOCK>, <SHIRT>). Under this account, repeatedly accessing the same concepts (e.g., by naming multiple pictures in the same category) causes them to accumulate activation, which they spread to related lemmas (both target and non-target). Although non-target lemmas do not actively inhibit the target (or each other), lemma selection is still competitive because the probability that a response-appropriate lemma is chosen for production is proportional to its share of the total activation of all salient lemmas (for details, see Roelofs, 1992; Levelt, Roelofs, & Meyer, 1999), so greater activation of non-target lemmas leads to longer naming latencies.

---

[2] The same pattern could be observed if, e.g., the activation levels of both target and non-target lemmas increased, but the activation of the target lemma increased more.

While it is difficult to derive predictions regarding the effects of sentential constraint on object naming from Belke's (2013) model because it has not yet been computationally implemented, the fact that cumulative semantic interference results from competition during lemma selection under her account potentially suggests that, like Howard et al.'s (2006) model, the duration of lemma selection and the degree of cumulative semantic interference should be positively related. As a result, any manipulation that speeds lemma selection should lead to a decrease in semantic interference.

A third 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 their model, there is no inhibition 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, while connections between active semantic representations and non-target lemmas (<CLOTHING>-*shirt*, <CLOTHING>-*glove*) are weakened, in proportion to those lemmas' 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).

Like the other models described above, implementing sentential constraint was outside the scope of Oppenheim et al.'s (2010) model. However, unlike those models, it provides a way to tease apart the possible effects of reading high-cloze 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. If the sentences decrease non-target lemma activation (*shirt*, *glove*), the connections between the shared semantic representation (<CLOTHING>) and those lemmas will be weakened less (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, decreasing non-target lemma activation ultimately leads to a smaller cumulative semantic interference effect. Conversely, if high-cloze sentences increase the activation of non-target lemmas, the amount of cumulative semantic interference will increase. If the activation of non-target lemmas is unaffected, the amount of cumulative semantic interference will not change.

Taken together, these models of cumulative semantic interference can potentially shed light on how prediction affects the activation of both expected and unexpected lemmas. All three models can account for a pattern of data in which high-cloze sentences speed lemma selection (thereby reducing naming latencies) while decreasing semantic interference: 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), while Howard et al. (2006) and Belke

(2013) could claim that sentences increased the difference in activation between target and non-target lemmas during selection without being able to determine the cause more specifically. However, if a reduction in naming latencies is accompanied either by an increase in semantic interference – perhaps because some of the activation induced by sentences is broad and category-specific – or no difference in semantic interference, only Oppenheim et al.'s error-based learning 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.

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 high-cloze sentence) interferes with naming *shirt* in a sentence context and whether this interference differs in magnitude from that observed in bare contexts.

Method

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.

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.

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). 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 could potentially belong to the category <BUILDING>.

For each picture, 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 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).

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.[3] 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. Across subjects, every critical picture was presented eight times in every combination of ordinal position and context.

---

[3] The 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 bear on the question of how strongly constraining sentences affect non-target activation, they will not be discussed further.

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), 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.

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 fixation point: +) 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 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 speech, or when the subject began an utterance with a filler word (e.g., “Um”). 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.

Analysis

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, harmonic mean 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.[4] 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, 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

---

[4] In addition to this multiplier, predictors were sometimes linearly scaled due to issues with model convergence. This scaling did not affect the results other than facilitating convergence, and all reported data show unscaled estimates and standard errors.

converge, all random slopes of covariates were removed to facilitate convergence; such exceptions are noted. Finally, to identify outliers, separate models were 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 the context of trial *n* 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 purposes 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 harmonic mean 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 harmonic mean 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.

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 4.1, and back-transformed subject means for theoretically relevant conditions are shown in Figure 4.1. 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 context, respectively. No other main effects or interactions were significant. In particular, the interaction between ordinal position and 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 harmonic mean 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

Table 4.1. 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.

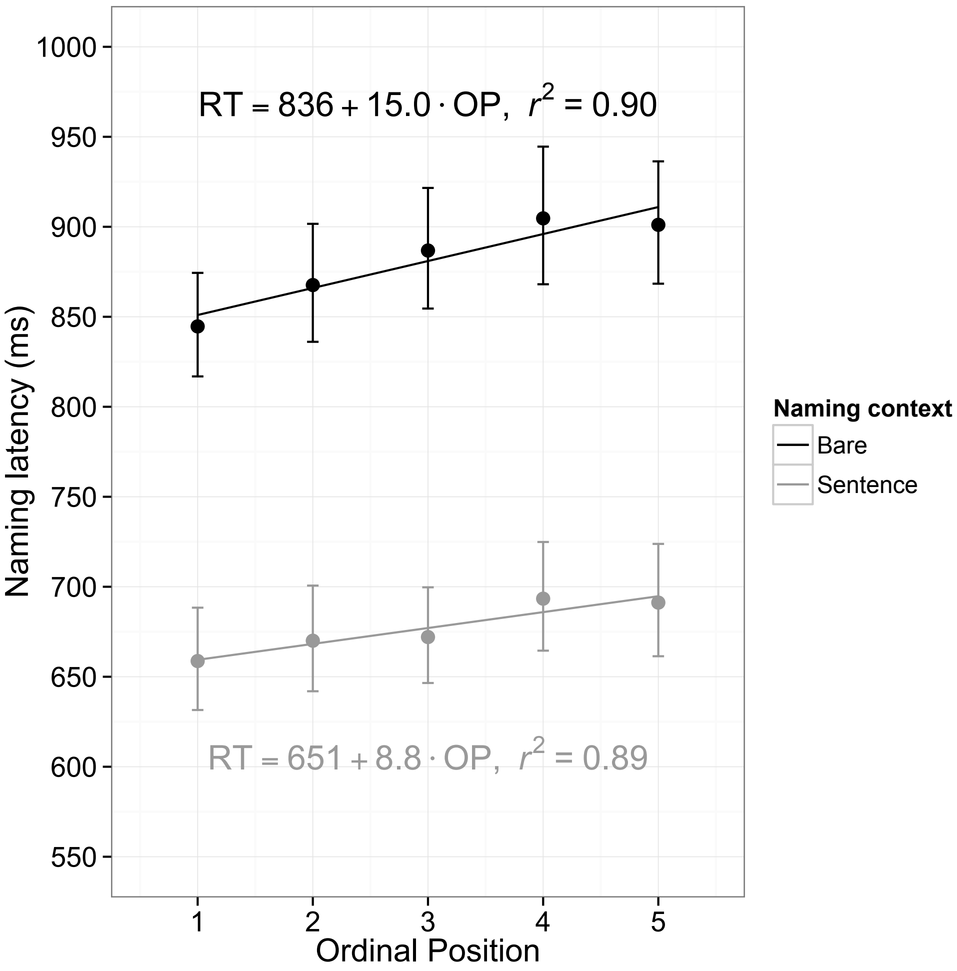


Figure 4.1. Experiment 1 back-transformed picture naming latencies shown as a function of ordinal position and context. Linear effects of ordinal position (OP) on reaction times (RT) are denoted by best-fit lines and their equations. Error bars represent 95% confidence intervals.

(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 (8.8 ms per position) was only 59% as large as the effect of ordinal position in bare contexts (15.0 ms per position), as shown in Figure 4.1. 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.

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, models in which cumulative semantic interference arises due to competitive lemma selection, instituted either via inhibition (Howard et al., 2006) or a competitive selection rule (Belke, 2013), are 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 leaving non-target lemma activation unchanged. Thus, the data favor a word-specific account of lexical prediction in which expecting a word causes extra activation to accrue only to the target lemma. Furthermore, they support a model of cumulative semantic interference in which the interference does not arise as the result of competition or inhibition between lemmas during selection.

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 well (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*).

Method

Subjects. Sixty new subjects from the same population as Experiment 1 participated in Experiment 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), and the microphone was connected to the button box indirectly via a Marantz PMD661 voice recorder.

Materials. All pictures and sentences were identical to those used in Experiment 1.

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 in

Experiment 1, the second block was designed to examine effects that are not relevant to the theoretical questions addressed in this paper, so the analysis and description of counterbalancing will be restricted to the first block (see Footnote 3). 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, 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.

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.

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 the Experiment 1 Analysis section (57). (Note that some trials violated multiple criteria.)

The model is summarized in Table 4.2, and back-transformed subject means for theoretically relevant conditions are shown in Figure 4.2. 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, context, and trial number, respectively. No other main effects or interactions were significant, including the interaction between ordinal position and context, $t = -0.054$. (Note that the effect of trial n-1 context did not distinguish between matching and mismatching sentences, but – as in Experiment 1 – only between bare and sentence contexts.) 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, $\beta = 0.14$, $SE = 0.060$, $t = 2.32$.

Discussion

As expected, the combined effect of adding a pre-exposure phase and mismatching 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.

Table 4.2. 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).

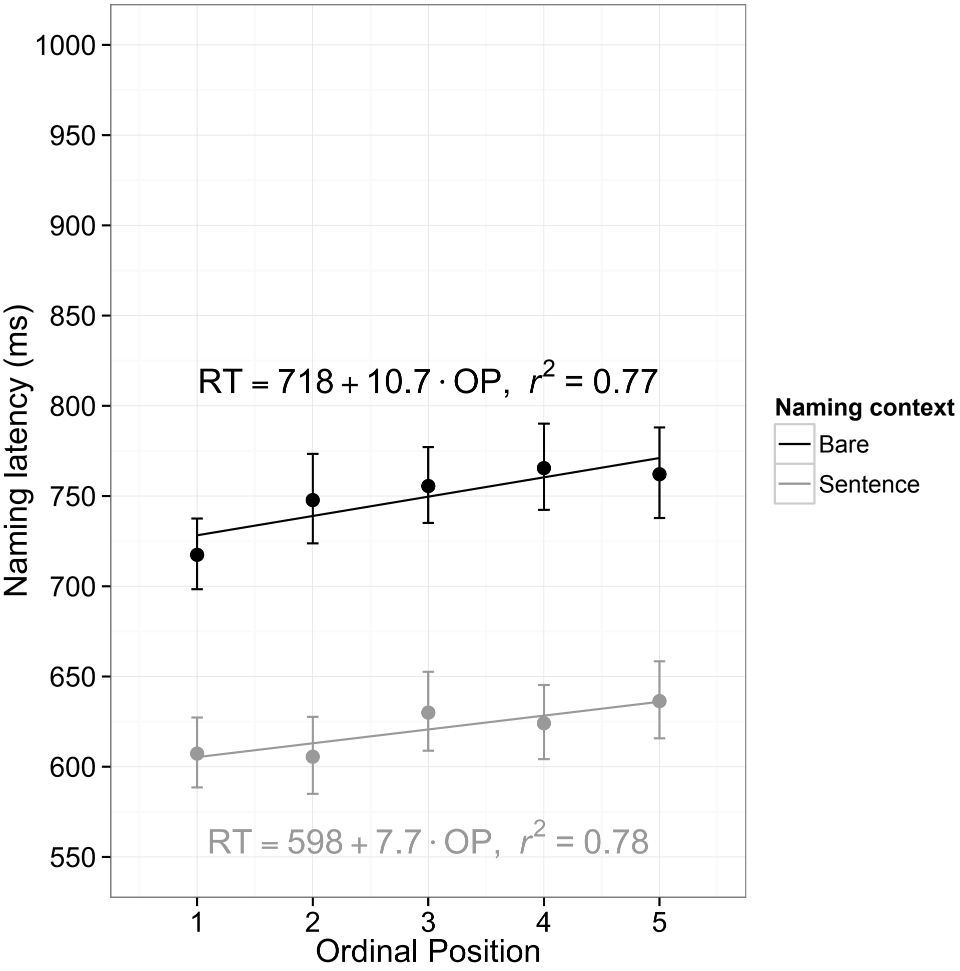


Figure 4.2. Experiment 2 back-transformed picture naming latencies shown as a function of ordinal position and context. Linear effects of ordinal position (OP) on reaction times (RT) are denoted by best-fit lines and their equations. Error bars represent 95% confidence intervals.

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, demonstrating that strongly constraining sentences increase target lemma activation while leaving non-target lemma activation unchanged. 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. Because the effects of sentential constraint must be interpreted through the lens of whichever model(s) can account for the data, the answers to these questions bear on the interpretation of how sentences affect processing.

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* relative to presenting *sock* in a bare context, the <ANIMAL>- *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 will be accompanied by a reduction in semantic interference. However, there is one modification that could potentially allow it to do so. One distinguishing feature of the 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 parameters, 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. To the extent that Belke's (2013) model could similarly account for the data from Experiments 1 and 2 by positing that sentence contexts increase the activation of conceptual representations more than bare contexts do, that account makes the same prediction. This contrasts with the prediction of Oppenheim et al. (2010) that 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 by the contexts in which they 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) and Belke (2013) 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 source of semantic interference is irrelevant, only Oppenheim et al.'s (2010) model can account for the results, allowing the previous conclusions regarding the effects of constraint on target and non-target activation to stand.

Method

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

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

Materials. All pictures and sentences were identical to those used in Experiments 1 and 2.

Design. During the experiment, subjects were presented with two blocks of 94 trials each (188 trials total) in an order determined by one of 20 stimulus lists. In Experiments 1 and 2, all five pictures within a category were presented in the same context to a given subject. In each block of Experiment 3, four pictures within a category were presented in one context (the "standard" condition), and the other picture – which was 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. 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 in each block, and each category was presented 8 times in each of 10 positions in each block.

However, in contrast to Navarrete et al. (2010), a counterbalancing error led to an unbalanced assignment of stimuli to conditions, both across and within subjects. In each block, pictures varied in how often they appeared across subjects in bare and sentence contexts (range: 20-60 presentations per context instead of 40), in combinations of context and ordinal position (range: 0-16 presentations per combination instead of 8), and in deviant conditions (range: 0-32 presentations instead of 16). Furthermore, in each block, not every category was presented equally often with each of the four combinations of ordinal position and context (range: 16-24 presentations per combination instead of 20). Most problematically, across blocks, pictures were sometimes presented twice to the same subject in the same context. In theory, this might not present a problem for pictures that were presented in bare contexts twice, but presenting a strongly constraining sentence to the same subject for the second time should greatly facilitate naming latencies. Indeed, in the second block, the 32 subjects with this counterbalancing problem named pictures in sentence contexts 116 ms faster when they had previously seen those contexts (repeated: 573 ms; unrepeated: 689 ms), compared with 41 ms faster for repeated bare contexts (unrepeated: 830 ms; repeated: 871 ms).

In dealing with this counterbalancing error, we wanted to analyze as much of the collected data as possible without skewing the results. Simply excluding trials from the second block when they were presented in the same context as the first block would have eliminated 75.0% of the deviant trials (and 18.4% of the standard trials) presented to 32 subjects, leading to an imbalance in data between subjects and greatly reducing the power to detect transfer of interference between contexts in the second block. Furthermore, data from the second blocks of Experiments 1 and 2 (as well as 3, based on the analysis above) suggested that even within a given context, naming latencies for a given picture may vary depending on the context in which that picture was named in the first block (see Footnote 3), yielding an additional source of variability in the half of the experiment with reduced power. Therefore, between analyzing both blocks for only 60% of subjects and analyzing only the first block for all subjects, we chose the latter approach, leaving it to mixed-effects models to handle imbalances in the data.

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

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., does naming *sock* in a bare context slow the subsequent naming of *shirt* in a bare context, and does naming *sock* in a sentence context slow 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., is *shirt* 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 should be slower than standard trials, whereas among pictures named in sentence contexts, deviant trials 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, regardless of the contexts in which those previous pictures were named (e.g., does naming *sock* in a bare context slow the naming of *shirt* equally in bare and sentence contexts, and does naming *sock* in a sentence context slow the naming of *shirt* equally in bare and sentence contexts?). This analysis (henceforth the “omnibus analysis”) included all data from all conditions.

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.

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, context, and their interaction; covariates included the context of trial *n*-1, its interaction with the context of trial *n*, 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 the Experiment 1 Analysis section (64).

The model is summarized in Table 4.3, and back-transformed means for theoretically relevant conditions are shown in Figure 4.3. (The data in this analysis are represented in the figure 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 trial *n*. No other main effects or interactions were significant, including the interaction between ordinal position and trial *n* context, $t = -0.31$.

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 harmonic mean RTs and no covariates, $\beta = 0.18$, $SE = 0.077$, $t = 2.37$. The apparent tenuousness of this effect, which was observed in every analysis in both Experiments 1 and 2, 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 ordinal position did not interact with context in the main analysis (even numerically trending toward a *larger* effect of ordinal position for the bare context) and that the lack of significance seems to be due to a combination of the RT transformation and the addition of a covariate, the effect of ordinal position within bare contexts is most likely real. Thus, these results should be taken to indicate that in the present experiment, 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.

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 context, deviancy ("deviant" or "standard"), and their interaction; covariates included the context of trial *n*-1, its interaction with the context of trial *n*, 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 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 the Experiment 1 Analysis section (30).

The model is summarized in Table 4.4, and back-transformed means for theoretically relevant conditions are shown in Figure 4.3. (The data in this analysis are represented in the figure 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 context. No other main effects or interactions were significant, including the

Figure 4.3. Experiment 3 back-transformed picture naming latencies shown as a function of ordinal position and context. 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). Linear effects of ordinal position (OP) on reaction times (RT) for pictures in standard conditions are denoted by best-fit lines and their equations. Error bars represent 95% confidence intervals.

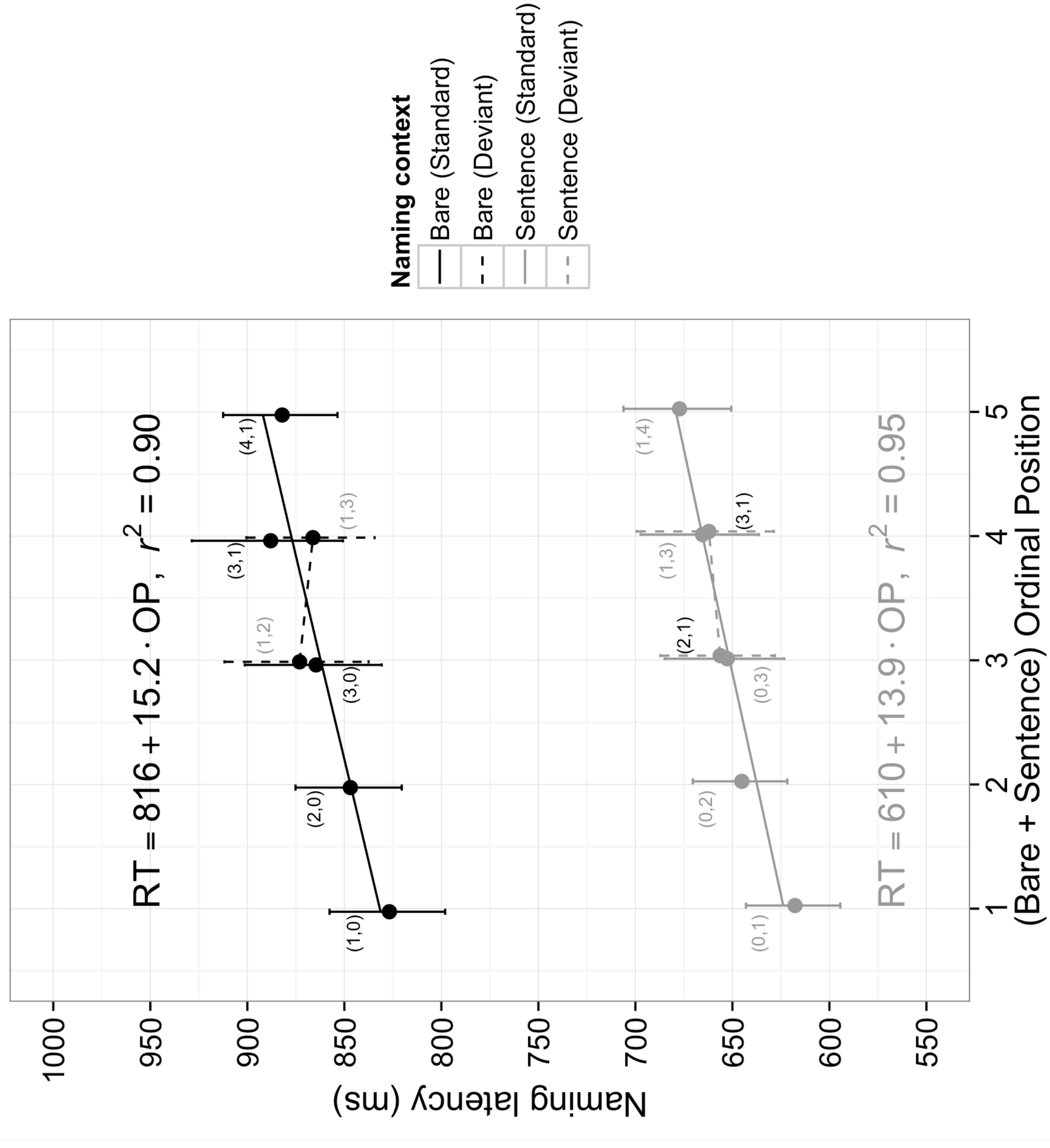

Naming context
Bare (Standard)
Bare (Deviant)
Sentence (Standard)
Sentence (Deviant)
RT = 816 + 15.2 · OP, r² = 0.90
RT = 610 + 13.9 · OP, r² = 0.95
(1,0)
(2,0)
(3,0)
(1,2)
(3,1)
(1,3)
(4,1)
(0,1)
(0,2)
(0,3)
(2,1)
(1,3)
(3,1)
(1,4)
Naming latency (ms)
1000
950
900
850
800
750
700
650
600
550
(Bare + Sentence) Ordinal Position
1
2
3
4
5

Table 4.3. 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}	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).

Table 4.4. 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).

Table 4.5. 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).

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, β = -0.29, *SE* = 0.26, t = -1.08, or sentence contexts, β = -0.01, *SE* = 0.34, t = -0.03. The effect on bare contexts was marginally significant when analyzed using untransformed RTs, β = -41.45, *SE* = 22.82, t = -1.82 (trending toward slower naming latencies for standard trials), but not when harmonic mean RTs were used without covariates, β = -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 results was only marginally significant, and even then only by using one out 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.

Omnibus analysis. This 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. So, for example, for a picture presented in a sentence context, if two same-category pictures were previously named in sentence contexts while 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, context, the interaction of bare ordinal position and context, and the interaction of sentence ordinal position and context. Covariates included the context of trial n-1, its interaction with the context of trial n, 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 the Experiment 1 Analysis section (76).

The model is summarized in Table 4.5, and back-transformed subject means for theoretically relevant conditions are shown in Figure 4.3. (The data in this analysis are represented by all points in the figure.) 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 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 harmonic mean 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, $\chi^2(1) = 7.11$, $p = .0077$, 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 harmonic mean 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.

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 picture 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 (presumably by speeding lemma selection), 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) and Belke (2013) would have to 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 predict. Thus, models of cumulative semantic interference in which the interference arises from competitive lemma selection are unable to account for the data. (Note that these findings do not bear on the question of whether or not lemma selection is competitive; they indicate only that cumulative semantic interference is not the result of a competitive selection process.) This means that strongly constraining sentences increase target lemma activation while leaving non-target lemma activation unchanged, 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.

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 4.6, and back-transformed means for theoretically relevant conditions are shown in Figure 4.4. 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 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 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 sentence harmonic mean 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, the

Table 4.6. 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 were allowed to vary by all random factors (experiment, subjects, semantic categories, pictures).

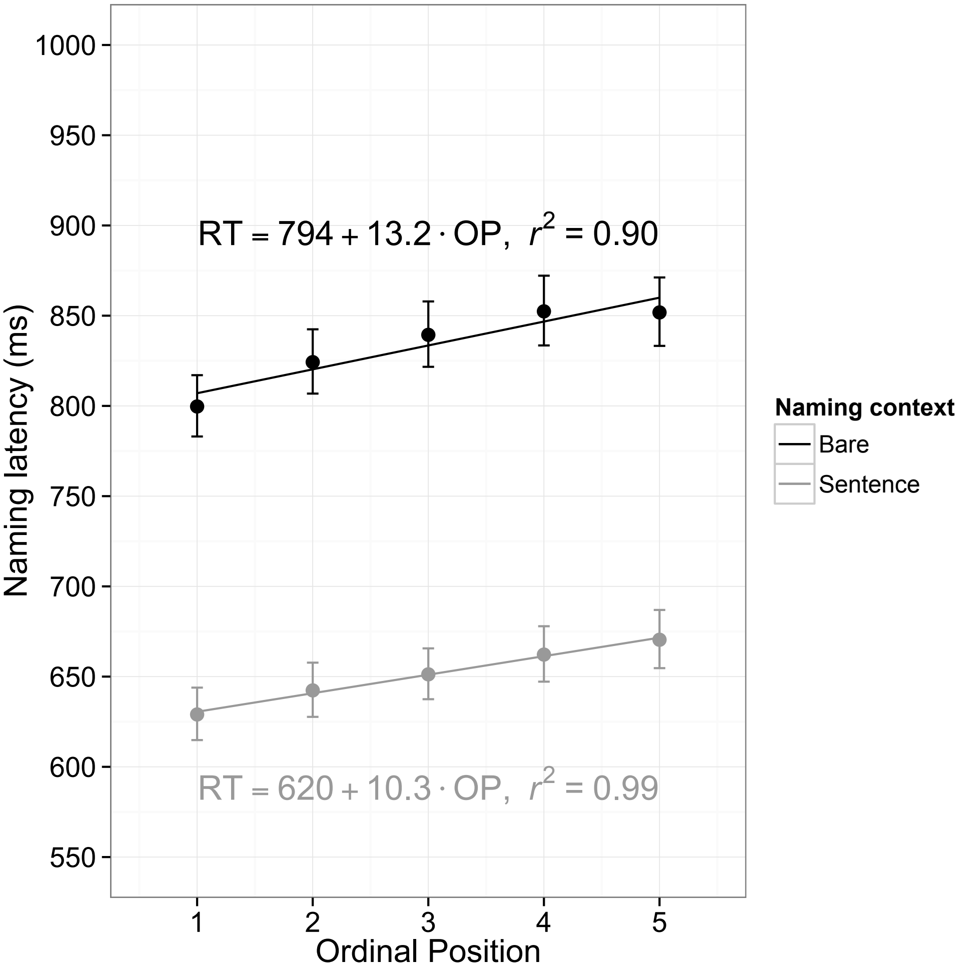


Figure 4.4. Experiment 1-3 back-transformed picture naming latencies shown as a function of ordinal position and context. Linear effects of ordinal position (OP) on reaction times (RT) are denoted by best-fit lines and their equations. Error bars represent 95% confidence intervals.

interference effect was still numerically smaller in sentence contexts than in bare contexts when they were considered separately. 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 trimmed data in the untransformed RT analysis. One subject was dropped for having no usable data in one combination of ordinal position and context. Of the other 219 subjects, 115 subjects (52.5%) 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, $p = 0.50$. 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, the non-parametric statistics reinforce the conclusion that the effect of ordinal position was not modulated by context.

General Discussion

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.

Models of cumulative semantic interference claiming that the interference emerges from competition during lemma selection (Belke, 2013; Howard et al., 2006) cannot account for such a pattern of data in their present form. According to these models, speeding lemma selection increases the difference in activation between the target lemma and non-target lemmas, either by increasing target activation, decreasing non-target activation, or both. Cumulative semantic interference emerges because non-target lemmas are more active relative to the target lemma at higher ordinal positions, so either increasing target activation or decreasing non-target activation should reduce interference. Contrary to this prediction, the amount of interference was the same in bare and sentence contexts. Furthermore, one potential assumption that could allow the models 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. Thus, the models cannot be used to explain the implications of the experimental results for effects of sentential constraint.

In contrast, Oppenheim et al. (2010) can do so. 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 was unchanged. Such a result is consistent with other behavioral and electrophysiological studies demonstrating no effect of sentential constraint on the processing of incongruous words (Federmeier & Kutas, 1999; Federmeier et al., 2007; Kutas & Hillyard, 1984; Lau et al., 2009; Stanovich & West, 1979, 1981); more generally, it supports the conclusion that strongly constraining sentences have a narrow scope of facilitation.

Necessary assumptions

The logic of the experiments, according to which the models of cumulative semantic interference proposed by Howard et al. (2006) and Belke (2013) were 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 and 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, every model could account for the results and no conclusions could be drawn about the effects of sentential constraint on target and non-target lemma activation.

Prior research supports the assumption that strongly constraining sentences affect lemma selection, however. As Griffin and Bock (1998) noted, sentential constraint affects performance even in object-identification tasks that do not require wordform retrieval (e.g., Kroll, 1990). 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) 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 the present 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.

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 the scope seems almost *too* narrow. Kutas and Hillyard (1984) and Fedemeier et al. (1999) showed that such 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 coactivate each other during naming, which is what gave rise to the interference in the first place. However, despite this semantic relatedness, the sentence that preceded the word “sock” did not further increase the activation of *shirt*; if it had, picture naming in a sentence context would have generated more semantic interference than picture naming in a bare context.

What accounts for this discrepancy in results? Perhaps the most intriguing possibility is that the degree of semantic relatedness required to ensure 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 (*pines*) 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 activate multiple words.

Another 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; instead, the vast majority were related to water ("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 coactivation among category members, we computed, for each picture in the norming study, the proportion of non-target responses that belonged to the same semantic category as the target (see Appendix). 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 trimmed, 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 stimuli presented only in sentence contexts – 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. That the numeric difference is so small and even runs counter to the predicted direction 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 every model of cumulative semantic interference predicts 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 the strongly constraining sentences did indeed pre-activate word lemmas, 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. Alternatively, the invariance of semantic interference to shared feature activation could be taken as evidence 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). More work is needed to determine which of these possibilities (if any) accurately explains why the scope of facilitation was narrower than might have been expected.

Future directions

On the basis of various issues raised in the General Discussion, at least three lines of research could help to establish the reliability and generalizability of the conclusion that strongly constraining sentences do not affect the activation of non-target lemmas. First, to address concerns about picture presentation 'washing out' effects of sentences, a version of the task could be run without pictures, in which subjects simply fill in the blank on each trial as quickly as possible. If sentences do not affect the activation of non-target lemmas, no interference should accumulate in such a paradigm. Second, shared feature activation could be systematically manipulated. Ideally, it would be possible to craft two sentences for each item that are balanced on cloze probability of the target response but differ in the extent to which non-target responses belong to the same semantic category. By controlling an important source of variation between sentences, this would increase the likelihood of being able to detect downstream effects of prediction. Third, the degree of sentential constraint could be varied. If more weakly constraining sentences have a broader scope of activation (e.g., Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985), they should generate more cumulative semantic interference.

Conclusions

Comprehenders generate expectations about upcoming speech and text on the basis of linguistic input. Depending on how strongly context constrains these expectations, they may take the form of lexically specific predictions. Speakers might generate and evaluate multiple, graded predictions simultaneously, 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 effect of strongly constraining sentences was to increase target activation. Thus, there was no evidence that comprehenders predict more than one word at a time.

Acknowledgments

This work was supported by the NIH under Grant Numbers R01 HD051030, NIDCD011492, and DC000041. We are grateful to Gary Oppenheim for helpful discussions and to Sarah Bae, Kristi Cheng, Cameron Hays, Elizabeth Jimenez, Danielle Lew, Cheryl Ma, Jessica Ma, Gabriela Meckler, Sheena Moore, Brittany Nielsen, Shivani Patel, and Kurina Wolff for data collection.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Kleinman, D., Runnqvist, E., & Ferreira, V. S. The dissertation author was the primary investigator and author of this material.

Appendix

Critical Stimuli Used in Experiments 1-3

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.

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 water stuck in her	0.73	0.73	0.22
	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

Appendix, Continued

Semantic category	**Picture**	**Sentence frame**	**Exact match**	**Concept match**	**Other same-category responses**
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
	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
Clothes	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
	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
Farm animals	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 get home and sleep in his own	0.98	0.98	0.00
	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
	table	After cooking supper, Jan set the food on the dinner	0.99	0.99	0.00

Appendix, Continued

Semantic category	**Picture**	**Sentence frame**	**Exact match**	**Concept match**	**Other same-category responses**
House parts	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
	window	The room was getting stuffy, so Cody opened a	0.97	0.97	0.03
Musical instruments	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 ability to sing and play the	0.21	0.21	0.76
	violin	Joshua lifted the bow high into the air and touched it to the strings of his	0.47	0.48	0.03
Tableware	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

Appendix, Continued

Semantic category	**Picture**	**Sentence frame**	**Exact match**	**Concept match**	**Other same-category responses**
Transport	bus	The football team traveled to the away game together on a	0.93	0.93	0.03
	car	When Rita ran out of gas, she had to ask people to help push her	0.97	0.98	0.01
	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

References

Alario, F.-X., & Moscoso del Prado Martín, F. (2010). On the origin of the "cumulative semantic inhibition" effect. *Memory & Cognition*, *38*, 57-66. doi:10.3758/MC.38.1.57

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

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

Baayen, R. H., Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12-28.

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

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., …Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344-380. doi:10.1016/j.jml.2004.03.002

Belke, E. (2013). 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*, *69*, 228-256. doi:10.1016/j.jml.2013.05.008

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *Series B*, *26*, 211-252.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257-271. doi:10.3758/BF03204507

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1121. doi:10.1038/nn1504

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75-84. doi:10.1016/j.brainres.2006.06.101

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-

495. doi:10.1006/jmla.1999.2660

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187-1199. doi:10.1037/0278-7393.28.6.1187

Fischler, I. S., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*, 1-20. doi:10.1016/S0022-5371(79)90534-6

Fischler, I. S., & Bloom, P. A. (1985). Effects of constraint and validity of sentence contexts on lexical decisions. *Memory & Cognition*, *13*, 128-139. doi:10.3758/BF03197005

Forster, K. I. (1981). 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*, *33*, 465-495. doi:10.1080/14640748108400804

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*, 89-106. doi:10.1162/jocn.1996.8.2.89

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313-338. doi:10.1006/jmla.1997.2547

Hagoort, P., Baggio, G., Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (4th ed., pp. 819-936). Boston, MA: MIT Press.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464-482. doi:10.1016/j.cognition.2005.02.006

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57-83. doi:10.1016/j.cognition.2012.10.013

Kroll, J. F. (1990). Recognizing words and pictures in sentence contexts: A test of lexical modularity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 747-759. doi:10.1037/0278-7393.16.5.747

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potential reflect semantic incongruity. *Science*, *207*, 203-205. doi:10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potential during reading reflect word expectancy and semantic association. *Nature*, *307*, 161-163. doi:10.1038/307161a0

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain & Language*, *111*, 161-172. doi:10.1016/j.bandl.2009.08.007

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-39. doi:10.3115/992628.992631

Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, *134*, 279-289. doi:10.1016/j.actpsy.2010.02.009

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, *4*, 15-22. doi:10.1162/jocn.1992.4.1.15

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

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 377-392. doi:10.1017/S0140525X12003238

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510-532. doi:10.1037/0033-2909.114.3.510

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107-142. doi:10.1016/0010-0277(92)90041-F

Schuberth, R. E., & Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 27-36. doi:10.1037/0096-1523.3.1.27

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 344-354. doi:10.1037/0278-7393.14.2.344

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232-252. doi:10.1016/0749-596X(85)90026-9

Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, *7*, 77-85. doi:10.3758/BF03197588

Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 658-672. doi:10.1037/0096-1523.7.3.658

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*, 297-314. doi:10.1016/j.jml.2012.12.002

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415-433.

Van Berkum, J. J. A. (2009). The neuropragmatics of “simple” utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory*, eds. U. Sauerland & K. Yatsushiro, pp. 276-316. Basingstoke, UK: Palgrave Macmillan.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443-467. doi:10.1037/0278-7393.31.3.443

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176-190. doi:10.1016/j.ijpsycho.2011.09.015

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potential to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*, 165-168. doi:10.16/S0304-3940(03)00599-8

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). 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*, *39*, 483-508. doi:10.1016/S0010-9452(08)70260-0

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). 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*, *16*, 1272-1288. doi:10.1162/0898929041920487

CHAPTER 5:

GENERAL DISCUSSION

Together, the studies in this dissertation investigated how preparing to plan the next word, or expecting to hear a particular word next, affects the activation of words in the lexicon.

Study 1 showed that speakers cannot select a word for production while they are simultaneously engaged in an attention-demanding task, indicating that word selection requires domain-general attentional resources. It also served as a demonstration that when one stage of a task (word selection) requires domain-general attentional resources but another (word reading) does not, this difference in automaticity can affect performance in dual-task compared with single-task settings.

Study 2 relied on both of these results to investigate how planning ahead affects the words that speakers say. Instead of a difference in automaticity between word selection and word reading, the critical difference was between word selection (which requires central attention) and the activation of lexical candidates for production (which does not). It showed that this difference buys the production system more time to activate potential lexical candidates when speakers plan words in advance relative to planning words in isolation, and that this extra time benefits weakly active lexical candidates more than strongly active ones.

Study 3 investigated what kinds of predictions comprehenders make about upcoming words on the basis of prior linguistic input. When a strongly constraining sentence suggests that the word “sock” is likely to appear, the names of other articles of clothing are likely to be at least partially supported by the context as well. Nevertheless, the results of Study 3 indicated that comprehenders did not modulate the activation of words from the same semantic category as the best sentence completion.

These results reveal that speakers and listeners benefit both from their ability to plan ahead and, perhaps counter-intuitively, from the limitations of this ability. The benefits of being able to plan ahead are obvious: Being able to prepare for upcoming words in advance reduces the processing burden for both speakers and listeners, permitting them to amortize the costs of planning and processing each word over a longer period of time. This allows speakers to recognize and signal upcoming processing difficulty (Clark & Fox Tree, 2002), to gauge how soon they can begin to speak while still maintaining fluency (e.g., Griffin, 2003), and to adjust the division of their attentional resources between planning the next word and planning upcoming words as needed (e.g., Malpass & Meyer, 2010). For listeners, planning ahead makes it easier to understand speech at the rate it is produced and to integrate words into the discourse context. These skills are especially useful for linguistic development, as the ability to process speech rapidly may make it easier for a child to extract more information from the speech stream, increasing their vocabulary and the complexity of their utterances (Fernald, Perfors, & Marchman, 2006).

Although it is beneficial, planning ahead is not a cost-free endeavor, and constraints on attention and working memory limit the amount of advance planning that speakers and listeners can do. In addition to protecting the processing of the current word from interference from upcoming words, the studies in this dissertation suggest that these limitations may be beneficial as well. By asymmetrically benefiting weakly active lexical candidates, attentional bottlenecks in word production facilitate linguistic diversity, increasing the variety of words that speakers use. In comprehension, it is possible that activating predicted lexical candidates, like activating lexical candidates for production, requires some attentional resources. That could explain why comprehenders only predict a single word from a strongly constraining context: Making predictions requires effort. If the comprehension system constantly retunes itself on the basis of the successes and failures of past predictions (e.g., Jaeger & Snider, 2013), making an incorrect prediction can affect processing difficulty both when the prediction is disconfirmed and whenever a similar semantic context arises in the future. By limiting themselves to a single word-specific prediction, comprehenders may be attempting to strike an optimal balance between processing already-spoken words and predicting as-yet unspoken words while limiting the amount of reweighting that their comprehension system will have to do.

How do the findings of Studies 1-3 collectively constrain their interpretation?

Study 1 (Experiment 2) showed that in a dual-task paradigm (tone categorization followed by picture naming), cumulative semantic interference (CSI) was unaffected by task overlap. Study 2 showed that words can continue to accrue activation while attentional resources are occupied. This means that in Study 1, appropriate picture names likely accrued activation while subjects were busy preparing a response to the tone – but, as noted, this extra activation did not modulate the amount of interference. What implications does this have for models of CSI and for word production more generally?

One possibility is that the attentional bottleneck had an effect similar to the strongly constraining sentences from Study 3. Those sentences increased the activation of the target word while leaving non-target activation unaffected, which, according to Oppenheim, Dell, and Schwartz (2010), should have facilitated word selection without modulating CSI – the observed pattern of data. If the extra pre-selection planning attributable to the tone task in Study 1 caused activation to accrue only to the target picture name and not to its semantic competitors, Oppenheim et al.'s model would correctly predict that CSI should not have been affected by task overlap.

However, such an explanation would also need to account for the fact that the Stroop effect is unaffected by attentional manipulations in dual-task experiments (Fagot & Pashler, 1992; Piai, Roelofs, & Schriefers, in press; but see Magen & Cohen, 2002, 2010) and the fact that, despite some complications described in Study 1 (Ayora et al., 2011; Dell'Acqua, Job, Peressotti, & Pascali, 2007; Kleinman, 2013), picture-word interference is often unaffected as well (Piai & Roelofs, 2013; Piai et al., in press; Schnur & Martin, 2012; van Maanen et al., 2012). Applied to these experiments, the hypothesis described above would posit that the target word (and only the target word) could receive activation when attention was occupied. Setting aside the improbability of this assumption – how could the production system have identified the target word prior to selection? – these interference effects are generally considered strong evidence for competition during word selection (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 2003), which means that increasing the activation of the target should reduce both Stroop interference and picture-word interference. For the hypothesis to be maintained, it would need to be shown that semantic interference does not result from competition.

As this is a heavy burden of proof, it may be more straightforward (as briefly noted in the Study 2 General Discussion) to assume that activation does *not* reliably spread to lemmas when attentional resources are occupied. This would neatly explain why semantic interference effects, which have a lemma selection locus, are not generally modulated in dual-task situations. To also account for the results of Study 2, it would be necessary to assume that activating *lexical concepts* requires fewer attentional resources than selecting one. So, for example, a speaker who wants to say “cat” as word *n*+1 could activate the concepts <CAT> and <DOG> in advance, but could not select one until after attention-demanding planning for word *n* was complete. This assumption allows the essential logic that motivated Study 2 to remain the same while being moved to an earlier step in the chain of word production processes. As such, it represents the best way to reconcile the results of Studies 1-3 with prior research using a single set of theoretical assumptions.

Summary

In naturalistic speech, no word is an island. Each one is produced and comprehended within a rich discourse context according to pragmatic, syntactic and semantic constraints. Just as speakers use this contextual information to select upcoming words, listeners use it to predict what those upcoming words will be. At the same time, the ability of speakers to plan ahead is limited by bottlenecks in word production that prohibit the selection of multiple words in parallel, and comprehenders may suffer costs for incorrect predictions. Together, the benefits and limitations of planning ahead have important effects on the words that speakers say and the words that comprehenders predict.

References

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: a dual-task study. *Frontiers in Psychology*, *2*, 57:1-10. doi:10.3389/fpsyg.2011.00057

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*, 73-111. doi:10.1016/S0010-0277(02)00017-3

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722. doi:10.3758/BF03196827

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1058-1079. doi:10.1037/0096-1523.18.4.1058

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98-116. doi:10.1037/0012-1649.42.1.98

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, *10*, 603-609. doi:10.3758/BF03196521

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57-83. doi:10.1016/j.cognition.2012.10.013

Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1860-1877. doi:10.1037/a0033095

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-39. doi:10.3115/992628.992631

Magen, H., & Cohen, A. (2002). Action-based and vision-based selection of input: Two sources of control. *Psychological Research*, *66*, 247-259. doi:10.1007/s00426-002-0099-0

Magen, H., & Cohen, A. (2010). Modularity beyond perception: Evidence from the PRP paradigm. *Joural of Experimental Psychology: Human Perception and*

Performance, *36*, 395-414. doi:10.1037/a0017174

Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 523-537. doi:10.1037/a0018522

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

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica*, *142*, 332-342. doi:10.1016/j.actpsy.2013.01.006

Piai, V., Roelofs, A., & Schriefers, H. (in press). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88-125. doi:10.1037/0033-295X.110.1.88

Schnur, T. T., & Martin, R. (2012). Semantic picture-word interference is a post-perceptual effect. *Psychonomic Bulletin & Review*, *19*, 301-308. doi:10.3758/s13423-011-0190-x

van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*, 62-101. doi:10.1111/j.1551-6709.2011.01213.x