NORTHWESTERN UNIVERSITY

# The Role of Domain-General Executive Functions, Articulation, and Conceptualization during Spoken Word Production

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Angela Fink

EVANSTON, ILLINOIS

June 2016

ABSTRACT

The role of domain-general executive functions, articulation, and conceptualization during spoken word production

Angela Fink

In this dissertation, we present three empirical studies exploring the relationship between the central planning processes of spoken word production—lexical selection, phonological encoding, and phonetic encoding—and three other cognitive processes traditionally considered separate or peripheral to this core system. Study 1 examines the role of domain-general executive functions, which facilitate thoughts and actions by directing attention and/or cognitive resources toward the task at hand, in resolving conflict during lexical selection. A growing body of behavioral and neurological evidence suggests that inhibitory executive functions, known to suppress non-target representations, can help manage conflict among co-active lexical representations. Across 4 experiments, we attempt to support this hypothesis by demonstrating that engagement of inhibition can modulate the difficulty of lexical selection, indexed by response time data. Study 2 investigates interactions between lexical selection and subsequent articulatory processing. Specifically, we examine word duration data from study 1 and a collaborator's experiments, testing whether the difficulty of lexical selection and the timing of response initiation influence speakers' articulatory outcomes. Finally, study 3 probes the relationship between spoken production and its lead-in process, conceptualization. A series of post hoc analyses on the response time data from study 1 explores the underlying structure of the semantic representations that send activation to the production system. Based on these three lines of research, we argue for more cognitively integrated theories of spoken word production.

## ACKNOWLEDGEMENTS

To my advisor Matt Goldrick, the incredibly adept juggler of oh-so-many hats: thank you for offering me unwavering support throughout my time at Northwestern. You taught me to think critically, analyze in a principled way, express myself clearly, and seek insight from others. Early in the program, you offered me opportunities; later on, you gave me room to discover some of my own. No matter where I land next, I aspire to follow in your footsteps as a conscientious colleague who leads by example and inspires achievement.

Thank you to Ann Bradlow and Sid Horton, for serving on my committee and providing constructive comments several times over. Thank you to the members of the Sound Lab, for sharing your expertise and keeping me honest. Thanks to all my gradmates, but most especially my cohort: Michael Blasingame, Samantha Heidenreich, Julie Matsubara, and David Potter. Without you to commiserate and blow off steam with, who knows if I would have made it through. Thanks to Angela Cooper and Erin Gustafson for the many girls' nights that kept me sane while dissertating.

A special thank you to my all relatives living in Chicago, both now and then. To Gina and John Fink; Michael, Toni, and Anna Fink; Jack and Joanne Almiro; Julie and Kaitlyn Grace; to Dan, Peter, and Nancy Kujawinski, I thank you for making me feel at home and for always keeping me fed. Thanks to Alina, Blaze, and Sylvie Kujawinski for brightening my time here with your smiles and imaginations. Thanks to Liza Kujawinski-Behn for setting an example of success and enthusiastically cheering me through the home stretch. Above all, thank you to Jo Kujawinski for looking after me in countless ways, sharing in the excitement of my successes, and bolstering me during times of crisis.

I am endlessly grateful to my Nerinx girls: Carolyn Duranowski, Katie Fortune, Amanda Hayden, Sarah Hitzel, Dailee Joyce, Keegan Kelley, and Rebecca King. When I felt overwhelmed, you gave me the solid, familiar ground of friendship to stand on. No matter how long it took me to reach out, you were always ready to respond. I am truly lucky to know such strong and loving women; it is a privilege to call you my friends.

Thanks to everyone at Glenwood Dance Studio, for creating a sense of community and showing me how to escape any mental turmoil through movement. Thanks also to the Ravenswood gang, for leading me on adventures in the great outdoors and at the gaming table. It has been a hilarious, heartwarming experience getting to know you all.

Finally, I owe the largest debt of gratitude to my immediate family, who have grown me into the person I am today. To my mom, Jean Fink, thank you for having such high expectations and always holding me accountable. Thank you for loving me and showing me with your every action how much you care. To my dad, David Fink, thank you for your quiet constancy. Your ability to calmly listen gives me space to discover what I really think and feel, and your subtle, playful wit brings me joy. To my brother, Patrick Fink, thank you for diligently maintaining the connections between us and among all our family members. Your kind-hearted nature and inquisitive mind never cease to amaze me.

To my newest family member, my fiancé Charles Welling, I want to say thank you for so many things. For your unfettered honesty and your insistence that I speak my mind too; for your open-mindedness, your willingness to listen, and your keen desire for knowledge; for your passion for the outdoors and your spirit of adventure; for the way you make me laugh, I love and thank you.

Thank you all.

# TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

## 1.1 Introduction

Ultimately, research on spoken word production seeks to understand the full range of representations and cognitive processes that support naturalistic spoken communication. However, such research necessarily begins with a more tractable problem: specifying the domain-specific cognitive architecture and mechanisms that generate single word utterances. This question may seem narrow, but it is hardly trivial. Spoken production entails at least three core levels of linguistic representation and processing (e.g., Levelt, Roelofs, & Meyer, 1999). First, the process of *lexicalization* or *lexical selection* maps the semantic representation(s) of a concept onto a word-level, or lexical, representation (e.g., the features <furry>, <four-legged>, <canine> activate DOG). Next, during *phonological encoding*, that lexical representation is mapped onto its phonemic sound structure (e.g., DOG activates /d/-/ɔ/-/g/). Finally, *phonetic encoding* adds additional detail to this segmental representation, taking into account co-articulation and contextual variation in the target phonemes (e.g., /d/-/ɔ/-/g/ becomes [dɔğ]). This complex framework is further complicated by a lack of consensus on the independence vs. integration of these component processes (see Ernestus, 2014, for a review of abstract, exemplar, and hybrid models of language processing). Clearly, even with a narrow focus on production-internal cognitive processes, research on spoken word production covers an impressive scope.

However, in recent years, the field has increasingly attended to the relationship of spoken production processes with other cognitive systems (see Fink & Goldrick, 2015, for a brief review and discussion). For instance, there is mounting evidence that domain-general cognitive control processes, also referred to as *executive functions* (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), may play a critical role in language processing, particularly under conditions of conflict (Novick, Trueswell, & Thompson-Schill, 2010). Furthermore, there is growing evidence for interactions between central production processes (i.e., lexical selection and phonological and phonetic encoding) and the peripheral processes that interface with this system: conceptualization and articulation. Before production planning begins, amodal conceptual representations get mapped onto lexicalized semantic representations, sometimes called lexical concepts, during *conceptualization* (Levelt et al., 1999). The dynamics and output of this process can influence speakers' production performance (e.g., Alario & del Prado Martin, 2010; Belke, 2013). After (a portion of) the speech plan is fully specified, it is implemented by the motor system during *articulation*. Articulatory/acoustic variation can reveal a great deal about speech planning, like the representation and influence of lexical information (e.g., lexical frequency: Bell, Gregory, Girand, & Jurafsky, 2009; Gahl, 2008) and the temporal overlap between planning processes (Kello, 2004; Kello, MacWhinney, & Plaut, 2000).

In this dissertation, we explore the interactions of these arguably separate or peripheral cognitive processes with the central processes of spoken word production. As a window into these interactions, we focus specifically on the stage of lexical selection and a well-established effect arising during that process known as *cumulative semantic interference*. Such interference is commonly observed in a paradigm known as continuous picture naming (Howard, Nickels, Coltheart, Cole-Virtue, 2006), where speakers perform the simple task of naming one picture after another. Crucially, the stimulus list comprises many different semantic categories, and multiple studies have shown that response times

(RTs) increase linearly with each subsequent naming of a given semantic category (Belke, 2013; Belke & Stielow, 2013; Howard, et al., 2006; Navarrete, Mahon, Caramazza, 2010; Oppenheim, Dell, Schwartz, 2010; Schnur, 2014). We explore how semantic interference effects interact with executive functions, articulation, and semantic structure in turn.

In the remainder of this introductory chapter, we set the stage for the empirical work reported in this dissertation. For each proposed interaction, we begin by reviewing the supporting evidence. We rely primarily on behavioral data, but also draw from neuro-imaging and neuropsychological data to strengthen some arguments. Next, we outline the logic and design of each empirical study. We conclude by foreshadowing the results of each study, ultimately tying them together to provide a broad view of production processing.

## 1.2 Executive Functions can (but do not Necessarily) Support the Resolution of Linguistic Conflict

Every stage of production processing is susceptible to conflict, due to the co-activation of multiple representations. For instance, as mentioned above, lexical selection becomes more challenging when semantically related lexical representations, i.e., *semantic neighbors*, become co-active and vie for selection (e.g., retrieving CAT is harder shortly after naming DOG; Wheeldon & Monsell, 1994). Domain-specific mechanisms, like an internal monitor (Levelt, 1983), may be sufficient for managing this linguistic conflict during production. Considering that language processing is a highly practiced behavior, it seems reasonable to assume that such specialized mechanisms are well equipped for handling everyday linguistic conflict.

Nonetheless, a growing body of research indicates that domain-general executive functions, especially those that fall under the umbrella of *inhibition*, participate in linguistic conflict resolution, particularly during lexical selection. Supporting evidence comes from behavioral studies (Belke & Stielow, 2013; Crowther & Martin, 2014; Hsu & Novick, 2016; Shao, Meyer, Roelofs, 2013; Shao, Roelofs, & Meyer, 2012; cf. Alario, Ziegler, Massal, and de Cara, 2012), neuroimaging studies of typical adult speakers (de Zubicaray, McMahon, & Howard, 2015; de Zubicaray, Wilson, McMahon, & Muthiah, 2001; Kan & Thompson-Schill, 2004; Ries, Kazmark, Navarrete, Knight, & Dronkers, 2015; Schnur, Schwartz, Kimberg, Hirschorn, Coslett, & Thompson-Schill, 2009; Shao, Meyer, Acheson, & Roelofs, 2014), and neuropsychological studies of impaired speakers (see Novick et al., 2010 for a review; Schnur, Schwartz, Brecher, & Hodgson, 2006; Schnur et al., 2009). In study 1, we will review some of the most compelling evidence to date.

For now, we note that behavioral studies have generally relied on correlational evidence to establish a link between inhibitory executive functions, which suppress non-target representations (Miyake et al., 2000), and lexical selection. For example, Shao et al. (2012) used correlational analyses to test whether three executive functions—inhibition, working memory updating, and task switching (Miyake et al., 2000)—factor into standard object and action naming. They selected one task apiece to index participants' abilities using each executive function. The results revealed that participants' performance on the inhibition task (stop-signal; Logan, 1994) predicted their mean object and action naming RTs, while performance on the working memory updating task (operation span; adapted from Turner & Engle, 1989) predicted the shape of their RT distributions. Performance on the task switching paradigm (switching between classifying stimulus color vs. shape) had no relationship to the production data. The authors concluded that domain-general executive functions support lexical selection quite broadly, even in the absence of experimentally induced linguistic conflict.

The first study of this dissertation aims to provide novel behavioral evidence for a causal relationship between inhibition and the resolution of lexical conflict. In addition, we seek to make this hypothesis more precise by probing the role of two distinct inhibitory executive functions in lexical selection. The details of this study are laid out below.

### 1.2.1 Outline of study 1

In this study, we contrast the inhibitory executive functions of *response inhibition* and *(proactive) interference resolution*, which previous work has shown to be statistically and functionally distinct (Friedman & Miyake, 2004; Persson, Welsh, Jonides, & Reuter-Lorenz, 2007). Response inhibition, sometimes referred to as selective inhibition (e.g., Shao et al., 2013; Shao et al., 2014), facilitates performance by suppressing pre-potent but incorrect response candidates. The Stroop task is a classic paradigm for engaging response inhibition, given that participants must override the dominant response of reading written words in order to name the text color (MacLeod, 1991). In contrast, interference resolution supports task performance by biasing activation away from any previously—but no longer—relevant representations. This executive function is tapped by list recall tasks, where interference resolution is needed to channel activation away from list A representations in order to allow recall of list B. Either of these inhibitory processes might provide a benefit during lexical selection, depending on the dynamics of lexical co-activation.

To probe the role of these two executive functions in lexical selection, we utilize a relatively new paradigm called the *process-specific negative transfer* paradigm, designed to test whether a target cognitive process is shared across two tasks (Nelson, Reuter-Lorenz, Persson, Sylvester, & Jonides, 2009; Persson, Larsson, & Reuter-Lorenz, 2013; Persson et al., 2007). This paradigm uses a pre- vs. posttest design: baseline performance in task A is assessed at pretest, task B is intensively practiced during treatment, and then posttest performance on task A is analyzed for treatment effects. The critical assumption is that treatment effects will only emerge when a specific executive function (or other cognitive process) is shared across tasks A and B. This approach has previously provided evidence of a selective role for interference resolution, but not response inhibition, in resolving conflict during semantic memory retrieval (Persson et al., 2007).

In a similar fashion, we used the negative transfer paradigm to assess the role of these target executive functions during lexical selection. In experiments 1 and 2, we employed the continuous naming paradigm (Howard et al., 2006) at pre- and posttest to generate conflict among semantic neighbors. Experiment 1 featured an interference resolution task during treatment, while experiment 2 involved a response inhibition treatment task. We expected that either inhibitory executive function might elicit treatment effects, depending on the dynamics of lexical co-activation during continuous naming. As an experimental control, we replaced that pre- and posttest task with continuous picture classification (Belke, 2013) in experiments 3 and 4. During this task, semantic neighbor co-activation is known to benefit response selection, such that RTs linearly decrease each time a member of a given semantic category is classified (Belke, 2013). We expected that if the target executive functions are recruited specifically to resolve lexical conflict, then they should have no influence on participants' classification performance. Once again, an interference resolution treatment was utilized in experiment 3, before a response inhibition treatment in experiment 4.

To foreshadow the results of study 1, they defied expectations almost across the board. Significant treatment effects were observed in experiment 2, but not in experiment 1. These data are interpreted as evidence that response inhibition, but not interference resolution, can help resolve lexical conflict. However, significant treatment effects were also observed in experiments 3 and 4, calling our original conflict-based account into question. Drawing on recent evidence of short-term cross-talk between executive function and domain-specific tasks (Hsu & Novick, 2016; Shell, Linck, & Slevc, 2015), we offer a conflict adaptation explanation of these findings. Specifically, we argue that the target executive functions are not recruited by default during these spoken production tasks. Instead, the experience of concentrated conflict during treatment triggered their engagement, which carried over to the subsequent posttest. In other words, we argue that the treatment effects were a by-product of contextually-induced executive function up-regulation, rather than evidence for the automatic recruitment of domain-general control during lexical conflict.

### 1.3 Lexical Effects on Articulation are Sensitive to Response Selection and Initiation

In our second study, we explore interactions between lexical selection and articulation. We sometimes describe such interactions as "long distance," reflecting their span across multiple levels of representation (i.e., lexical, phonological, and phonetic). This description portrays the interactions as planning-mediated: it assumes that lexically-conditioned articulatory/acoustic variation results when disruptions during lexical selection are passed down through subsequent production processes via cascading activation. However, such interactions might also arise in other ways. For instance, articulation might begin before production planning is complete (e.g., Kawamoto, Liu, & Kello, 2015), allowing difficulties in ongoing planning to directly influence articulatory outcomes after articulation has been initiated (Kello, 2004; Kello et al., 2000; but cf. Damian, 2003). In this case, we expect that the timing of response initiation relative to planning might influence whether direct interactive effects are observed. We explore these possible sources of interaction using word durations to index the speed of articulatory processing. Critically, we test for lexical effects on word durations when RT—i.e., the timing of response initiation—is taken into account as a covariate.

Previous studies using word durations to test for long distance interactions between lexical selection and articulation have yielded inconsistent results. Several have successfully demonstrated effects of lexical manipulations on durations (Balota, Boland, Shields, 1989; Buz & Jaeger, 2015; Gahl, Yao, & Johnson, 2012; Kello, 2004; Kello et al., 2000). For instance, Kello et al. (2000) found that under time pressure, Stroop interference not only yielded increased latencies but also caused speakers to lengthen the duration of target words. However, other studies have failed to detect such interactions (Damian, 2003; see Heller & Goldrick, 2015, for null results with vowel durations). Indeed, Damian (2003) attempted a direct replication of Kello et al.'s findings, but failed to replicate their results. Furthermore, in two other paradigms, he failed to find any evidence that speakers increase durations under conditions that disrupt lexical access.

In study 2, we aim to reconcile these inconsistent findings in two ways. First, we follow Kello and colleagues in proposing that the timing of response initiation and therefore the extent of temporal overlap between planning and articulatory processes is not fixed (Kello, 2004; Kello et al., 2000). Instead, it may vary across and within individuals, depending on task demands and differences in cognitive processing. If interactive effects arise (at least in part) due to temporal overlap between planning and articulation, then such variation in response initiation might prevent group-level interactive effects on word durations from being observed. Similarly, we note that individual variation in sensitivity to lexical manipulations may also mask the presence of long distance interactions in the production system. If some portion of a participant sample fails to experience much lexical co-activation in response to an experimental manipulation, then by definition we cannot expect to detect related effects of such co-activation on articulation. In the next section, we explain how we empirically tested these proposals.

**1.3.1 Outline of study 2**

In this study, we began by analyzing word duration data gathered during study 1. To avoid contamination from the executive function treatments in that study, we utilized the data from pretest blocks only. Experiment 1 examined the data from all participants who completed the continuous picture naming task (Howard et al., 2006), while experiment 2 examined the data from the continuous classification task (Belke, 2013). As reviewed above, the sequential presentation of pictures from the same semantic category is known to induce co-activation of targets' semantic neighbors; this neighbor co-activation causes semantic interference during naming and semantic facilitation during classification. By testing for lexical effects on word durations when lexical co-activation hinders vs. helps lexical selection, we hoped to constrain the conditions necessary for observing such long distance interactive effects.

Experiment 3 explored word duration data gathered during a related paradigm called blocked-cyclic naming (Damian, Vigliocco, & Levelt, 2001; Oppenheim, in prep). In this task, participants name blocks of pictures that are either blocked by semantic category or mixed. Within each block, they encounter a limited set of semantically related or unrelated items, respectively, which are repeated in random orders across several cycles of naming. Numerous studies have shown slower RTs in the semantically blocked compared to mixed contexts, indicating (non-cumulative) semantic interference (e.g., Breining, Nozari, Rapp, 2015; Crowther & Martin, 2014; Navarrete, Del Prato, Mahon, 2012). Interestingly, the repeated appearance of each block's limited response set arguably allows participants to engage cognitive control to suppress non-target responses (Belke & Stielow, 2013). Therefore, analysis of word durations from this paradigm allows us to test for interactive effects in the context of top-down processes, helping us to further constrain their scope.

Across these three experiments, we constructed statistical models that explicitly accounted for inter- and intra-individual variation in the timing of response initiation, as well as inter-individual variation in sensitivity to lexical selection difficulties. In terms of response timing, we included a by-participant predictor of overall response speed, to capture participants' general tendencies toward earlier vs. later response initiation. If the timing of response initiation influences the pace of articulation, then we expect a main effect of overall speed on word durations. We also included a trial-level RT predictor, to capture any finer-grained correspondences between response timing and articulation. If real-time adjustments to participants' response decision criteria impact their articulations, then we predict a relationship will emerge between trial-level RTs and durations.

In terms of response selection difficulty, we included a main effect of ordinal position (experiments 1 and 2) or semantic context (experiment 3) to test whether the semantic neighbor effects impacted articulation overall. If we observe semantic interference and facilitation in durations when RT is a covariate in the model, then we can infer that lexical co-activation directly influenced articulation after response initiation. We also created a by-participant predictor for the size of semantic neighbor effects on RTs, in order to accommodate variation in participants' sensitivity to lexical disruptions. Because this predictor reflects the dynamics of planning processes, any effect of it on word durations would reflect a truly long distance interaction via cascade.

To foreshadow our findings, we observed an array of results that support the hypothesis of a flexibly interactive production system. In all three experiments, we observed a positive effect of overall response speed, such that slower responders (i.e., those with longer mean RTs) tended to produce longer utterances than faster responders. This result suggests that speakers set a common pace for all aspects of production processing, including the peripheral process of articulation. Second, we found significant effects of trial-level RT in experiments 2 and 3, albeit in different directions. While a negative relationship emerged during continuous naming and classification, such that faster responses had longer durations, a positive relationship was found during blocked cyclic naming. We explain the contrasting direction of these effects as a by-product of methodological differences across the two experiments. More importantly, the observation of any significant effects of trial-level RT on word durations reinforces the idea that flexible adjustments to the timing of response initiation contribute to articulatory variation.

With respect to response selection difficulty, the results revealed an overall effect of semantic neighbor co-activation only within experiment 1: semantic interference caused longer word durations across ordinal positions within a category, parallel to the incremental increase observed in RTs (though smaller in size). This finding implies a direct interaction between lexical selection and articulation. Finally, a subtler interactive effect emerged in experiment 3: an interaction between semantic context and RT interference size revealed that semantic interference affected word durations only among participants who showed large lexical disruptions in RTs. This finding suggests that lexical co-activation can directly influence articulation, but only when it is intense enough to require additional processing after response initiation. Furthermore, it demonstrates the utility of an individual variation approach for increasing the sensitivity of group-level analyses.

### 1.4 Investigation of Semantic Memory Representations Informs Production Theories

In the final study of this dissertation, we explore the interface of spoken production and its lead-in process, conceptualization. Specifically, we investigate the structure of the semantic memory representations that this process operates over, asking whether they occupy a one-dimensional space or are organized in a hierarchical structure. In other words, we ask whether semantic relatedness is represented as a continuum of stronger and weaker relationships, or whether sets of related items cluster together to form supracategories. Although conceptualization arguably constitutes the first step in spoken word production (e.g., Levelt et al., 1999), it typically receives short shrift in the literature as a peripheral process, much like articulation. Production theories frequently make simplifying assumptions about the structure of semantic representations (Dell, 1986; Levelt et al., 1999), or they remain agnostic (Howard et al., 2006; Oppenheim et al., 2010).

However, the prominence of semantic relatedness manipulations as a tool for examining production processes suggests that these representations deserve further scrutiny.

With such background in mind, Alario and del Prado Martin (2010) conducted post hoc analyses of Howard et al.'s (2006) continuous picture naming study, to test whether evidence of semantic complexity lay hidden within the data. The authors had noticed that many of the semantic categories constructed for that study shared common features (e.g., *farm animals* and *zoo animals*), as if they belonged to higher level supracategories (*mammals*). This observation led them to wonder if such hierarchical structure is encoded in semantic memory representations and, if so, whether it has detectable consequences on semantic neighbor effects. To test this hypothesis, Alario and del Prado Martin employed a then state-of-the-art statistical technique: linear mixed effects regression analysis (e.g., Baayen, 2008). This approach not only revealed that the size of semantic interference varied across categories, but also that that variation reflected an underlying supracategory structure. Concretely, the authors found independent effects of an item's ordinal position within its category and a category's ordinal position within its supracategory. They interpreted these results as evidence that semantic interference accumulated across related co-categories, as well as within them. Our final study attempts to replicate this evidence using different stimuli and participants and to extend the findings to continuous classification.

### 1.4.1 Outline of study 3

Once again, we utilized the pretest data from study 1 for additional post hoc analyses. Following Alario and del Prado Martin (2010), we implemented a two-step procedure in order to test for supracategory effects. First, we constructed baseline models of the pretest continuous naming and classification data, before adding in random effects for semantic category. We expected that if any supracategory structure was hidden within our stimulus set, then inclusion of these random effects would capture previously unexplained variance and improve the models' fits. Second, we explicitly built supracategory structure into the models, based on our subjective identification of co-categories within the set. If semantic neighbor effects accumulated both within and across co-categories, then we expected to observe two additive ordinal position effects like Alario and del Prado Martin.

In the end, study 3 failed to replicate the original study: we observed no evidence of supracategory structure within our continuous naming and classification data. Inclusion of random slopes revealed no significant variation in semantic neighbor effects across categories, and the supracategory predictor indicated no accumulation of those effects across co-categories. These divergent results most likely arose because of methodological differences between our study and Alario and del Prado Martin's.

Despite these null results, other subtle but evocative findings emerged from these analyses. In both the picture naming and classification data, inclusion of a random intercept did improve model fit. This variation in mean RT indicates that not all semantic categories are equal in terms of accessibility. Intriguingly, within only the classification data, the random intercepts and random ordinal position slopes were significantly correlated, such that categories with slower mean RTs tended to shower greater semantic facilitation. These data suggest that semantic facilitation—and not semantic interference—is somewhat time sensitive, a conclusion that has implications for theories of cumulative neighbor effects.

## 1.5 Conclusion

Across three empirical studies, this dissertation pushes the boundaries of research on spoken word production. We look beyond the core production processes of lexical selection, phonological encoding, and phonetic encoding, asking how the architecture and mechanisms of seemingly separate or peripheral cognitive processes contribute to the timing and acoustic attributes of spoken words. In study 1, we discover that domain-general processes can support the resolution of linguistic conflict, but that their recruitment is not automatic. In study 2, we unveil a complex interactive relationship between lexical selection and articulation, which is highly sensitive to variations in cognitive processing across and within individual speakers. Finally, study 3 demonstrates that the structure and accessibility of semantic representations can modulate the behavioral impact of semantic neighbor co-activation. Taken together, these studies draw our attention to the wide range of cognitive representations and processes that can influence spoken production, while simultaneously constraining the necessary conditions for observing such interactions. The data presented in the following chapters therefore challenge us to adopt a broad view of the production system and, indeed, of language processing in general.

CHAPTER 2

## 2.1 Introduction to Study 1

Although most speakers converse fluently with little effort, speech production is an impressive cognitive feat. It involves several stages of complex cognitive processing, including conceptualization of the intended message (e.g., activating the semantic features *small, furry, pet,* and *feline* in response to a picture of a cat), retrieval of a word-level representation to convey that message (*lexical selection;* e.g., choosing CAT from an array of active lexical candidates such as CAT, DOG and RAT), and encoding and articulation of the sound structure of the selected word (e.g., activating the phonemes /k/-/ae/-/t/ and producing the appropriate speech gestures; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Crucially, each stage of processing is susceptible to conflict, due to the co-activation of multiple representations. For example, lexical selection becomes more difficult when semantically related lexical representations, often referred to as *semantic neighbors*, are highly co-active and compete for selection (e.g., retrieving CAT when DOG was just named; Wheeldon & Monsell, 1994). The current study tests whether several specific, domain-general cognitive control processes, also known as *executive functions*, play a role in managing such lexical conflict.

Executive functions facilitate thoughts and actions by directing attention and/or cognitive resources towards the task(s) at hand (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). According to Miyake and colleagues' prominent theory, there are three primary executive functions: *updating* of the representations held in working memory, *switching* between different tasks or operations, and *inhibiting* non-target representations. While multiple executive functions may factor into spoken production (Shao, Roelofs, &

Meyer, 2012), we follow a growing body of research in positing that domain-general inhibition may be instrumental, specifically in the resolution of lexical conflict induced by the activation of semantic neighbors (Belke, & Stielow, 2013; Crowther & Martin, 2014; de Zubicaray, McMahon, & Howard, 2015; de Zubicaray, Wilson, McMahon, & Muthiah, 2001; Kan & Thompson-Schill, 2004; Ries, Kazmark, Navarrete, Knight, & Dronkers, 2015; Schnur, Schwartz, Brecher, & Hodgson, 2006; Schnur, Schwartz, Kimberg, Hirschorn, Coslett, & Thompson-Schill, 2009; Shao, Meyer, Acheson, & Roelofs, 2014; Shao, Meyer, Roelofs, 2013; see Novick, Trueswell, & Thompson-Schill, 2010 for a review; but see Alario, Ziegler, Massal, and de Cara, 2012).

Previous psycholinguistic studies testing this hypothesis in individuals without neurological impairments have relied primarily on correlational evidence (Alario et al., 2012; Crowther & Martin, 2014; Shao et al., 2012; Shao et al., 2013). With the exception of Alario et al. (2012), these studies have provided evidence that some aspects of participants' inhibition abilities correlate with their spoken production performance (see below for more detail). While these findings suggest that domain-general inhibitory control and production processes are related, they do not provide conclusive evidence that such correlations reflect a causal relationship between executive functions and the resolution of lexical conflict. We extend this previous work using a relatively new approach called the negative transfer paradigm (Persson, Larsson, & Reuter-Lorenz, 2013; Persson, Welsh, Jonides, & Reuter-Lorenz, 2007) to try and establish such a causal relationship.

The negative transfer paradigm was developed based on the assumption that repeatedly engaging in a task decreases participants' ability or likelihood of using the executive functions associated with that task. If two tasks recruit a common executive function, then intensively engaging in one should reduce performance on the other, i.e., it should generate *negative transfer.* However, if the first task involves an executive function not required by the second task, then no transfer should occur. This type of dissociation is taken as evidence that negative transfer is process-specific, as opposed to reflecting general cognitive fatigue.

Persson et al. (2007) utilized this approach to test whether two different inhibitory executive functions, *proactive interference resolution* and *response inhibition*, are involved in semantic memory retrieval. Friedman and Miyake (2004) had previously argued that these inhibitory executive functions are statistically and functionally distinct on the basis of large-scale latent variable analyses. Interference resolution supports task performance by biasing activation away from previously but no longer relevant representations. For instance, in a list recall task, interference resolution might channel activation away from list A representations, so as to reduce interference during recall of list B. Response inhibition aids performance by suppressing pre-potent but incorrect response candidates. The Stroop task is a classic paradigm for engaging response inhibition; participants' automatic response of reading written words interferes with their ability to name the text color, requiring inhibition for target production (Stroop, 1935). Using the negative transfer paradigm, Persson and colleagues showed that intensive engagement with an interference resolution task induced a performance decrement on a subsequent semantic memory task, while engagement with a response inhibition task did not.

In the current study, we investigated whether these same inhibitory executive functions contribute to the resolution of conflict during lexical selection. We employed the continuous picture naming paradigm (Howard, Nickels, Coltheart, Cole-Virtue, 2006) to generate conflict among semantic neighbors. This task elicits a well-established effect known as cumulative semantic interference: as speakers perform the simple task of naming one picture after another, response times (RTs) increase linearly with each subsequent naming of a given semantic category (Belke, 2013; Belke & Stielow, 2013; Howard, et al., 2006; Navarrete, Mahon, Caramazza, 2010; Oppenheim, Dell, Schwartz, 2010; Runnqvist, Alario, Strijkers, & Costa, 2012; Schnur, 2014). While there remains some debate over the origin of this effect (conceptual vs. lexical; Belke, 2013), we follow Oppenheim et al. (2010) and others in attributing it to conflict among co-active lexical representations.

Under the hypothesis that domain-general control processes are recruited to help resolve lexical conflict, we expect to observe negative transfer from inhibitory executive function tasks to continuous picture naming. Specifically, we expect negative transfer to manifest as increased cumulative semantic interference during continuous naming. On the other hand, if the domain-general hypothesis is incorrect and mechanisms within the language domain are sufficient for resolving lexical conflict, then we predict no modulations of semantic interference following intensive engagement with executive function tasks.

As a control, we also tested for transfer from the target executive functions to a second production task where lexical conflict is absent: continuous picture classification. Participants in this task are required to classify a sequence of pictured objects as natural vs.

manmade (Morrison, Ellis, & Quinlan, 1992). Critically, the inverse of cumulative semantic interference is observed: RTs decrease linearly with each subsequent classification of a given semantic category (Belke, 2013). Assuming that domain-general executive functions are recruited by the production system specifically for conflict resolution, then we expect no transfer effects from inhibition tasks to picture classification.

In the remainder of this section, we review the literature that informed this investigation. First, we summarize previous behavioral studies of the relationship between inhibition and lexical selection. We then revisit the logic of the negative transfer paradigm and discuss the mechanisms thought to underlie negative transfer effects. Finally, we outline the specific tasks used here to index interference resolution and response inhibition, providing more concrete predictions for the four experiments presented below.

### 2.1.1 Previous evidence that inhibition resolves lexical conflict

As noted above, previous psycholinguistic studies investigating the role of inhibition in lexical selection have relied primarily on correlational analyses. For instance, Shao et al. (2012) tested for a relationship between speakers' inhibition abilities and their picture naming speeds. They indexed inhibition abilities using the stop-signal task (Logan & Cowan, 1984). During this task, participants are trained to provide a habitual response (e.g., push the correct button to indicate an arrow's direction), then asked to withhold their responses on a subset of trials when an auditory "stop-signal" is played. Shao and colleagues observed a significant correlation between participants' stop-signal RTs and their mean object and action naming RTs. Stop-signal RTs also correlated with the shapes of those naming distributions. The authors concluded that inhibition plays a role even in basic spoken word production, where lexical conflict should be minimal.

Shao et al. (2013) performed a similar analysis, this time manipulating the level of lexical conflict during production using a picture-word interference paradigm. In this task, participants name pictures while ignoring superimposed words. Responses are reliably slower when the distractor word is semantically related to the pictured object than when it is unrelated (Rosinski, Golinkoff, & Kukish, 1975). The authors replicated their previous finding that participants' stop-signal RTs correlated with their mean picture naming RTs. However, stop-signal performance did not correlate with the magnitude of picture-word interference, nor with the shape of the picture naming distributions, as revealed by Delta plot analyses. Instead, the latter measures of picture-word interference size and the (slowest) Delta plot segments correlated with each other. Shao et al. concluded that nonselective inhibition (indexed by stop-signal RTs) and selective inhibition (indexed by picture-word interference and the Delta plots) have partially separable effects on spoken production.

Crowther and Martin (2014) tested for relationships between three different executive function abilities and production performance in a blocked cyclic picture naming task (where semantically related items occur in blocks; Belke, Meyer, Damian, 2005). In terms of executive functions, they targeted working memory capacity, response inhibition, and proactive interference resolution. Working memory capacity reflects the maximum amount of information an individual can hold and manipulate in (working) memory while performing a task. To measure this capacity, the authors used a word span task (e.g., Baddeley & Hitch, 1974), which required participants to repeat back auditorily presented lists of words in the correct order and without omission. As explained earlier, response inhibition allows individuals to override automatic or dominant responses, while proactive interference resolution helps them ignore previously (but no longer) relevant information. Response inhibition was indexed using the traditional color-word Stroop task, while interference resolution ability was measured using a recent negatives task (Monsell, 1978). In the latter task, participants had to report if a probe word was present in a preceding array. Crucially, on half of the rejection trials ("recent negative" trials), the probe had appeared in the immediately preceding list, inducing proactive interference. Finally, the blocked cyclic naming task required participants to name pictures in semantically homogeneous vs. heterogeneous blocks. This paradigm is known to elicit (non-cumulative) semantic interference, such that RTs are slower in homogeneous than heterogeneous blocks, presumably because of neighbor co-activation (e.g., Belke et al., 2005). Within a given block, a closed set of items is presented repeatedly across four different cycles, with each item appearing once per cycle. We note that the use of a limited response set in this task is thought to increase speakers' ability to engage inhibition, because they can apply it specifically to the non-target members of the set (Belke & Stielow, 2013). This attribute of blocked cyclic naming will factor into our discussion of the current study. Furthermore, this design allowed Crowther and Martin to measure several effects for each block type: the increase of RTs across trials, the change in RTs across cycles within a block, and the change in RTs across blocks.

While Crowther and Martin (2014) reported a number of significant correlations, we note those that suggest a role for inhibitory executive functions in naming. Specifically, the size of participants' Stroop interference effects correlated reliably with the growth of RTs across cycles in semantically homogeneous blocks. In other words, Stroop interference correlated with semantic interference. The authors interpreted this as evidence that response inhibition is required for managing semantic neighbor co-activation. Interestingly, the magnitude of participants' proactive interference effects correlated reliably with the decrease in RTs across cycles in semantically heterogeneous blocks. In other terms, proactive interference correlated with repetition *priming*. This novel piece of evidence, which indicates a relationship between an inhibitory executive function and a facilitation effect during production, will be echoed in the results below.

A final study to employ the correlational approach was conducted by Alario et al. (2012), who probed the relationship between inhibitory control and naming performance in school-age children (ages 7-10). Response inhibition abilities were indexed using a Simon task (Simon, 1990). When a colored disc was presented to the right or left of a fixation cross, participants had to indicate the color of the disc with a right or left button press, while ignoring the disc's location. Production processes were again measured using blocked cyclic picture naming. Ultimately, Alario et al.'s analyses revealed no significant correlations between children's response inhibition abilities and their production performance. This finding lends no support to the hypothesis under investigation here. However, the fact that executive functions and their neural substrates are relatively slow developing compared to other cognitive processes (see Chrysikou, Weber, & Thompson-

Schill, 2014, for discussion) may explain these null results. If children do not have fully developed cognitive control systems, then they may be unable to recruit executive functions during speech production in the same manner as adults.

On the whole, this growing body of research suggests a relationship between inhibition abilities and spoken production, not only under conditions of lexical conflict, but whenever previous experience influences current processing (e.g., the link between inhibition and repetition priming in Crowther & Martin, 2014). This view is strengthened by neuroimaging data showing increased recruitment of the prefrontal cortex (PFC), thought to subserve executive functions, when lexical selection difficulty is manipulated (de Zubicaray et al., 2015; de Zubicaray et al., 2001; Kan & Thompson-Schill, 2004; Schnur, et al., 2009; Shao et al., 2014). Recent neuropsychological studies provide converging evidence supporting this association: patients with selective impairments to the neural network underlying executive functions show increased sensitivity to lexical conflict in certain tasks (Ries et al., 2015; Schnur et al. , 2006). This work is particularly important, because it establishes a causal role for inhibitory executive functions in resolving lexical conflict. The purpose of the current research is to demonstrate that this causal relationship extends to typically functioning speakers as well. To that end, we employ the negative transfer paradigm, which allows us to manipulate the accessibility or utility of specific executive functions.

### 2.1.2 Mechanisms underlying negative transfer

Persson and colleagues (2007; 2013) drew on the resource depletion framework (e.g., Baumeister, Bratslavsky, Muraven, & Tice, 1998) in developing the negative transfer paradigm. This framework stems from the literature on self-regulation, i.e., cognitive processing that supports goal-directed behavior, particularly when the goal is to overcome impulses and avoid problematic activities (e.g., overeating). A key assumption of the resource depletion framework is that self-regulation relies on a pool of limited cognitive resources, which can be diminished by repeated use. This assumption generates the prediction that performing one self-regulatory task may cause a performance decrement on a second self-regulatory task, if the first expends all available resources. In other words, it generates the prediction of negative transfer.

Given the many parallels between research on self-regulation and cognitive control (see Hofman, Schmeichel, & Baddeley, 2012, for a review), Persson and colleagues applied this negative transfer logic to executive functions. Specifically, they devised a pre- vs. posttest design to test whether repeatedly engaging a particular executive function would trigger performance changes in a second task requiring that executive function, regardless of domain. In their first experiment, Persson et al. (2007) tested whether the executive function of interference resolution plays a role in semantic memory retrieval. They used the verb generation task (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) to manipulate the difficulty of semantic retrieval. This task requires participants to generate a verb in response to a written noun. On easy trials, the verb stimulus maps onto a clearly dominant response (e.g, SCISSORS – "cut"); on difficult trials, it evokes a range of possible responses (e.g., BALL – "throw", "kick", "bounce"). RTs are reliably slower in the difficult condition, due to conflict among the competing response candidates. Between the verb generation pre- and posttests, participants underwent an interference resolution treatment: they performed a recent negatives task (see above) for approximately 20 minutes. The authors found that the difference between easy and difficult verb generation trials was reliably increased at posttest, compared to a baseline group who underwent sham treatment.

Critically, Persson et al.'s (2007) experiment 1 finding of negative transfer from interference resolution to semantic retrieval contrasted with the results of experiment 2. Participants again performed the verb generation task at pre- and posttest; however, they were assigned the stop-signal task at treatment, in order to engage the executive function of response inhibition. In this case, the authors observed no reliable increase in the verb generation effect at posttest. They concluded that interference resolution, not response inhibition, plays a selective role in semantic retrieval, and that negative transfer can be process-specific. Continuous engagement with an executive function depletes the resources supporting that specific control process, rather than inducing generalized cognitive fatigue. In their follow-up work, Persson et al. (2013) extended this argument on the basis of neuroimaging data. They observed that following an interference resolution treatment, fMRI scans revealed decreased activity in the region of the prefrontal cortex associated with that executive function, compared to the baseline group. They interpreted these data as evidence of process-specific resource depletion at a neural level.

### 2.1.3 Using negative transfer to investigate the role of inhibitory executive functions in lexical selection

Over the course of four negative transfer experiments, we integrated these two sets of previous findings, using the negative transfer paradigm to test whether interference resolution or response inhibition play causal roles in resolving conflict during lexical selection (see Table 2.1). Experiment 1 constituted a replication of Persson et al. (2007)'s first experiment, except that their verb generation task was replaced with continuous picture naming. The recent negatives task was utilized to engage interference resolution. Treatment participants experienced proactive interference in their intervening task, because a subset of rejection probes had appeared on immediately preceding trials; baseline participants encountered no such recency manipulation and therefore no proactive interference. If interference resolution influences lexical conflict, we expect negative transfer within the treatment group only. Concretely, the size of their semantic interference effects during continuous picture naming should increase at posttest compared to pretest. If a different executive function moderates lexical conflict, or if no domain-general processes are required, we expect no transfer to occur.

Experiment 2 followed the same format as experiment 1, except that the intervening task was chosen to engage the executive function of response inhibition. As explained above, the picture-word interference task (Rosinski et al., 1975) evokes a form of Stroop interference, which arises when the simultaneous presentation of a target and a distractor forces participants to suppress a highly trained response (the reading of written words) in order to provide the less automatic target response (picture naming; MacLeod, 1991). Psycholinguists continue to debate the locus of picture-word interference (within lexical selection vs. in a pre-articulatory response buffer; see Spalek, Damian, & Boelte, 2012, for a review), yet there is some consensus that response inhibition helps adjudicate among the competing responses (e.g., Novick et al., 2010; Shao et al., 2013). We selected this verbal task for the response inhibition treatment in order to maximize the likelihood of significant transfer. If response inhibition is recruited to manage lexical conflict in continuous naming, then treatment participants (but not baseline) should show increased semantic interference at posttest. On the other hand, if a different process is sufficient for handling lexical co-activation, then no such transfer should appear.

| | | Executive Function Treatment | |
|---|---|---|---|
| | | Interference resolution (Recent negatives) | Response inhibition (Picture-word interference) |
| Pre- and Posttest Effect of Semantic Neighbors | Interference (Naming) | Experiment 1<br>Prediction: Increased interference at posttest<br>Result: No effect | Experiment 2<br>Prediction: Increased interference at posttest<br>Result: Posttest slowing, decreased interference overall |
| | Facilitation (Classification) | Experiment 3<br>Prediction: No effect<br>Result: Posttest slowing, decreased facilitation at posttest | Experiment 4<br>Prediction: No effect<br>Result: decreased facilitation at posttest |

Table 2.1 2 x 2 design of negative transfer experiments, with predictions based on the assumption that lexical conflict engages executive functions, which can be depleted.

Experiments 3 and 4 were designed as controls for experiments 1 and 2, respectively. In particular, these experiments were intended to show that any negative transfer to continuous naming emerged because the target executive function helped resolve lexical conflict. To that end, we replaced the continuous picture naming task with continuous picture classification, where the co-activation of semantic neighbors facilitates performance rather than interfering with it. Assuming that interference resolution (experiment 3) and/or response inhibition (experiment 4) are recruited by the production system solely under conditions of conflict, then we expect no posttest change in classification performance after an executive function treatment.

To foreshadow the results, they do not provide support for the hypothesis that the inhibitory executive functions of interference resolution and response inhibition are recruited because of lexical conflict during continuous naming. Indeed, the pattern of results was practically the inverse of what we expected. We present the data from all four experiments together, deferring the majority of discussion until the reader has a full view of the data. Then we consider these results in light of some recent work and situate them within a conflict adaptation framework.

## 2.2 Experiment 1: Negative Transfer from Recent Negatives to Picture Naming

### 2.2.1 Methods

*Participants*

We recruited 60 participants at Northwestern University (NU) using the Linguistics Department subject pool and flyers around campus. Participants from each source received course credit or $10 compensation, respectively. They reported learning no language other than English before age 5 and no history of color blindness or language/cognitive impairment.

*Materials and Design*

*Speech production task: Continuous picture naming.* As in Howard et al. (2006), participants named a sequence of pictures as quickly and accurately as possible. 90 colored line drawings were selected from Rossion and Pourtois' (2004) database, which colorized and normed Snodgrass and Vanderwart's (1980) classic black and white line drawings. The selected items came from 18 categories (5 items each; see Appendix) and had an average word frequency of 65.3 words per million (SUBTLEX database; Brysbaert & New, 2009).

A master stimulus list was created such that 9 categories appeared at pretest and 9 at posttest, thus avoiding item repetition within participants. The 9 categories within a test were subdivided into 3 blocks, where items were drawn from 3 categories in rotation (e.g., a block containing birds, fruits, and vehicles might begin OWL - APPLE - CAR - PEACOCK - ORANGE - PLANE). As a result, items from the same semantic category were presented with a consistent lag of 2 intervening items between them. This short-lag design deviates from Howard et al.'s original study, but subsequent work has demonstrated that it elicits comparable semantic interference effects (Runnqvist et al., 2012; Schnur, 2014). The master stimulus list was then manipulated to create 9 additional versions, which counterbalanced the assignment of items to pre- vs. posttest and ordinal position within a category. Between participants, every item appeared once in each ordinal position at both pre- and posttest. Participants were randomly distributed across the 10 stimulus lists.

Within trials, we followed Howard et al.'s (2006) design. A fixation cross appeared in the center of the screen for 500 ms, followed by a blank interval of 250 ms. The target appeared onscreen and remained visible for 2000 ms, during which time the participant named the item aloud. The screen then blanked again for an inter-trial interval of 500 ms. This task took approximately 6 minutes to complete, collapsing across pre- and posttest.

*Executive function task: Recent negatives.* This paradigm, also called an "item recognition" task, asks participants to hold a set of items in memory and identify whether a subsequent probe belongs to that set. We followed Persson et al. (2007; 2013) in using letter stimuli to engage verbal working memory, directly replicating their design. On each trial, an array of four lowercase letters appeared in a square arrangement around a central fixation cross, remaining onscreen for 1500 ms. The screen blanked for a delay of 3000 ms, before a capital letter probe appeared for 1500 ms in the center of the screen. Participants were instructed to indicate whether the probe belonged to the current trial's set. They did so by pressing "yes" or "no" (buttons 1 and 8) on a Cedrus RB-840 Response Box using their right or left index fingers, respectively. An inter-trial interval of 1500 ms followed before the next trial began. Participants encountered 144 trials total, divided into three 48-trial blocks with automatically timed one-minute breaks in between. Completion of this task took about 20 minutes.

Across blocks, half of all trials required participants to accept the probe as a set member ("yes" trials), and half required rejections ("no" trials). The rejection trials were divided into three types. Low interference rejections (“low no”) were relatively easy, because the probe did not appear in the memory set of the current or two preceding trials. For intermediate interference rejections (“medium no”), the probe was not in the current memory set, but it did appear in the immediately preceding trial. Proactive interference from the previous trial therefore made the probe more difficult to reject. Finally, high interference rejections ("high no") involved probes that appeared in the memory sets of the two preceding trials. We expected proactive interference to be strongest on such trials. Following Persson et al. (2007; 2013), we manipulated rejection type to create baseline and treatment versions of this task. While baseline participants exclusively encountered low interference rejections, treatment participants saw equal numbers of low, medium, and high interference rejections.

We implemented some additional constraints on this design. Within each block, and excluding the letters "i" and "l," each letter from the Roman alphabet appeared as a probe 1-2 times. Across a given stimulus list (one each for the baseline and treatment conditions), every letter appeared in memory sets approximately the same number of times (7-9). Finally, for "yes" trials, the location of the lowercase match for the probe was counterbalanced to appear in each quadrant of the screen an equal number of times (6 per block).

*Procedure*

Participants provided informed consent and completed a language background questionnaire before working through the three-step negative transfer procedure. At pretest, participants performed continuous picture naming to provide a baseline measure of their sensitivity to the semantic interference effect. Next, they performed the recent negatives task, which was intended to engage the executive function of interference resolution. Finally, participants completed a second, posttest round of continuous picture naming, thereby allowing us to test for treatment-specific negative transfer effects. Experimental sessions lasted about 45 minutes total.

*Data processing*

Response times in all spoken tasks were extracted from stereo recordings. These contained audio markers time-locked to stimulus onsets on the right channel and participant speech on the left. After segmenting the recordings at the audio markers, speech onsets were detected using intensity thresholds. Each trial was equalized to an average root mean square intensity of 0.02 Pascal. The Praat Intensity function then estimated the intensity contour of the normalized signal. Speech onsets were located by sampling this contour at 1-millisecond increments to detect when the normalized signal passed a 55 dB threshold. The first author manually corrected these boundaries to avoid incorrect triggers due to lip smacks, breathing, and/or low amplitude onsets.

*Analysis*

All linear mixed effects models (Baayen, Davidson, & Bates, 2008; R, package lme4, Bates, Maechler, Bolker, & Walker, 2009) were constructed in accordance with the recommendations of Bates, Kliegl, Vasisth, and Baayen (2015). First, categorical variables were contrast coded (e.g., the baseline condition was assigned the value -0.5 and the treatment condition was 0.5). Next, all variables were centered by subtracting the grand mean from each individual value, thereby avoiding correlations between the variable coefficients and the model intercept. Pair-wise correlations among all fixed effects were also considered in order to avoid issues of multicollinearity; no pair-wise correlation exceeded the threshold of $r$=0.3.

Model selection began with construction of a maximal model, including random by-participant and by-item intercepts, random by-participant and by-item slopes for all appropriate variables (e.g., no by-participant slope for a variable manipulated between participants), and correlations among these random effects. If the maximal model failed to converge, random correlations were eliminated; in all cases, this allowed the model to converge. A Principle Components Analysis (PCA; R, package RePsychLing, function rePCA) was then run on the model structure to test for overparameterization, which can impact interpretability (Bates et al., 2015). The PCA reveals how much variance is captured by each random effect. If any of the by-participant or by-item effects explained a near-zero amount of variance, they were removed.

Once the PCA no longer indicated overparameterization, the model was further simplified in a stepwise fashion. Random interaction slopes were removed in order of complexity (i.e., all three-way interactions removed before any two-ways interactions), until nested model comparison indicated a significant decrease in overall model fit. At that point, correlations among the remaining random effects were added back into the model. These were retained only if model comparison indicted a significant improvement in fit.

### 2.2.2 Results

*Recent negatives*

*Model structure.* This analysis modeled button press RTs among treatment participants. Fixed effects included response type ("yes"=-0.5 vs. "no"=0.5), rejection difficulty ("high lo"=0.5 vs. "low no"=-0.5), block number (1-3), and all possible interactions. Random effects included intercepts for both subjects and items, by-subject slopes for block and response type, plus correlations among all random effects.

| | *Block 1* | | *Block 2* | | *Block 3* | |
|---|---|---|---|---|---|---|
| | Treatment | Baseline | Treatment | Baseline | Treatment | Baseline |
| *Yes* | 873 (8) | 864 (9) | 874 (8) | 859 (9) | 888 (10) | 884 (10) |
| *Low no* | 876 (15) | 890 (9) | 856 (15) | 900 (10) | 907 (17) | 918 (10) |
| *High no* | 980 (16) | - | 896 (14) | - | 935 (16) | - |

Table 2.2 Mean (and standard error) RTs for Experiment 1 recent negatives task.

*Results.* The results confirmed that the recent negatives task elicited proactive interference from participants in the treatment condition. They made slower button presses when rejecting probes on "high no" trials than on "low no" trials ($\beta$=0.06, se=0.01, $\chi^2(1)$=31.76, p<0.001). This difference decreased in magnitude after the first block (rejection type x block; $\beta$=-0.04, se=0.01, $\chi^2(1)$=7.06, p<0.01), but it remained robust. Treatment participants also showed numerically smaller RTs on the easy acceptance ("yes") trials compared to baseline participants (see Table 2.2), suggesting that proactive interference slowed processing in general. These findings support the argument that the recent negatives treatment engaged the target executive function, interference resolution.

*Continuous picture naming*

Initial filler trials from each block of continuous naming were excluded from analysis. Removal of errors (including incorrect responses, non-canonical responses, dysfluencies, and technical errors) eliminated 9.8% of the data. RTs below 200 ms and above 2000 ms were also removed, as were RTs more than three standard deviations from a given subject's mean (1.4%).

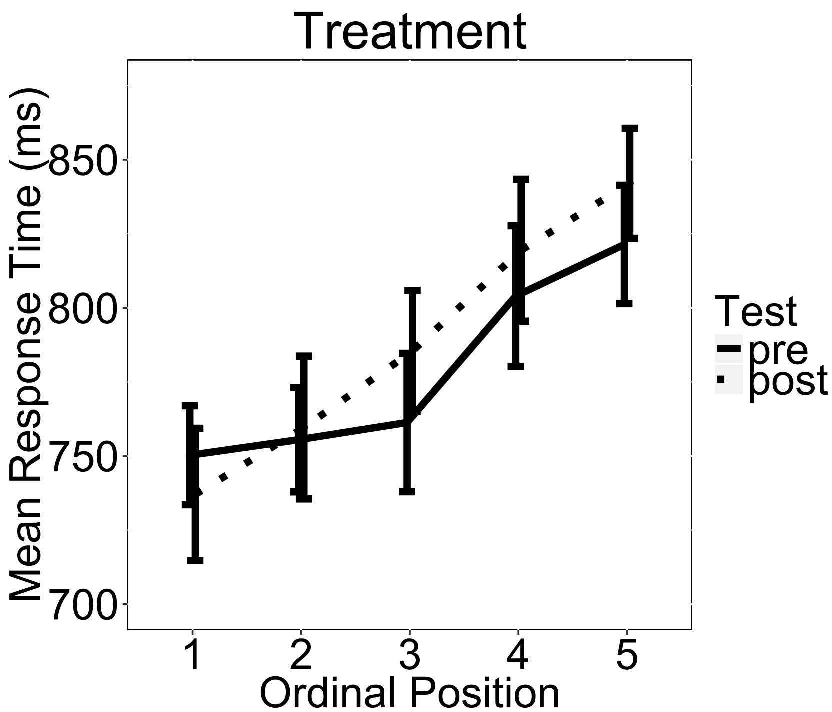
*Model structure.* Three models were constructed to test for effects of the interference resolution treatment on picture naming performance. The first model included the data from all 60 participants (i.e., baseline and treatment) who completed the naming task. Fixed effects of interest included condition (baseline=-0.5, treatment=0.5), ordinal position within a category (1-5), block within a test session (1-3), test (pre=-0.5, post=0.5), and all possible interactions among these variables. The variable class (manmade=-0.5, natural=0.5) and its interactions were also included as a control. The random effects structure included by-participant and by-item intercepts; by-participant slopes for block, test, class, and the block by test interaction; by-item slopes for condition, ordinal position, and test; plus correlations among these random effects.

To unpack effects observed in the first model, subset models of each condition were constructed. For the treatment model, the fixed effects were identical to the main model, except that condition was replaced with a new variable: proactive interference size. To derive this variable, a simple linear regression was built for each treatment participant. The dependent variable was rejection RT in the recent negatives task, with rejection type ("low no" vs. "high no") as the single predictor. The beta estimate was then extracted to index that participant's proactive interference effect size. When added to the treatment subset model, this variable was allowed to interact with all other effects of interest. Random effects included by-participant and by-item intercepts; by-participant slopes for test and the test by block interaction; and a by-item slope for ordinal position.

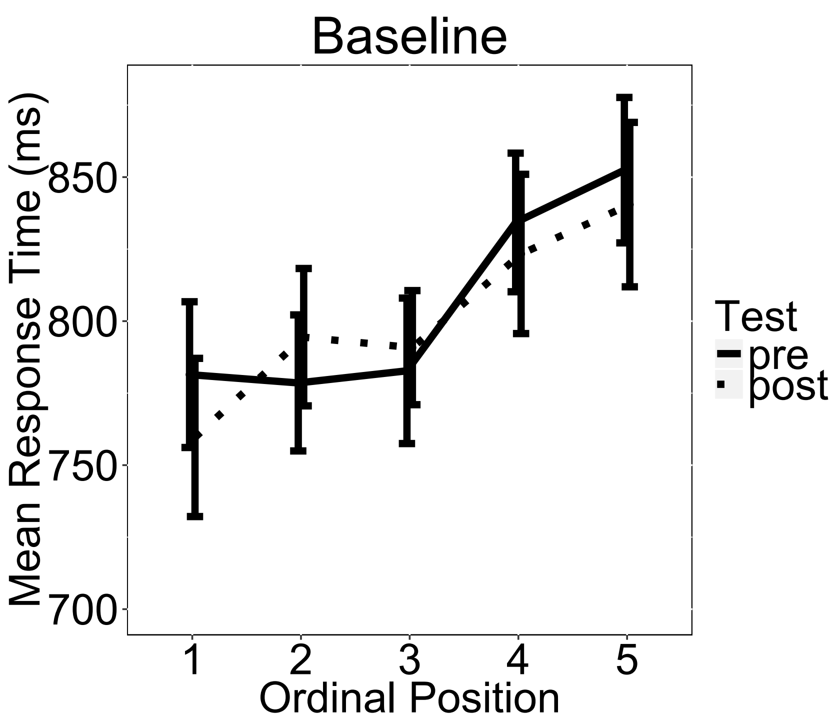
For the baseline model, the fixed effects were again identical to the main model, except that the condition variable and its interactions were simply removed. Random effects included by-participant and by-item intercepts; by-participant slopes for block, test, class, and the block by test interaction; by-item slopes for ordinal position and test; and correlations among all random effects.

*Transfer effects: No increase in interference at post-test.* The results of the full model replicated the standard finding of semantic interference: participants responded slower each time they named an item from the same semantic category ($\beta=0.03$, se=0.003, $\chi^2(1)=56.02$, $p<0.001$; fig. 2.1). There was also an effect of block, indicating that participants' responses became slower across blocks ($\beta=0.02$, se=0.004, $\chi^2(1)=16.25$, $p<0.001$). A significant interaction between these two variables revealed that the magnitude of semantic interference decreased across blocks within a test ($\beta=-0.02$, se=0.003, $\chi^2(1)=24.71$, $p<0.001$).

The only reliable effect of condition (baseline vs. treatment) was a three-way interaction between condition, block, and test ($\beta=-0.04$, se=0.02, $\chi^2(1)=4.27$, $p<0.05$). This interaction reflects variation in the general block effect. It appears that treatment participants got slightly slower across blocks at both pre- and posttest; baseline participants only showed this pattern at pretest. All other main effects and interactions failed to reach significance ($\chi^2s(1)\leq2.46$, $p\geq0.11$). Critically, condition did not interact with ordinal position, providing no evidence that the recent negatives treatment impacted semantic interference, i.e., providing no evidence of negative transfer.




(A)




(B)

Figure 2.1 Experiment 1 RTs increased with each category member, indicating semantic interference. Interference size (slope) remained constant across test and condition.

Subset models helped verify the claim that the interaction of ordinal position and test did not differ across the baseline and treatment conditions. Consistent with the main model, the treatment model showed overall effects of ordinal position ($\beta=0.03$, se=0.004, $\chi^2(1)=50.74$, $p<0.001$) and block ($\beta=0.02$, se=0.006, $\chi^2(1)=7.13$, $p<0.01$), plus a significant interaction between these variables ($\beta=-0.01$, se=0.004, $\chi^2(1)=12.00$, $p<0.001$). Proactive interference size had no impact on the results; neither the overall effect nor any of its interactions reached significance ($\chi^2(1)\leq1.34$, $p\geq0.24$).

The baseline model also confirmed effects of ordinal position ($\beta=0.02$, se=0.004, $\chi^2(1)=35.71$, $p<0.001$), block ($\beta=0.02$, se=0.008, $\chi^2(1)=9.15$, $p<0.01$), and their interaction ($\beta=-0.02$, se=0.004, $\chi^2(1)=18.21$, $p<0.001$). In addition, they showed a significant two-way interaction between block and test ($\beta=0.03$, se=0.02, $\chi^2(1)=4.42$, $p<0.05$). The presence of this final interaction in the baseline condition only confirms the three-way interaction of condition, block, and test in the full model.

*No correlation between strength of interference effects across tasks.* Although the absence of transfer effects from the recent negatives task to picture naming suggests that interference resolution is not shared across tasks, we sought to strengthen this conclusion using correlational analysis. Specifically, we tested the correlation between proactive interference and semantic interference effect sizes in the treatment condition. In order to have a measure of semantic interference that was not contaminated by the executive function treatment, we assessed participants' semantic interference at pretest only. Both effect size measures were calculated using by-participant simple linear regressions, from which beta estimates were extracted. For the proactive interference measure, button press

RTs were predicted by rejection type; for the semantic interference measure, vocal RTs were predicted by ordinal position within a category. The result of this analysis corroborates the regression analyses, showing a non-significant correlation between effect sizes in the two tasks ($r$=0.16, n=30, p=0.41).

### 2.2.3 Summary

Experiment 1 successfully replicated previous behavioral findings of semantic interference during continuous picture naming (e.g., Howard et al., 2006) and proactive interference during a recent negatives task (e.g., Persson et al., 2007). However, we found no evidence of selective negative transfer from the recent negatives task to picture naming. We also observed no significant correlation between participants' performance on the two tasks. In short, experiment 1 provided no evidence that domain-general interference resolution plays a role, causal or otherwise, in managing lexical conflict during continuous picture naming.

## 2.3 Experiment 2: Negative Transfer from Picture-Word Interference to Picture Naming

Because negative transfer effects are assumed to be process-specific, the possibility remains that the second inhibitory executive function, response inhibition, helps resolve lexical conflict. To test this hypothesis, the picture-word interference task was placed between two sessions of continuous picture naming. If response inhibition is shared across the picture-word and continuous naming tasks, then we expect negative transfer, i.e., greater semantic interference after treatment. Unlike in experiment 1, we did not compare the performance of experiment 2 treatment participants to a baseline group who performed a reduced conflict version of that treatment. Instead, we compared treatment participants from experiments 1 and 2 in a combined analysis. If the response inhibition treatment in experiment 2 induced negative transfer, contra the results of experiment 1, then we expect a significant interaction between ordinal position, test, and experiment. We present the results of this combined analysis first, before examining the results within experiment 2, in order to parallel the analyses from experiment 1. This approach also ensures that we include the "control" condition (experiment 1) in our first pass analysis, before zoning in on the data from this new treatment condition.

### 2.3.1 Methods

*Participants*

Thirty additional participants were recruited from the NU Linguistics Department subject pool or via flyer. They received course credit or $10 compensation for their time, respectively. As above, participants learned no language other than English before age 5, and they reported no relevant cognitive impairments.

*Materials and Design*

*Speech production task.* The continuous picture naming materials used here were identical to experiment 1.

*Executive function task: Picture-word interference.* In this paradigm, participants were required to name pictures while ignoring superimposed distractor words, which were either semantically related or unrelated to the target (Rosinski et al., 1975). Many parameters of this task have been varied in previous work, including the time between the target and distractor onsets (stimulus onset asynchrony/SOA; e.g., Schriefers, Meyer, &

Levelt, 1990), the modality of the distractor (written word vs. picture; Damian & Bowers, 2003), the nature of the relationship between the target and related distractors (e.g., semantic & phonological in Schriefers et al., 1990; categorical & "has-a" semantic relationships in Costa, Alario, Caramazza, 2005), and finer-grained characteristics of the distractors (lexical frequency; Miozzo & Caramazza, 2003). We strove to choose task parameters that would (a) minimize the complexity of the task and (b) maximize the likelihood of engaging the same control process as continuous picture naming (assuming that any control process is engaged by lexical conflict in that task).

All trials began with a 500 ms fixation cross, followed by a 250 ms blank screen. The target picture and distractor word were then presented with a 0 ms SOA, meaning that they appeared simultaneously. All related distractors were in a semantic relationship with the pictured object (e.g., LEG and FOOT are members of the body parts category). Participants had up to 2000 ms to name the picture, at which time the screen blanked for a 1500 ms inter-trial interval.

Stimuli were developed from a set of 69 pairs of semantically related items, which had been normed for name agreement during a previous study in our lab ($n$=20 participants; all items elicited 75% name agreement or higher). Photographs were drawn primarily from the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010), with a few supplements from the internet. The average frequency of the items was 70.3 words per million (SUBTLEX database; Brysbaert & New, 2009). While none of these items appeared in the continuous picture naming task, some belonged to the same semantic categories used in that task[1]. This was unavoidable, given constraints of object imageability, semantic relatedness, etc.

A master stimulus list was created in the following manner. Each of the 69 item pairs was divided into target 1 (e.g., "locker") and target 2 (e.g., "backpack"). Thirty-five of the pairs were randomly assigned to group A, and 34 were assigned to group B. In block 1, the group A pairs were scrambled to create unrelated stimuli, with target 1 items appearing as pictures and target 2 items as words (a picture of a LOCKER with the word "mouse" superimposed; a picture of a CANDY with the word "backpack" superimposed). Group B items stayed in their original combinations to create related stimuli. All 69 stimuli were then pseudo-randomly ordered to prevent more than 3 consecutive trials of the same type. Block 2 was created in a similar fashion, except that target 2 items appeared as pictures (BACKPACK-"dolphin") and target 1 items were words (HAIR-"locker"). In the second half of the picture-word interference task, groups A and B were switched so that every picture would be named a second time in the opposite condition (block 3: LOCKER-"backpack"; block 4: BACKPACK-"locker"). As a result, all targets appeared 4 times total, once in each condition of this 2 x 2 design crossing semantic relatedness (unrelated vs. related) by modality (picture vs. word). A second version of the stimulus list was created by switching the order of the experimental halves, and list assignment was counterbalanced between participants.

[1] Five items were excluded from analysis in experiments 2 and 4 to correct design errors. "Anchor" was removed from PWI because it was a filler in CPN. "Rooster" and "jeep" were removed from PWI and "chicken" and "car" from CPN because of their synonymy.

*Procedure.* The procedure for experiment 2 was the same as experiment 1, except for the single change of the intervening executive function task from recent negatives to picture-word interference.

### 2.3.2 Results

*Picture-word interference*

*Model structure.* Vocal RT during treatment was the dependent measure. Fixed effects included semantic relatedness (unrelated=-0.5 vs. related=0.5), block within half (1-2), and half (1=0.5 vs. 2=-0.5). Random effects included intercepts for both subjects and items; by-subject slopes for block, half, and their interaction; by-item slopes for semantic relatedness, half, and their interaction; plus correlations among all random effects.

| | *Half 1* | | *Half 2* | |
|---|---|---|---|---|
| | Block 1 | Block 2 | Block 3 | Block 4 |
| *Related* | 880 (9) | 880 (8) | 788 (7) | 802 (7) |
| *Unrelated* | 861 (8) | 857 (8) | 786 (7) | 786 (7) |

Table 2.3 Mean (and standard error) RTs for Experiment 2 picture-word interference task.

*Results*. Results confirm the presence of a semantic relatedness effect in the picture-word interference data. Across the board, participants were slower to name the target when the superimposed distractor word was semantically related to the pictured object compared to when it was unrelated ($\beta$=0.02, se=0.003, $\chi^2(1)$=9.03, p<0.01). Responses were also faster overall in the second half of the experiment compared to the first ($\beta$=0.10, se=0.01, $\chi^2(1)$=52.37, p<0.001). This finding most likely indicates repetition priming, because all items appeared in both halves of the task, once in the related condition and once in the unrelated condition. Although that the semantic relatedness effect was numerically diminished at the start of the second half (Table 2.3), the interaction of condition and half failed to reach significance ($\chi^2(1)=0.21$, $p>0.05$). We conclude from these data that the picture-word interference task generated sufficient conflict between semantically related pictures and words to engage the target executive function of pre-potent response inhibition.

*Continuous picture naming*

Initial trials from each block were excluded from analysis. Removal of errors (including incorrect and non-canonical responses, dysfluencies, and technical errors) eliminated 10.7% of the data. RTs below 200 ms and above 2000 ms were also removed, as were RTs more than three standard deviations from a given subject's mean (4.2%).

*Model structure*

Two models were constructed to test for negative transfer effects from the picture-word interference task to continuous picture naming. As explained above, the first analysis combined the continuous naming RTs from experiment 2 with those from the treatment condition in experiment 1 ($n$=60). Fixed effects of interest included ordinal position (1-5), block within test (1-3), test (pre=-0.5 vs. post=0.5), intervening executive function task (recent negatives=-0.5 vs. picture-word interference=0.5), and all possible interactions. Item class (natural=0.5 vs. manmade=-0.5) and its interactions were also included as a control. Random effects included intercepts for both participants and items; by-participant slopes for block, test, class, and the block by test interaction; by-item slopes for ordinal position, test, executive function task, the ordinal position by test interaction, and the test by executive function task interaction.

In the second analysis, the continuous picture naming RTs from the 30 participants in experiment 2 were modeled alone. The fixed effects structure was the same as above, except that the predictor coding intervening executive function task was replaced with a by-participant measure of the amount of picture-word interference in the intervening task (beta estimates from by-participant linear regressions). The random effects structure included intercepts for both participants and items; by-participant slopes for block, test, and class; a by-item slope for test; and correlations among these terms.

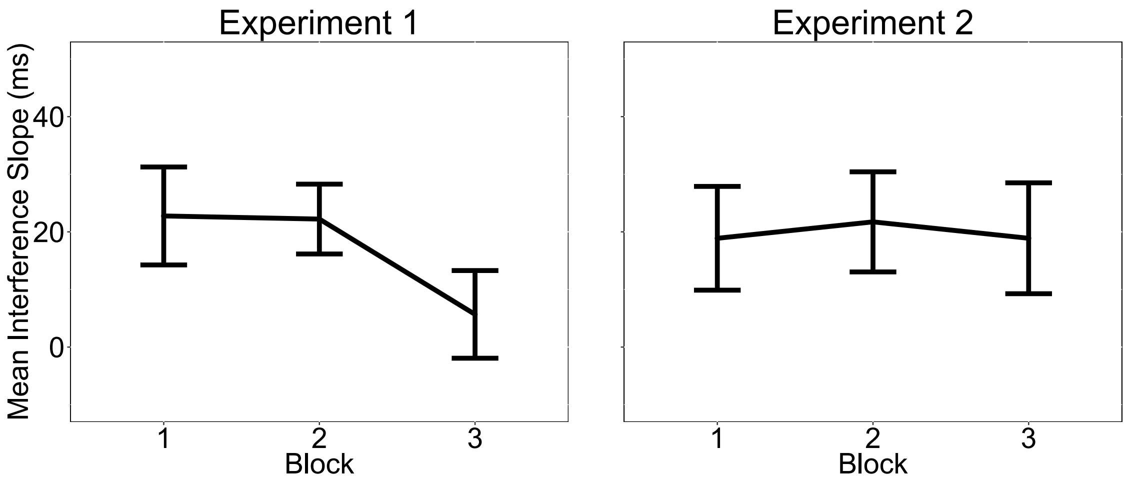
*Comparison of transfer effects in experiments 1 and 2: Positive transfer in experiment 2*

When the treatment participants from both experiments were combined into one analysis, we observed reliable semantic interference ($\beta=0.03$, se=0.003, $\chi^2(1)=59.57$, $p<0.001$), slowing across blocks within a test ($\beta=0.01$, se=0.005, $\chi^2(1)=5.01$, $p<0.05$), and posttest slowing ($\beta=0.02$, se=0.009, $\chi^2(1)=6.38$, $p<0.05$). All treatment participants also showed decreasing semantic interference across blocks within a test ($\beta=-0.02$, se=0.003, $\chi^2(1)=21.22$, $p<0.001$).

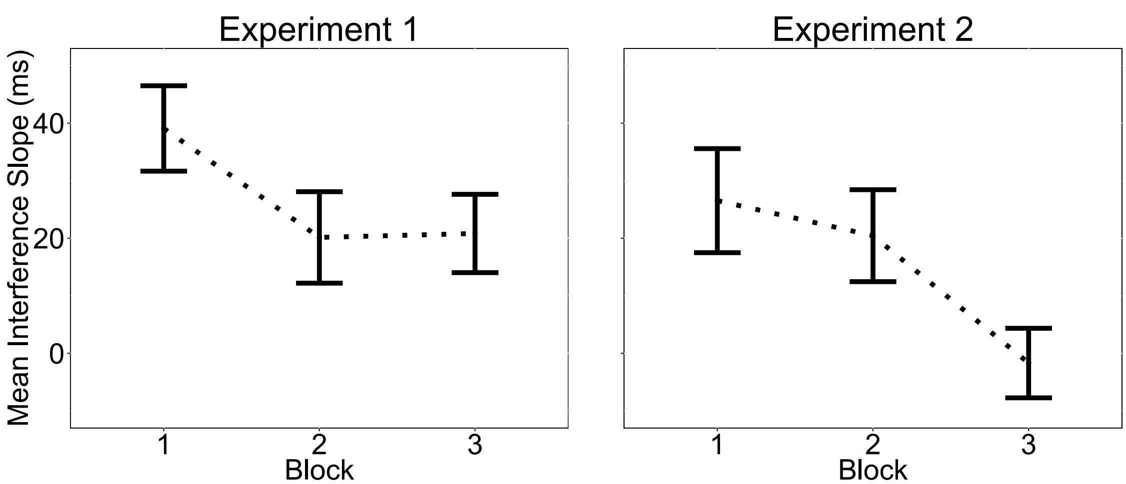
However, two differences emerged across treatment groups[2]. Importantly, a significant interaction of ordinal position with executive function task revealed that the magnitude of semantic interference was significantly smaller in experiment 2 than experiment 1 ($\beta=-0.009$, se=0.004, $\chi^2(1)=4.59$, $p<0.05$; fig. 2.3). The unexpected direction

[2] N.b. a model of pretest data only from both treatment groups confirmed the absence of any preexisting differences between them ($\chi^2(1)\leq2.16$, $p\geq0.14$).

of this effect suggests a form of *positive transfer* within experiment 2, rather than a performance decrement.




(A)




(B)

Figure 2.2 Semantic interference effect size among treatment participants at pretest (A) and posttest (B).

The second difference across experiments was revealed by a marginal four-way interaction between ordinal position, block within test, test, and intervening executive function task ($\beta$=-0.02, se=0.01, $\chi^2(1)$=3.37, p=0.07). Figure 2.2 shows that this interaction is driven by block 3 performance. While experiment 1 treatment participants showed greater semantic interference during block 3 of posttest compared to block 3 of pretest, experiment 2 participants showed the opposite pattern. In other words, participants who engaged interference resolution during the intervening task (experiment 1) persisted in showing semantic interference until the end of posttest. However, participants who engaged response inhibition during the intervening task (experiment 2) showed little semantic interference by the end of posttest. Once again, this pattern hints that the picture-word treatment may have induced positive rather than negative transfer.

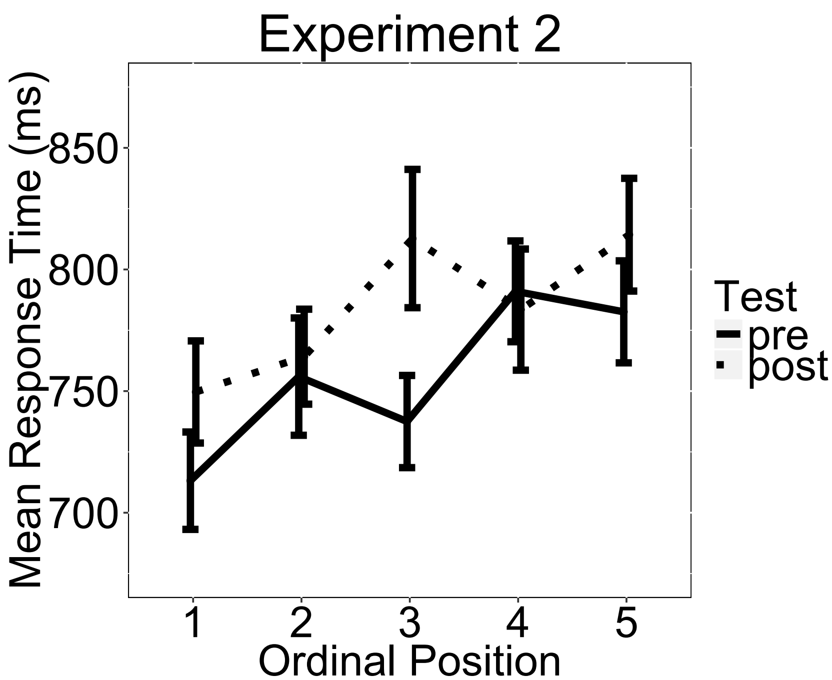



Figure 2.3 Increasing RTs across ordinal positions within category confirm semantic interference in experiment 2. Interference size (slope) remained constant across tests, but RTs slowed overall at posttest.

*Positive transfer effects within experiment 2*

Results within experiment 2 confirmed that participants experienced semantic interference during naming ($\beta$=0.02, se=0.003, $\chi^2(1)$=39.09, $p<0.001$; fig 2.3). Importantly, a reliable effect of test revealed that participants' responses were slower overall at posttest compared to pretest ($\beta$=0.03, se=0.01, $\chi^2(1)$=4.70, $p<0.05$). This finding suggests that completion of the picture-word interference task impacted production processing during posttest. No other main effects were significant ($\chi^2(1)\leq1.28$, $p\geq0.26$).

A significant interaction emerged between ordinal position and block, indicating that semantic interference shrank across blocks within a test ($\beta$=-0.01, se=0.004, $\chi^2(1)$=10.98, $p<0.001$). However, a reliable three-way interaction between ordinal position, block, and test, revealed that this pattern was driven by posttest performance ($\beta$=-0.02, se=0.008, $\chi^2(1)$=7.60, $p<0.01$; fig 2.2). Finally, there was a marginal three-way interaction between block, test, and the magnitude of picture-word interference, suggesting that participants' response inhibition abilities may have modulated their tendency to slow across blocks and tests ($\beta$=.0009, se=0.0005, $\chi^2(1)$=3.01, $p$=0.08). All remaining interactions failed to reach significance ($\chi^2(1)\leq0.98$, $p\geq0.32$).

*No significant correlations between interference effects across tasks*

As in experiment 1, participants' effect sizes in the production and executive function tasks were estimated using beta coefficients from simple linear regressions of their RTs. Ordinal position predicted RTs in the continuous picture naming pretest, while semantic relatedness (related vs. unrelated) predicted picture-word RTs. Although the transfer effects described above suggest a link between these two tasks, no significant correlation was observed (*r*=-0.14, n=30, p=0.47).

### 2.3.3 Summary

Like experiment 1, experiment 2 replicated previous findings of semantic interference during continuous picture naming: participants responded slower each time they encountered a new member of a given semantic category. However, the results of the combined analysis showed that semantic interference was smaller overall when the intervening task was picture-word interference (experiment 2) as opposed to recent negatives (experiment 1). This finding indicates that the response inhibition treatment, but not the interference resolution treatment, benefited continuous picture naming performance. Continuous engagement of response inhibition appears to have improved participants' abilities to resolve conflict during lexical selection.

This interpretation is supported by the subset analysis of the experiment 2 data. While the results did not reveal a posttest change in the magnitude of semantic interference, they showed that the response inhibition treatment impacted performance in two other ways. First, the treatment produced posttest slowing, such that experiment 2 participants were generally slower to respond during the continuous picture naming posttest compared to pretest. Second, it caused the size of semantic interference to fluctuate across posttest blocks: while interference was exceptionally large at the start of posttest, it shrank to near-zero levels by the end. The latter finding in particular suggests that this treatment helped participants resolve competition among co-active neighbors.

## 2.4 Experiment 3: Negative Transfer from Recent Negatives to Picture Classification

As laid out in the introduction, experiments 3 and 4 were designed as controls for experiments 1 and 2, respectively. They aimed to support the claim that transfer effects on continuous naming arise specifically under conditions of conflict, when the production system recruits domain-general executive functions to resolve lexical co-activation. This argument predicts that in the absence of such conflict, no transfer effects should emerge. We therefore replaced the naming task with continuous picture classification, where semantic neighbor co-activation is known to facilitate performance rather than interfering with it. If inhibitory executive functions only influence neighbor effects when conflict is involved, then we expected not to observe transfer effects in experiments 3 or 4. However, to foreshadow the results, they did not support this prediction. A set of significant findings emerged, suggesting that a different account of the treatment effects in experiment 2 must be developed.

### 2.4.1 Methods

*Participants*

Sixty participants were recruited at NU, and they received either course credit or $10 compensation for their time. None had learned any language other than English before age 5, nor reported a history of any relevant cognitive impairment.

*Materials and Design*

*Speech production task: Picture classification.* Borrowing from Belke's (2013) design, we used identical stimuli across the picture naming and classification tasks. This allowed us to rule out item- and category-specific differences across tasks. In fact, all trial parameters were identical: we utilized the same experimental scripts for naming and classification, changing only the instructions. Here, participants were instructed to verbally classify objects as natural vs. man-made as quickly and accurately as possible.

We note that because of the fixed 2-lag design, responses were predictable within block (e.g., a sequence of birds, fruits, and vehicles would receive the responses NATURAL - NATURAL - MANMADE - NATURAL - NATURAL - MANMADE). However, participants could not simply repeat the same sequence throughout pre- or posttest, because each comprised three separate blocks with different category configurations and therefore different response patterns.

*Executive function task: Recent negatives*. The materials for this task were identical to those used in Experiment 1. Recall that two versions of the task were created: a treatment version designed to elicit substantial proactive interference (containing both "low no" and "high no" rejections), and a baseline version devoid of interference ("low no" rejections only).

### 2.4.2 Results

*Recent negatives*

*Model structure.* This analysis modeled button press RTs among treatment participants only. Fixed effects included response type ("yes"=-0.5 vs. "no"=0.5), rejection difficulty ("high lo"=0.5 vs. "low no"=-0.5), block number (1-3), and all possible interactions. Random effects included intercepts for both subjects and items; by-subject slopes for response type, block, and the interaction of rejection difficulty and block; plus correlations among all random effects.

| | *Block 1* | | *Block 2* | | *Block 3* | |
|---|---|---|---|---|---|---|
| | Treatment | Baseline | Treatment | Baseline | Treatment | Baseline |
| *Yes* | 835 (8) | 818 (8) | 840 (9) | 831 (9) | 846 (10) | 837 (10) |
| *Low no* | 844 (14) | 837 (8) | 833 (13) | 841 (9) | 859 (16) | 849 (9) |
| *High no* | 930 (17) | - | 878 (13) | - | 912 (17) | - |

Table 2.4 Mean (and standard error) RTs for Experiment 3 recent negatives task.

*Results*. The results presented here (see Table 2.4) are consistent with experiment 1. The recent negatives task elicited proactive interference from treatment participants, who were slower to reject probes on "high no" trials than on "low no" trials ($\beta$=0.06, se=0.01, $\chi^2(1)$=24.92, p<0.001). Participants also responded faster on "yes" trials compared to rejections ($\beta$=0.05, se=0.01, $\chi^2(1)$=21.53, p<0.001). Finally, treatment participants gave numerically slower responses to "yes" trials compared to baseline participants (Table 2.4), indicating that proactive interference slowed their processing overall. On the whole, these results support the assumption that the target executive function, interference resolution, was engaged.

*Picture classification*

Initial trials from each block were excluded from analysis. Removal of errors (including incorrect responses, dysfluencies, and technical errors) eliminated 1.3% of the data. Removal of RTs below 200 ms, above 2000 ms, or more than three standard deviations from a subject's mean eliminated an additional 1.1% of the data.

*Model structure.* As in experiment 1, three linear mixed effects models were constructed. The full model included both baseline and treatment data (*n*=60), while subset analyses explored the effects within each condition (*n*=30). Fixed effects in the full model included condition (baseline=-0.5, treatment=0.5), ordinal position within a category (1-5), block within a test session (1-3), test (pre=-0.5, post=0.5), class (manmade=-0.5, natural=0.5), and all possible interactions among these variables. Random effects included by-participant and by-item intercepts; by-participant slopes for ordinal position, block, test, class, and the two-way interactions of ordinal position by block, ordinal position by test, and block by test; and by-item slopes for condition, ordinal position, and the ordinal position by test interaction.

In the treatment model, the fixed effects structure was the same as the full model, except that the condition variable was replaced with a predictor estimating participants' proactive interference effect size during the recent negatives task (following experiment 1). The random effects structure included by-participant and by-item intercepts; by-participant slopes for ordinal position, block, test, class, and the block by test interaction; and a by-item slope for ordinal position. In the baseline model, the fixed effects were again identical to the main model, except that the condition variable and its interactions were removed. Random effects included by-participant and by-item intercepts; by-participant slopes for ordinal position, block, test, class, and the two-way interactions of ordinal position by block and block by test; plus a by-item slope for ordinal position.

*Transfer effects: Decrease in facilitation at post-test.* The full model replicated previous findings of semantic facilitation in classification, such that participants responded faster with each successive category member ($\beta$=-0.02, se=0.003, $\chi^2(1)$=33.92, p<0.001; fig. 2.4). As in experiments 1 and 2, participants responded slower in later blocks within a test

($\beta$=0.03, se=0.006, $\chi^2(1)$=21.91, $p<0.001$). They also responded slower during posttest compared to pretest ($\beta$=0.03, se=0.01, $\chi^2(1)$=7.59, $p<0.01$). This finding suggests that the recent negatives task impacted processing during the classification posttest, i.e., it indicates the presence of transfer. A marginal effect of item class shows that classification of natural items tended to be faster than classification of manmade items ($\beta$=-0.02, se=0.01, $\chi^2(1)$=3.66, $p=0.06$). Condition had no overall effect on response speed ($\chi^2(1)<0.001$, $p=0.996$).

Two interactions reliably modulated RTs. An interaction between block and test showed that participants' slowing across blocks was stronger at posttest than pretest ($\beta$=0.03, se=0.01, $\chi^2(1)$=5.80, $p<0.05$). Crucially, the interaction of ordinal position and test revealed that semantic facilitation was reduced at posttest compared to pretest ($\beta$=0.01, se=0.004, $\chi^2(1)$=5.95, $p<0.05$). This finding provides a second piece of evidence that performing the recent negatives task impacted the cognitive processes involved in classification. Unlike in experiment 2, this transfer effect was in the expected negative direction: a decrease in facilitation constitutes a performance decrement. However, the ordinal position by test interaction did not differ significantly across conditions ($\beta$=0.003, se=0.008, $\chi^2(1)$=0.10, $p>0.05$). As a result, we cannot make the strongest claim of selective negative transfer—there is not reliable evidence that the interference resolution manipulation caused a larger posttest performance decrement in the treatment group compared to baseline.

Nonetheless, the subset models provide some support for the claim that the effect of test differed across the treatment and baseline conditions. As in the full model, the treatment model showed effects of ordinal position ($\beta$=-0.02, se=0.003, $\chi^2(1)$=17.60, $p<0.001$), block ($\beta$=0.03, se=0.009, $\chi^2(1)$=9.13, $p<0.01$), and test ($\beta$=0.05, se=0.01, $\chi^2(1)$=9.14, $p<0.01$). The block by test interaction was once again significant ($\beta$=0.03, se=0.01, $\chi^2(1)$=5.24, $p<0.05$), as was the ordinal position by test interaction ($\beta$=0.03, se=0.01, $\chi^2(1)$=5.95, $p<0.05$), indicating that treatment participants showed a reliable reduction in semantic facilitation at posttest compared to pretest.[3] All other effects failed to reach significance ($\chi^2(1)\leq2.23$, $p\geq0.14$).

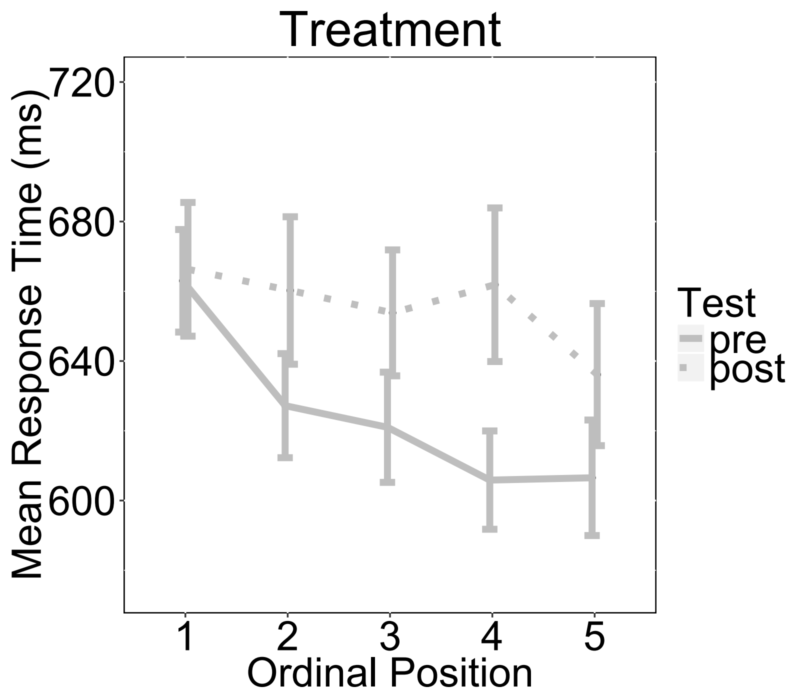
Turning to the baseline model, the results showed effects of ordinal position ($\beta$=-0.02, se=0.003, $\chi^2(1)$=23.49, $p<0.001$) and block ($\beta$=0.03, se=0.008, $\chi^2(1)$=13.50, $p<0.001$). Unlike in the treatment model, the effect of test was not reliable ($\beta$=0.01, se=0.02, $\chi^2(1)$=0.83, p=0.36). This divergence across conditions was not attested in the full model; however, this is unsurprising, given that the effects differ in magnitude but not direction across conditions. The effect of class was marginally significant ($\beta$=-0.02, se=0.01, $\chi^2(1)$=3.19, p=0.07), such that classification tended to be quicker for natural than manmade items.

---

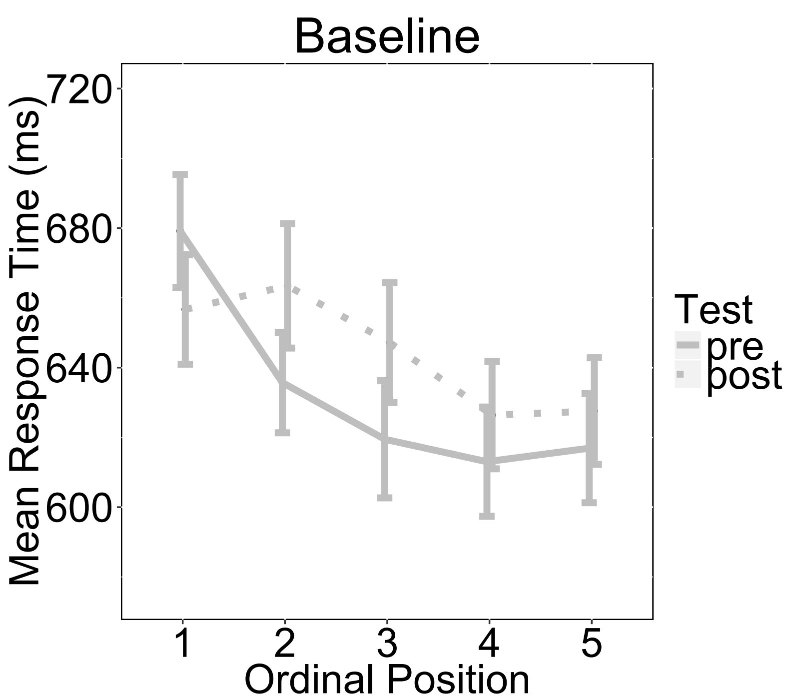
[3] In contrast to the picture naming results in experiment 1, the proactive interference predictor did impact the treatment model for experiment 3, albeit it in a complex way. There was no overall effect of proactive interference size on classification RTs ($\beta$=-0.005, se=0.006, $\chi^2(1)$=0.78, p=0.38), nor were the variable's second and third order interactions significant ($\chi^2(1)\leq1.80$, $p\geq0.18$). Instead, the four-way interaction between proactive interference size, ordinal position, block, and test proved reliable ($\beta$=-0.0004, se=0.0001, $\chi^2(1)$=6.05, $p<0.05$). We hesitate to over-interpret this complex result, which may merely reflect noise. At most, we note the possibility that participants' sensitivity to proactive interference may have predicted how much semantic facilitation they experienced in certain blocks.

Similar to the treatment model, baseline participants showed reduced semantic facilitation at posttest compared to pretest ($\beta$=0.009, se=0.004, $\chi^2(1)$=4.60, $p<0.05$), though the reduction was numerically smaller in this condition (from -14ms to -10ms per position) compared to treatment (from -14ms to -6ms per position; fig 2.4). Finally, there was a significant interaction between ordinal position, block, and test ($\beta$=-0.01, se=0.005, $\chi^2(1)$=3.75, p=0.053). This finding indicates that among the baseline participants, semantic facilitation changed in size across blocks during both pre and posttest, but in different ways. While facilitation decreased across pretest blocks, it was essentially absent at the start of posttest and gradually grew towards its original size. This pattern differs from the treatment condition, where facilitation remained reduced throughout posttest. The remaining interactions failed to reach significance ($\chi^2(1)\leq 1.72$, $p\geq 0.19$).

*No significant correlations between effects across tasks.* Using the same procedure as experiment 1, we tested the correlation between participants' recent negatives and picture classification performance. A negative relationship pairing larger (positive) proactive interference with larger (negative) semantic facilitation would corroborate the tentative evidence of negative transfer. However, while the data patterned in the expected direction, there was no significant correlation between treatment participants' proactive interference and semantic facilitation effect sizes ($r$=-0.18, n=30, p=0.34).




(A)




(B)

Figure 2.4 Decreasing RTs show semantic facilitation in experiment 3. Both conditions exhibited reduced facilitation at posttest; only treatment exhibited posttest slowing.

### 2.4.3 Summary

Taken together, the results of experiment 3 indicate transfer from the recent negatives task to continuous picture classification. First, all classification participants exhibited posttest slowing; this effect was only reliable within the subset model for the treatment condition. The absence of similar slowing in picture naming (experiment 1) rules out the possibility that intensive practice with the recent negatives task induced general cognitive fatigue. Instead, it appears that consistent engagement of interference resolution selectively impacted subsequent speech production when it relied heavily on conceptual processing (classification), but not when it depended on lexical retrieval (naming). Similarly, all classification participants showed reduced semantic facilitation at posttest, and this effect was numerically larger among treatment participants. Finally, treatment participants showed consistent reduction of semantic facilitation at posttest, whereas baseline participants recovered facilitation by the end of posttest. Because the executive function treatment had no such effect on semantic interference in naming, we again infer that interference resolution impacted this conceptually-driven production task, but not its lexically-driven counterpart.

## 2.5 Experiment 4: Negative Transfer from Picture-Word Interference to Picture Classification

Experiment 4, which tests whether response inhibition modulates semantic facilitation in classification, was analyzed in the same manner as experiment 2. We began with a combined analysis of the treatment groups from experiments 3 and 4, followed by a subset analysis of experiment 4 only. If the interference resolution (experiment 3) and response inhibition (experiment 4) treatments had differential impacts on neighbor co-activation during classification, then this combined analysis should reveal significant interactions with the experiment variable.

### 2.5.1 Methods

*Participants*

A final group of 30 participants was recruited from NU for course credit or $10 payment. All were native speakers of English only, and none reported any relevant cognitive impairment.

*Materials and design*

The materials for experiment 4 were recycled without change from the preceding experiments. Specifically, the picture-word interference task from experiment 2 was reused, as was the continuous classification task from experiment 3. These were inserted into the three-part negative transfer procedure utilized throughout this study.

### 2.5.2 Results

*Picture-word interference*

*Model structure.* This analysis modeled participants' vocal RTs during treatment[4]. Fixed effects included semantic relatedness (unrelated=-0.5 vs. related=0.5), block within half (1-2), and half (1=0.5 vs. 2=-0.5). Random effects included intercepts for subjects and

[4] One participant was excluded from this model, the within condition model, and the correlational analysis for failure to follow the picture-word interference instructions; many of their responses included indefinite articles before the target picture name.

items; by-subject slopes for block within half and half; and by-item slopes for semantic relatedness, half, and their interaction.

| | *Half 1* | | *Half 2* | |
|---|---|---|---|---|
| | Block 1 | Block 2 | Block 3 | Block 4 |
| *Related* | 869 (8) | 892 (8) | 814 (8) | 808 (8) |
| *Unrelated* | 849 (8) | 863 (8) | 787 (7) | 777 (7) |

Table 2.5 Mean (and standard error) RTs for Experiment 4 picture-word interference task.

*Results*. Analysis confirmed the presence of a relatedness effect: participants were slower to respond when the distractor word was semantically related to the target picture than when it was unrelated ($\beta$=0.03, se=0.007, $\chi^2(1)$=21.76, p<0.001). A reliable effect of repetition priming also emerged, indicating that participants responded faster during the second half of treatment ($\beta$=0.09, se=0.01, $\chi^2(1)$=29.49, p<0.001). All remaining effects failed to reach significance ($\chi^2(1) \leq 1.13$, $p \geq 0.29$).

*Continuous picture classification*

*Model structure.* As noted above, this analysis follows the same format as experiment 2. In the first model, we combined the experiment 4 data with those from the treatment condition of experiment 3. Fixed effects of interest included ordinal position (1-5), block within test (1-3), test (pre=-0.5 vs. post=0.5), intervening executive function task (recent negatives=-0.5 vs. picture-word interference=0.5), and all possible interactions. Item class (natural=0.5 vs. manmade=-0.5) and its interactions were included as a control. Random effects included intercepts for both participants and items; by-participant slopes for ordinal position, block, test, class, and the interactions of ordinal position by test and block by test; as well as by-item slopes for ordinal position, test, executive function task, and the test by executive function task interaction.

In the second model, the classification RTs from experiment 4 were analyzed alone. The fixed effects structure was the same as above, except that the predictor coding intervening executive function task was replaced with a by-participant measure of the amount of picture-word interference in the intervening task (beta estimates from by-participant linear regressions). The random effects structure included intercepts for both participants and items; by-participant slopes for ordinal position, block, test, class, and the two-way interactions of ordinal position by test and block x test; plus by-item slopes for ordinal position and test.

*Comparison of transfer effects in experiments 3 and 4: Decrease in facilitation at post-test*

When treatment participants from experiments 3 and 4 were combined, we observed semantic facilitation ($\beta$=-0.02, se=0.003, $\chi^2(1)$=39.49, p<0.001) and slowing across blocks ($\beta$=0.02, se=0.006, $\chi^2(1)$=11.84, p<0.001). All other main effects, including test and intervening executive function task, were not significant ($\chi^2(1) \leq 2.69$, $p \geq 0.10$).

A reliable two-way interaction between ordinal position and test confirmed that across both classification experiments, the magnitude of semantic facilitation was smaller (i.e., less negative) at posttest compared to pretest ($\beta$=0.02, se=0.004, $\chi^2(1)$=12.56, p<0.001). Importantly, this aspect of negative transfer was not dependent on the intervening executive function task (ordinal position x test x experiment n.s.; $\beta$=0.009, se=0.008, $\chi^2(1)$=1.17, p=0.28), suggesting a similar pattern of transfer across experiments

3 and 4[5]. The two-way interaction of block and test was also significant, indicating that the block effect—slowing across blocks—was driven by posttest performance (β=0.03, se=0.01, $\chi^2(1)=7.07$, $p<0.01$).

A marginally significant three-way interaction between ordinal position, block, and executive function task suggests that semantic facilitation tended to shrink across blocks in experiment 3, while growing across blocks in experiment 4 (β=-0.008, se=0.004, $\chi^2(1)=3.29$, p=0.07). Finally, the marginally significant three-way interaction of ordinal position, block, and test demonstrates that the posttest reduction of facilitation was numerically greatest in block 1 (β=-0.008, se=0.004, $\chi^2(1)=3.11$, p=0.08). All remaining interactions failed to reach significance ($\chi^2(1)\leq1.46$, $p>0.2$). Because the predictor for intervening executive function task had no major impact on the results of this combined analysis (i.e., it did not interact with the test variable), we conclude that transfer was fairly comparable across experiments 3 and 4.

*Transfer effects within experiment 4: Decrease in facilitation of post-test*

The results of the subset analysis confirmed reliable semantic facilitation within experiment 4, such that participants responded faster on each item within a semantic category (β=-0.02, se=0.003, $\chi^2(1)=27.36$, $p<0.001$; fig. 2.5). There was also an overall effect of block, indicating that participants responded slower across blocks within a test

[5] A model of pretest data only revealed a significant ordinal position by executive function task interaction, due to less semantic facilitation among experiment 3 than experiment 4 participants (β=-0.009, se=0.005, $\chi^2(1)=4.31$, $p<0.05$). However, given the absence of any test by executive function interactions in the main analysis, we refrain from examining this group-wise difference further.

($\beta$=0.02, se=0.007, $\chi^2(1)$=4.09, $p<0.05$). All remaining first order effects failed to reach significance ($\chi^2(1)\leq0.64$, $p\geq0.42$).

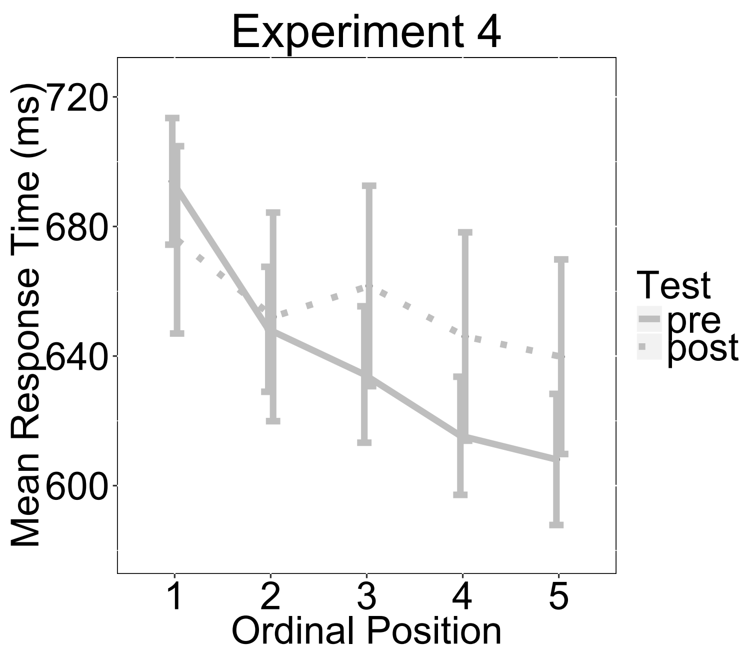



Figure 2.5 Reduction of RTs across category members indicates semantic facilitation in experiment 4. Effect size (slope) was reduced at posttest.

Three significant interactions were found within experiment 4. An interaction between ordinal position and test showed that semantic facilitation was reduced during posttest compared to pretest ($\beta$=0.02, se=0.007, $\chi^2(1)$=8.53, $p<0.01$). Under the account laid out in the introduction, this constitutes a performance decrement, i.e., a negative transfer effect. A two-way interaction between block and picture-word interference size also emerged: participants who experienced less interference in the intervening picture-word interference task showed more slowing across blocks in classification ($\beta$=-0.0007, se=0.0004, $\chi^2(1)$=4.12, $p<0.05$). This interaction suggests that participants who engaged response inhibition more strongly during the executive function task became more cautious over time during classification. We return to this argument in the discussion. The three-way interaction of ordinal position, block, and test also proved reliable ($\beta$=-0.01, se=0.007, $\chi^2(1)=4.31$, $p<0.05$). As explained in the previous section, this finding reflects how the posttest reduction of facilitation was most robust in block 1. The remaining interactions failed to reach significance ($\chi^2(1)\leq2.80$, $p\geq0.09$[6]).

*No significant correlations between effects across tasks.* Adopting the procedure from experiment 2, we tested the relationship between participants' picture-word interference and picture classification performance. Once again, no significant correlation emerged between participants' picture-word interference and semantic facilitation effect sizes ($r$=0.24, n=29, p=0.21).

**2.5.3 Summary**

The results of experiment 4 suggest that response inhibition can impact semantic facilitation during continuous picture classification. The primary finding to support this conclusion was the reduction in semantic facilitation at posttest, both in the combined analysis within experiment 3 and in the subset analysis. This result suggests that participants who continuously engaged response inhibition, like those who engaged interference resolution (experiment 3), experienced less benefit overall from neighbor co-activation at posttest than pretest.

---

[6] Similar to experiment 3, the four-way interaction of ordinal position, block, test, and picture-word interference size was marginally significant ($\beta$=-0.0005, se=0.0003,, $\chi^2(1)=2.80$, p=0.09).

## 2.6 General Discussion

Over four experiments, we investigated the role of two domain-general executive functions, interference resolution and response inhibition, in managing co-activation of semantic neighbors during production. In particular, experiments 1 and 2 addressed the central question of whether these executive functions are recruited to resolve conflict among neighbors during lexical selection. Following previous work using the negative transfer paradigm (Persson et al., 2007; 2013), we predicted that if either executive function is engaged by lexical conflict, then that executive function treatment should induce a posttest performance decrement in the form of greater semantic interference at posttest. The results contradicted this prediction of negative transfer.

In experiment 1, we observed no significant transfer effects from the recent negatives task, intended to engage interference resolution, to picture naming. We consider several possible explanations for this null result below. Ultimately, we conclude that interference resolution has no effect on lexical co-activation, at least in the continuous picture naming paradigm. Furthermore, this result suggests that cumulative semantic interference does not constitute a form of proactive interference from previously named category members.

In experiment 2, we did find evidence of transfer from the picture-word interference task, intended to engage response inhibition, to continuous naming. In comparison to the interference resolution treatment, the response inhibition treatment induced a type of performance improvement. Although treatment participants from experiments 1 and 2 showed equivalent performance at pretest, the latter group exhibited less semantic interference overall. In other words, participants who intensively engaged response inhibition were better at managing semantic interference. This result implicates response inhibition in the resolution of lexical conflict; however, the unexpected direction of the effect calls our original hypothesis into question and deserves further consideration. Below, we discuss these data within a conflict adaptation framework, presenting some recent studies corroborating the positive impact of executive function treatments.

Experiments 3 and 4 served as control conditions for experiments 1 and 2, testing the claim that domain-general executive functions would only be recruited in the context of conflict among linguistic representations. Continuous picture naming was replaced with continuous picture classification, where neighbor co-activation provides a benefit to production processing and produces (cumulative) semantic facilitation. We expected the absence of conflict during classification to produce null results; if inhibitory executive functions are only recruited by the production system in the context of processing difficulties, then we should observe no transfer effects from the executive function tasks to classification. Once again, the results contradicted our predictions.

In both experiments 3 and 4, the results revealed significant negative transfer from the target executive function task to classification. Experiment 3 showed that intensive engagement of interference resolution during the recent negatives task led to reduced semantic facilitation at posttest; experiment 4 showed a comparable reduction of facilitation after engagement of response inhibition during the picture-word interference task. Although it may seem counterintuitive, these data demonstrate that inhibitory executive functions can modulate facilitation effects during production. As with the results of experiment 2, we pull these findings together under a conflict adaptation account.

### 2.6.1 Reframing transfer effects as conflict adaptation

To understand this unexpected pattern of results, we utilize the conflict adaptation framework. This account claims that the experience of response conflict causes participants to become more cautious in making subsequent responses (Verguts, Notebaert, Kunde, & Wuehr, 2011). Such caution is arguably enacted in two different ways. First, participants engage in *task focusing* after conflict, meaning that they enhance executive functions to direct attention towards the relevant dimensions of stimuli (e.g., text color in the Stroop task) and away from the irrelevant/distracting dimensions (e.g., orthographic content in the Stroop task). Task focusing typically leads to *reduction* of the interference or incongruency effects caused by response conflict. The second way that participants respond to conflict is by slowing down subsequent responses in order to avoid making mistakes. This strategy of *post-conflict slowing* is similar to the phenomenon of post-error slowing (Verguts et al., 2011). While conflict adaptation is frequently investigated on a trial-to-trial basis, blockwise adaptations have also been explored (see Desender & Van Den Bussche, 2012, for discussion).

In support of this approach, several recent studies have shown short-term, positive effects of executive function engagement on language processing. Shell, Linck, and Slevc (2015) intertwined a picture-word interference task with cued language switching in bilingual speakers. They found that the presence of semantically related distractor words reduced the cost associated with language switching. Such a pattern could reflect conflict adaptation; experience with picture-word interference honed participants' response inhibition skills, which in turn helped them more efficiently switch languages. Similarly, Hsu and Novick (2016) recently reported a positive transfer effect in comprehension. Following incongruent Stroop trials (engaging response inhibition) participants made fewer comprehension errors on "garden path" sentences (where an incorrect structural analysis must be suppressed), compared to garden path sentences immediately following congruent Stroop trials. In other words, participants' garden path comprehension was improved by prior use of response inhibition to resolve conflict. Complementing the short-term positive transfer effects demonstrated by Shell et al. (2015) and Hsu and Novick (2016), there is a growing body of work showing long-term positive transfer via training (see Hussey & Novick, 2012, for a review). This work suggests that regular practice in engaging conflict resolution executive functions induces performance improvements across domains, including language processing.

We argue that the so-called positive transfer effects observed in experiment 2, along with the negative transfer effects observed in experiments 3 and 4, should be reframed as a form of conflict adaptation across blocks—and domains—at a time scale between the short- and long-term positive transfer effects reviewed above.

Experiment 2 participants experienced frequent response conflict during the intervening picture-word interference task, because half of the trials presented semantically related picture-word pairs. Given such exposure to response conflict, they may have become more cautious over time, engaging response inhibition and slowing their responses. If they carried this conflict adaptation over to the continuous naming posttest, it would explain the findings of positive transfer. Specifically, the application of response inhibition during continuous naming would stave off competition from co-active neighbors, reducing semantic interference as a result. Although we did not observe reduced interference within the subset analysis of experiment 2, the results of the combined analysis provide some support for this prediction: semantic interference was reduced overall within experiment 2 compared to experiment 1. Similarly, a decrease in response speed would buy participants more time to ensure selection of the target and not a co-active neighbor. The finding of reliable posttest slowing within the subset analysis of experiment 2 provides some evidence that participants adopted this strategy of post-conflict slowing (although there was no reliable difference in this effect across experiments).

Having reframed the positive transfer effects in experiment 2 as conflict adaptation, we turn to the results of experiments 3 and 4. We argue that the seemingly negative results of experiments 3 and 4 emerged because conflict adaptation was unnecessarily extended to the subsequent task. Participants engaged executive functions during treatment, and carryover of those inhibitory processes to classification inadvertently reduced the independent effect of semantic facilitation.

To unpack this argument, we consider the origins of semantic facilitation. During continuous classification, we assume that each successful classification strengthens the links between the relevant semantic features and the appropriate classification response (e.g., classifying EAGLE as “natural” strengthens the links between *animal, bird, feathered* and that response). Such strengthening speeds responses on subsequent within-category trials. As argued in the introduction, there should be no need for inhibitory executive functions during this task, because there is no representational or response conflict.

However, if another source of conflict (i.e., treatment) leads participants to engage executive functions, those control processes might interact with the production system, modulating semantic facilitation. Experiment 3 treatment participants gained lots of experience managing proactive interference from previous stimuli sets; experiment 4 participants gained similar experience suppressing pre-potent responses. These experiences most likely led them to rigorously engage interference resolution and response inhibition, respectively, in order to suppress the intruding representations. Crucially, we propose that these vigilant strategies were carried over to the classification posttest, where inhibition was applied to the semantic neighbors evoked by each picture. As a result, participants lost the benefit they had received from neighbor co-activation and exhibited less semantic facilitation. In contrast to experiment 2, where carryover of conflict adaptation across tasks/domains improved subsequent production performance, such carryover manifested as a performance decrement in experiments 3 and 4.

The recruitment of domain-general executive functions may also have been accompanied by post-conflict slowing in these experiments. A reliable effect of posttest slowing emerged within experiment 3 treatment participants, though not within experiment 4 or the combined analysis of both treatment conditions. In addition, the subset analysis of experiment 4 revealed that smaller picture-word interference in the intervening task predicted greater slowing across blocks during classification. From this, we infer that participants who more strongly engaged response inhibition tended to slow their classification responses over time. Such evidence may seem weak; however, it is possible that the effects of post-conflict slowing were masked by the impact of participants' heightened control (e.g., task focusing; see Verguts et al., 2011, for similar arguments).

Thus, we conclude that the conflict adaptation framework provides a ready account of the transfer effects observed in experiments 2-4. When experience with conflict leads participants to heighten control, such adaptation can either improve or worsen subsequent performance. If the following task evokes interference, up-regulated control will improve it; if the task is conflict-free, the additional control may hinder it. At least two mechanisms have been proposed to explain how conflict leads to recruitment of domain-general control. Botvinick, Braver, Barch, Carter, and Cohen (2001) propose that executive functions are engaged when a conflict monitoring mechanism detects response co-activation. Alternatively, Kurzban, Duckworth, Kable, and Myers (2013) argue that participants assess whether to allocate executive functions to an experimental task, to divide them between that task and available alternatives (e.g., daydreaming), or to disengage control altogether, depending on the relative costs vs. benefits of maintaining status quo performance. The current study cannot adjudicate between these two perspectives.

### 2.6.2 Response inhibition—not interference resolution—impacts cumulative semantic interference

Regardless of the mechanism underlying the conflict adaptation in experiments 2-4, the contrast between those results and experiment 1 has implications for theories of cumulative semantic interference. When viewed in isolation, the absence of effects may appear spurious; however, the results from the subsequent experiments largely debunk such claims. For instance, one might argue that the treatment version of the recent negatives task was not intense enough to induce conflict adaptation. After all, only one-third of rejection trials involved strong proactive interference (the “high no” trials); the remaining two-thirds triggered only moderate interference (“medium no” trials) or none at all (“low no” trials). While Persson and colleagues observed effects using these exact task parameters (2007), perhaps interference resolution must be more consistently engaged to elicit them in this particular population. This hypothesis predicts that we should never observe conflict adaptation from this recent negatives task to production. This prediction is contradicted by experiment 3, where the recent negatives task elicited changes in subsequent picture classification performance.

A second explanation for the lack of effects in experiment 1 is that continuous picture naming may be the wrong production task for detecting conflict adaptation. Co-activation may need to reach a certain level of intensity to trigger recruitment of domain-general executive functions (a la Botvinick et al., 2001). Under this assumption, the nature of response sets in continuous naming could prevent interference resolution from playing a role. Because each target can evoke an unbounded array of possible responses (e.g., a picture of a dog might activate DOG, CAT, RAT, PET, ANIMAL, etc.), co-activation is relatively diffuse and executive functions are therefore unnecessary. Belke and Stielow (2013) offered this type of explanation for their data, where a working memory load modulated the size of semantic interference in blocked naming, which involves a constrained response set, but not in continuous naming. Ries et al. (2015) provided a similar argument. The authors found that patients with lesions to the left inferior frontal gyrus, a region of the PFC thought to subserve executive functions, showed increased semantic interference in a blocked naming paradigm, but not in continuous picture naming. They argued that item repetition, rather than a bounded response set, intensifies lexical conflict in blocked cyclic naming vs. continuous naming. Ries and colleagues concluded that in the absence of task parameters enhancing lexical conflict (e.g., item repetition), language-specific selection mechanisms are sufficient for managing neighbor co-activation.

This intensity threshold account makes a straightforward prediction. If lexical co-activation is too diffuse in continuous picture naming for domain-general control to be required, then we should never observe conflict adaptation when continuous naming follows an intensive executive function task. This prediction is undermined by the results of experiment 2, where the intervening picture-word interference task produced changes in continuous naming performance.

Thus, we return to the original question of this investigation: which, if any, inhibitory executive functions are recruited by the production system to manage lexical co-activation? From the contrastive results of experiments 1 and 2, we conclude that domain-general response inhibition, but not interference resolution, can help reduce lexical conflict during continuous picture naming. However, it seems that the production system does not automatically recruit response inhibition whenever such conflict arises. Instead, conflict must be strong enough—either within the production task or, as we observed here, in the broader experimental context—before the system calls on additional control processes. The results of experiments 3 and 4 support this interpretation, demonstrating that outside sources of conflict can trigger engagement of executive functions during production, even when the dynamics of production do not demand them.

## 2.7 Conclusion

From the results of four negative transfer experiments, we draw several conclusions. The divergent results of experiments 1 and 2 (continuous naming) suggest that domain-general response inhibition, not interference resolution, may help resolve conflict among co-active lexical representations. They also support the idea that inhibitory executive functions are functionally separable. However, the significant results in experiments 3 and 4 (continuous classification) indicate that the engagement of such control by the production system is neither automatic, nor dependent on the presence of conflict within the production system itself. Despite the absence of lexical conflict during classification, both target executive functions—interference resolution and response inhibition—were triggered by the experience of heightened conflict during the intervening treatment tasks. We interpret these findings as evidence of conflict adaptation, which carried over from the intervening task to subsequent production. Taken together, the results of this study indicate the importance of tightly controlling laboratory experiences, to minimize the recruitment of non-target processes that may modulate the behavioral effects under investigation. More importantly, they constrain the scope of production theories, demonstrating that processes outside the production system may influence spoken performance, but only under specific conditions.

CHAPTER 3

## 3.1 Introduction to Study 2

A long-standing line of inquiry among speech and language scientists explores what, beyond phonological and phonetic structure, modulates variation in the spoken form of words. Researchers have identified numerous sources of variation and their corresponding acoustic consequences. These sources fall into at least two loose categories. First, variation can arise from the representation or retrieval of linguistic knowledge itself, as shown by effects of lexical frequency (Gahl, 2008; Bell, Gregory, Girand, & Jurafsky, 2009) and neighborhood density (Baese-Berk & Goldrick, 2009; Buz & Jaeger 2016; Goldrick, Vaughn, & Murphy, 2013; Munson & Solomon, 2004; Scarborough, 2004; Scarborough & Zellou, 2013; Wright, 2004). Second, variation can be triggered by the target's conversational context. Previous work has shown that the acoustics of a target change according to its contextual predictability (Bell et al., 2009), informativity (Seyfarth, 2014), givenness (Fowler & Housum, 1987; Kahn & Arnold, 2012; 2013; Lam & Watson, 2014), and the overall style of speech (Aylett & Turk, 2004; Baker & Bradlow, 2009). The current study investigates a third possible source of articulatory and therefore acoustic variation: the coordination of production processes. Specifically, this study tests whether the timing of response initiation relative to planning has systematic consequences on articulation, as indexed by word durations. In other words, we ask how the timing of a speaker's decision to start articulating a response impacts the duration of their production.

This question arises from the now commonplace assumption of some degree of interactivity in the production system, via the concept of cascade (e.g., Dell, 1986). Under this assumption, later stages of processing can begin before earlier stages are complete. Interactive frameworks therefore enable articulation to begin during ongoing planning. This account predicts that the relative timing of response initiation may be related to the articulation of a target in at least two ways. First, variation in response initiation might affect the degree of co-activation among the target and competitor representations at the start of articulation. This co-activation could affect the details of the speech plan sent to the motor system. Relatedly, the timing of response initiation may modulate the extent to which planning continues during execution of the target. The degree to which co-activation still needs to be resolved during motor execution may affect articulatory processing after response initiation. The current study seeks evidence for both types of interactive effects.

Importantly, there is currently no consensus among interactive production theories on how exactly the timing of response initiation is determined. One line of work posits the existence of a minimal planning unit (MPU), i.e., some fundamental constituent that must be fully specified before articulation can begin (see Kawamoto, Liu, & Kello, 2015, for a review). This *MPU hypothesis* suggests a fairly stable relationship between planning and articulation, such that responses are always initiated after a similarly sized unit has been planned (though the underlying planning may take more or less time). Another body of research assumes that there is no fixed metric for response initiation. Under this *flexible cascade hypothesis*, the amount of temporal overlap between planning and articulation is thought to change dynamically depending on task demands (Kello, MacWhinney, & Plaut, 2000; Kello, 2004). All else being equal, when response initiation occurs earlier, there is predicted to be greater overlap between planning and articulation. While the current study cannot definitively adjudicate between these two accounts, we seek to strengthen the support for the flexible cascade hypothesis by addressing three key questions.

*Does the timing of response initiation impact articulation?*

We begin with the fairly simple goal of establishing whether the relative timing of planning and articulation are linked. We achieve this by testing the relationship between response times (RTs) and word durations across three production tasks that vary in response selection difficulty. We utilize two different RT predictors for this purpose: a measure of speakers' overall response speeds (i.e., their mean RTs) and a measure of trial-level RT. Given the assumption of temporal overlap between planning and articulation, we expect that a relationship might emerge between them, such that either or both of the RT variables predict speakers' word durations. Critically, under the flexible cascade hypothesis, we expect that this relationship may vary across the three production tasks, according to their different processing demands (Kello et al., 2000; Kello, 2004).

*Does response selection difficulty modulate articulation after response initiation?*

Next, we investigate whether response selection difficulty has a "direct" effect on articulatory processing. In other words, we ask whether response selection difficulties continue to influence articulatory processing even after articulation has begun, e.g., if word durations are sometimes lengthened to allow time for residual planning. In this case, we expect to observe direct effects of lexical selection difficulties on word durations, even when response timing is already taken into account.

Previous evidence that lexical selection disruptions impact articulation has been equivocal, even when speeded manipulations force earlier responding. Kello et al. (2000)

found that under time pressure, Stroop interference—caused by incongruency between the ink color a speaker is instructed to name and the color word written in that ink (MacLeod, 1991)—not only yielded increased latencies but also caused speakers to lengthen the duration of target words. However, Damian (2003) failed to replicate these results; furthermore, in two other paradigms, Damian failed to find any evidence that speakers increase durations under conditions that disrupt lexical selection.

In the current study, we propose that consideration of individual variation may help reconcile this inconsistent evidence for direct lexical effects on articulation. In particular, we argue that participants may vary in their sensitivity to the experimental manipulations intended to trigger difficulties in lexical selection. If a participant's selection processes are relatively unaffected by a manipulation, then their articulation will be too, preventing the interactive potential of the production system from being detected. Thus, a sample group that includes less sensitive participants may fail to demonstrate direct lexical effects on articulation, even if the production system is capable of supporting them.

*Do RT-mediated and direct interactive effects on articulation interact?*

Finally, we further evaluate the flexible cascade hypothesis by testing whether the timing of response initiation modulates our ability to detect direct lexical effects on articulation. If the temporal overlap between planning and articulatory processes increases as RTs decrease, then direct lexical effects should be more likely to impact articulation when responses are earlier. In a series of five experiments, Kello (2004) provides compelling support for this prediction. After observing baseline performance in a standard word naming task (experiment 1), Kello pushed speakers towards greater overlap between processes by speeding up their responses in two tempo-naming experiments (experiments 2 & 3). He observed that as the tempo sped up and participants responded sooner, the effects of lexical manipulations (frequency, orthographic neighborhood size, and spelling-sound consistency) shifted from RTs to word durations. In other words, he found that earlier response initiation decreased the influence of high-level effects on a planning measure (RTs) and increased their influence on an articulatory measure (durations). These results further contrasted with experiments 4 and 5, where participants saw the stimulus at trial onset but withheld responding until a tempo-timed response cue (delayed tempo-naming). In this case, the provision of planning time between the stimulus and response cue eliminated lexical effects on both RTs and durations. Kello concluded that, in the absence of preparation time, earlier response initiation increases temporal overlap and therefore direct interactions between planning and articulation.

In the current study, we apply the same logic as Kello (2004) within trials rather than at the task level. Concretely, we examine whether trial-level RT affects the manifestation of direct lexical effects on articulation. As above, comparison of this interaction across three production tasks with different lexical selection demands will help assess the production system's flexibility in coordinating planning and articulation.

In the remainder of the introduction, we review the lines of research that motivated this investigation. First, we consider previous work on the timing of response initiation relative to speech planning. Second, we explore the literature on direct interactive effects between lexical selection and articulation. Finally, we pull these threads together in laying out predictions for three production experiments that manipulated the ease of lexical selection by inducing co-activation of semantic neighbors. These predictions are informed by recent work taking a similar approach by considering RT and duration data in tandem.

### 3.1.1 Evidence for the flexible coordination of planning and articulation

As noted above, the question of when articulation begins relative to planning has been addressed from various angles. One approach seeks to identify the MPU necessary before articulation can unfold. This approach has been made most explicit by Kawamoto and colleagues, who advocate the phoneme as the MPU in spoken production (see Kawamoto et al., 2015, for a review). That is, they propose that speakers can initiate responding as soon as the first phoneme of a target is phonetically encoded. The strongest evidence for a phonemic MPU is the finding of negative RTs when the initial phoneme has been primed. Specifically, Kawamoto, Liu, Lee, and Grebe (2014) presented participants with a reading aloud task. Critical trials began with the presentation of a target's initial phoneme only, before the rest of the word appeared 600 ms later. On approximately one quarter of such trials, Kawamoto et al. found negative RTs, i.e., response initiation before the entire word was presented. By definition, this result suggests that articulation can proceed when only a single segment has been fully planned, meaning that planning and articulation overlap extensively in time.

However, many production models take a different view, arguing that the MPU is a syllable or more. For example, Levelt and colleagues' (1999) seminal model floats the idea that a full phonological word—a prosodic unit that is bounded by pauses and can contain multiple lexical items—is planned before articulation begins. The group followed up on this idea in a study of word length effects in object naming (Meyer, Roelofs, & Levelt, 2003).

They predicted that if response initiation takes longer when more material needs to be planned (i.e., when the phonological word is larger), then speakers should be faster to name objects with short than long names. This prediction was confirmed under certain circumstances: participants named objects with monosyllabic names more quickly than objects with disyllabic names, but only when those objects were blocked according to name length. When the objects were intermixed, the length effect disappeared. Meyer et al. concluded that the phonological word can serve as the MPU when circumstances facilitate that strategy; otherwise, articulation may begin earlier, with only the first syllable's phonetic plan prepared.

Meyer et al.'s (2003) mixed results, which indicate that the timing of response initiation relative to planning may depend on task parameters, point toward a third perspective on this issue. While many researchers continue to debate the size of the MPU in spoken production, others question the existence of a fundamental planning unit in the first place. For example, Pluymaekers, Ernestus, and Baayen (2005) argue that phonetic plans are continuously assembled and articulated, rather than being articulated on a unit-to-unit basis. They present data from a corpus of spontaneous Dutch speech, testing for predictability effects on the durations of word stems and suffixes. The authors detected independent effects of repetition, predictability from the preceding word, and predictability from the following word. Critically, each of these effects modulated the duration of different items and different portions of those items (i.e., stems vs. suffixes), sometimes crossing morpheme boundaries and sometimes being confined to a single part. Based on these data, Pluymaekers et al. conclude there is no standard unit of planning during production, implying that the timing of response initiation is variable and dependent on speakers' decision criteria for when to begin articulation.

Taken together, this sampling of work on the coordination of planning and articulation demonstrates the lack of consensus in the field. Not only does the MPU appear to vary across studies, but it can also vary within-study depending on task parameters (Meyer et al., 2003). We interpret these findings as evidence that speakers' response decision criteria may be fairly flexible, subject to a range of possible factors including motivation, attention, and planning difficulty, which is a primary focus here. This proposal echoes the work of Kello and colleagues (2000; Kello, 2004), who argue that the temporal overlap between planning and articulatory processes is not a fixed feature of the production system's architecture. Instead, they propose that the extent of cascade and its behavioral manifestation fluctuate dynamically across trials, individuals, and tasks.

### 3.1.2 Evidence for direct lexical effects on articulation

One factor that should influence the coordination of planning and articulation is the difficulty of response selection. In the current study, we focus on process of lexical selection, where speakers much choose the intended word from a field of competitors (e.g., select DOG from among DOG, CAT, RAT) before encoding its sound structure. As noted above, selection difficulties may influence not only the timing of response initiation (and thus the articulatory plan available when a response is initiated), but also the dynamics of subsequent articulatory processing.

A number of studies have shown that cascade enables interactions between consecutive stages of production processing, such that difficulties at one stage influence the stage immediately following. Semantic neighbor effects confirm that co-active semantic representations influence lexical selection (as illustrated by semantic interference effects across tasks: picture-word interference, Rosinski, Golinkoff, & Kukish, 1975; blocked-cyclic picture naming, Belke, Meyer, Damian, 2005; continuous picture naming, Howard, Nickels, Coltheart, Cole-Virtue, 2006). Mixed error effects demonstrate that lexical co-activation influences phonological planning (and that phonological planning can also influence lexical selection; Goldrick, 2006). Finally, articulatory and acoustic blends during error production reflect cascading activation from phonological planning to phonetic planning (Goldrick & Blumstein, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007).

Recent data indicates that long distance interactions can occur as well, such that lexical-level activation affects articulatory outcomes. For example, McMillan, Corley, and Lickley (2009) demonstrated that the lexicality of an error outcome determines whether the target vs. error's phonetic properties are more strongly realized. Using articulatory data, they found that for errors like *gome* → *dome*, onset articulation closely approximated a canonical /d/ onset due to the strong influence of the real word outcome; for errors like *gofe* → *dofe*, the error onset was less /d/-like because of the weaker influence of the nonword outcome. In a similar vein, Goldrick, Baker, Murphy, and Baese-Berk (2011) tested whether the word frequency of the target vs. error outcome influences the phonetic attributes of the error produced. Acoustic data showed that low frequency items benefitted from enhanced phonetic processing relative to high frequency items: low frequency targets exerted a strong influence on the acoustics of error outcomes, while low frequency error outcomes were robust to intrusions from the target. Together, these data show that the relative lexical activation of the target vs. competitors can dictate which gradient phonetic properties are realized during production.

Nonetheless, when word durations are utilized to index articulation, the evidence for long distance interactions has been less consistent. Kello et al. (2000) utilized the Stroop manipulation (MacLeod, 1991) to test for effects of disrupted lexical selection on word durations. While the authors replicated the finding of semantic interference in RTS—slower responses on incongruent trials (RED written in blue ink) compared to congruent trials (RED written in red ink)—they failed to observe any effect on durations. After implementing a response deadline, interactive effects emerged. When participants suffered impaired lexical selected on incongruent trials, they not only slowed their responses but also increased word durations.

However, in a direct replication, Damian (2003; experiment 3) failed to find any effect of Stroop interference on word durations, even under time pressure. He implemented two additional paradigms to rigorously test for long distance interactive effects. In experiment 1, Damian used the picture-word interference paradigm (Rosinski et al., 1975), where picture naming is slower in the context of semantically related distractor words than in the context of unrelated words. He expected that semantic interference might be more robust during the naming of imageable objects in this task than during the naming of color words in the Stroop task; therefore, this paradigm might prove more successful at pushing the production system towards more cascade. Participants with and without a response deadline failed to show any effect on word durations, despite robust RT interference effects. In experiment 2, participants named pictures that were either unrelated or blocked by semantic category (Damian, Vigliocco, & Levelt, 2001), with lexical selection being disrupted in the latter condition due to strong competition from category-related neighbors. Once again, only RTs were modulated by the presence of semantic interference, even under time pressure.

Similarly mixed results have been reported when examining the influence of lexical neighborhood density on articulation. Buz and Jaeger (2015) explored how phonological neighborhood density influences word durations during picture naming. They found that trial-level RT predicted articulatory outcomes, such that slower responses exhibited longer durations. They also observed an independent effect of neighborhood density on durations (over and above RT effects). These data support the argument that lexical disruptions can influence articulation both before and after response initiation, i.e., that planning-mediated and direct interactive effects on articulation can occur in the same study. However, Heller and Goldrick (2014; 2015) found no evidence of a direct interactive effect. They explored whether a target picture's grammatical category (noun vs. verb) constrained the influence of its phonological neighbors (e.g., SAT and CAR are phonological neighbors of CAT) during subsequent phonological and phonetic planning. While, like Buz and Jaeger, they found a positive effect of trial-level RT on durations, they found no direct effect of grammatical category-constrained neighborhood density.

These conflicting duration results suggest that direct interactions between lexical planning and articulation are necessarily weak and difficult to detect. Such subtlety may inherently arise from the architecture of the production system, assuming some representational distance between lexical and phonetic nodes in the network. Concretely, if co-activation at the lexical level must travel through at least two subsequent levels of representation (phonological and phonetic) in order to impact articulation, then lexical disruptions may be washed out by noise and other inputs to the system before they can impact motor execution. However, we focus here on an additional possibility raised in the introduction, that individual variation may prevent detection of group-level effects.

A role for individual variation in determining long distance interactions falls naturally out of the flexible cascade hypothesis, which argues that the manifestation of interactivity is dependent on the coordination of production processes (Kello et al., 2000; Kello, 2004). In other words, the account assumes that long distance interactions occur due to temporal overlap between planning and articulation, and that the extent of that overlap can be flexibly determined by speakers' response decision criteria. This logic not only predicts that interactive effects may vary across task conditions, as Kello and colleagues have demonstrated, but also across individual speakers.

In particular, we argue that speakers vary in their behavioral sensitivity to experimentally induced lexical disruptions. Such differences may be transient (e.g, the participant is not attending to the task) or intrinsic to speakers' cognitive processes (e.g., the participant is skilled at resolving co-activation). Regardless, they might explain why the effects of long distance interactions between lexical selection and articulation have been unstable across previous studies. For instance, many manipulations increase the difficulty of lexical selection by exposing participants to semantic neighbors of the target (e.g., naming DOG is harder after naming CAT; Wheeldon & Monsell, 1994). For a participant who is less sensitive to these manipulations, co-activation is more likely to be resolved by response initiation, providing little opportunity for it to influence articulation after that point.

### 3.1.3 Predictions for the current study

In the remainder of the introduction, we outline the three experiments presented below and predict how the timing of response initiation and the difficulty of lexical selection might impact word durations. In experiment 1, we utilized the continuous picture naming paradigm, where speakers simply name one picture after another (Howard et al., 2006). Crucially, the stimulus list comprised many different semantic categories, with members of the same category separated by several trials. Multiple studies have shown that RTs increase linearly with each subsequent naming of a given semantic category, indicating cumulative semantic interference (e.g., Belke, 2013; Belke & Stielow, 2013; Howard, et al., 2006; Navarrete, Mahon, Caramazza, 2010; Oppenheim, Dell, Schwartz, 2010). Experiment 2 used the same stimuli and procedure, but it required participants to perform continuous picture classification instead (natural vs. manmade; Belke, 2013). This task is known to elicit cumulative semantic facilitation, where RTs decrease with each classification of a given category. Finally, experiment 3 implemented the blocked cyclic naming paradigm (Damian et al., 2001), where participants name blocks of pictures that are either blocked by semantic category or mixed. Numerous studies have used this particular paradigm to show slower RTs in semantically blocked compared to mixed contexts, indicating semantic interference (e.g., Breining, Nozari, Rapp, 2015; Crowther & Martin, 2014; Damian et al., 2001; Navarrete, Del Prato, Mahon, 2012). Across these experiments, we generated three sets of predictions.

*Predicted effects of overall response speed*

First, we considered whether participants' average response speeds (i.e., mean RT) affect their word durations. Under the hypothesis that participants might use top-down mechanisms to set a consistent rate of processing throughout production, we predict a positive correlation between overall speed and durations. If participants try to maintain a steady pace from planning to articulation, then we expect generally faster responders to show shorter durations overall.

*Predicted effects of trial-level RT*

Below, we present a range of predictions for the possible effect of trial-level RT on word durations. These predictions assume that the relationship between these planning and articulatory measures depends on how participants' response decision criteria are adjusted in the face of lexical selection difficulties. While we will ultimately conclude that lexical selection difficulty is not the sole determinant of the timing of response initiation (see General Discussion), this simplifying assumption allows us to make clear predictions.

*Non-adaptive responders.* When many semantic neighbors are co-activated and selection is more difficult, a participant may choose not to adjust the timing of response initiation, e.g., if they decide to ignore their previous decision threshold or make it more lax in order to maintain the same pace. Assuming RTs are held fairly constant in this way, we expect no overall effect of trial-level RT on word durations. However, this may lead to greater overlap between planning and articulation (see below).

*Adaptive responders*. Under an alternative account, when a participant is faced with difficult lexical selection, they may adjust the timing of response initiation in order to maintain the same decision threshold and allow more time to meet it. One major effect of this would be to reduce the interaction between planning and articulation (see below). In terms of the relationship between RT and duration, at least three alternative outcomes could be predicted:

- Subsequent articulation might proceed as if co-activation had never occurred. This account predicts no primary effect of trial-level RT on durations.
- Next, delayed response initiation might lead to expedited articulation, to avoid exceeding the finite time available for responding. Under this account, we expect a negative effect of trial-level RT on durations, such that faster responses tend to have longer word durations.
- Finally, delayed response initiation might also lead to slower articulation, if the speaker extends their caution during planning to response execution. In this case, we expect a positive correlation between trial-level RT and word durations, such that faster responses have faster durations.

*Predicted effects of response selection difficulty*

Our third set of predictions examines whether response selection difficulty has a direct impact on articulation, above and beyond any effects mediated by RTs. In other words, we explore whether lexical co-activation influences articulatory processing after response initiation. To that end, we sought an overall effect of each experimental manipulation on word durations when RT was accounted for as a covariate. If ongoing planning after response initation directly affects articulation, then the neighbor effects in our three paradigms should not be reducible to RT effects alone.

*Interactions with response selection difficulty*

As reviewed above, previous research demonstrates that direct lexical effects on word durations are hard to detect. To improve the sensitivity of our analyses, we factored in individual variation in sensitivity to the experimental manipulations. Specifically, we tested whether the size of participants' neighbor effects in RTs predicted their duration effect sizes. Such an interaction might provide insight into the source of lexical effects that influence articulation after a response has been initiated. In particular, it would suggest that the same lexical co-activation that disrupts planning continues spreading to subsequent processes even after articulation has begun.

We also examined the interaction between trial-level RT and response selection difficulty. If responders are not adaptive, maintaining a constant rate of production in spite of selection difficulties, there is a greater chance that co-activation will need to be resolved after response initiation, predicting an interaction between trial-level RT and the size of semantic interference/facilitation in durations. Concretely, we expect trials with faster RTs to show larger neighbor effects in durations, as there is greater overlap between planning and articulation processes on these trials. Alternatively, if responders are adaptive, delaying responses would allow more time to resolve co-activation, eliminating effects of response difficulty on durations (producing no interaction of RT with response selection difficulty).

## 3.2 Experiment 1: Continuous Picture Naming

Out of our three experiments, the first one uses the most naturalistic paradigm: continuous picture naming (Howard et al., 2006). Participants simply named a sequence of pictured objects one after another, where that sequence has been implicitly manipulated to include multiple members of each semantic category. The word duration results from this experiment seem most likely to generalize to other production tasks, compared to the more artificial manipulation used in experiment 2 (classification) and the explicit manipulation in experiment 3 (blocked-cyclic naming).

### 3.2.1 Methods

*Participants*

We recruited 90 participants at Northwestern University (NU) using the Linguistics Department subject pool and flyers around campus. Each group of participants received course credit or $10 compensation, respectively. They reported learning no language other than English before age 5 and no history of cognitive impairment.

*Materials and Design*

The data analyzed here were originally reported by Fink and Goldrick (in prep), where the authors tested for effects of two executive function treatment tasks on picture naming RTs. To avoid contamination of the duration data by those treatments, we analyzed durations from pretest only here.

Participants performed a version of the continuous picture naming task (Howard et al., 2006), where they named a sequence of pictures as quickly and accurately as possible. 90 colored line drawings were drawn from Rossion and Pourtois' (2004) database, depicting 5 items each from 18 semantic categories (see Appendix; average word frequency of 65.3 words per million; SUBTLEX, Brysbaert & New, 2009). Between participants, the stimulus lists were counterbalanced to ensure that each item appeared at both pre- and posttest, as well as in every ordinal position within its category (1-5). As a result, all items are represented in the current data set, though each participant only saw half of them.

Within a participant, the 9 categories presented during pretest were subdivided into 3 blocks, with items drawn from 3 categories in rotation (e.g., a block containing birds, fruits, and vehicles might begin OWL - APPLE - CAR - PEACOCK - ORANGE - PLANE). As noted in the original study (Fink & Goldrick, in prep), this consistent lag of 2 trials between category members deviates from Howard et al.'s original schema, but it is known to elicit the standard cumulative semantic interference effect (Runnqvist, Strijkers, Alario, & Costa, 2012; Schnur, 2014).

Trials were automatically timed, with stimuli remaining on screen until a 2 second response deadline. The pretest procedure was quite short, lasting just over 3 minutes. More detail about trial composition and the overall experimental procedure can be found in Fink and Goldrick (in prep).

*Data processing*

In all experiments, word durations were extracted from stereo recordings. These contained audio markers time-locked to stimulus onsets on the right channel, plus participant speech on the left. After segmenting the recordings into trials using the audio markers, speech onsets and offsets were identified using intensity thresholds. Each trial was first equalized to an average root mean square intensity of 0.02 Pascal. The Praat Intensity function then estimated the intensity contour of the normalized signal. Speech onsets were located by sampling this contour at 1-millisecond (ms) increments to detect when the normalized signal passed a 55 dB threshold. Speech offsets were located in the same fashion, except that sampling began from trial's end. Duration was defined as the difference between speech onset and offset. The first author manually corrected these boundaries to avoid false triggers due to lip smacks, breathing, and/or low amplitude segments.

### 3.2.2 Results

An initial process of error and outlier removal, based on RTs, was reported in the original study (Fink & Goldrick, in prep). Word durations were then log-transformed to compensate for positive skew, before we conducted a second round of duration based outlier trimming. Word durations more than 3 standard deviations away from a participant's mean were removing, eliminating an additional 0.3% of the pretest data.

*Model structure*

A single mixed-effects regression model was constructed to test for effects of response timing and response selection difficulty on word durations. Fixed effects of interest included a by-participant overall speed, trial-level RT, ordinal position within a semantic category (1-5), by-participant RT interference size, and the two-way interactions of ordinal position by trial-level RT and ordinal position by RT interference size. Several control variables were also included as fixed effects: experimental block (1-3), item class (manmade = -0.5, natural = 0.5), and a block by ordinal position interaction. All of these predictors were centered to avoid issues of co-linearity. Following the model selection procedure recommended by Bates, Kliegl, Vasishth, and Baayen (2015), random effects included intercepts for both participants and items, plus by-participant slopes for block and trial-level RT.

The by-participant predictors for overall speed and RT interference size were created using simple linear regressions, where participants' continuous naming RTs were predicted by ordinal position only. Specifically, intercepts and beta coefficients were extracted from those RT models and input into the current duration analysis. We note that a more conservative analysis would extract best linear unbiased predictors (BLUPS; Baayen, 2008) from a complete, mixed-effects model of the RT data. Such an approach estimates the influence of the target predictor, ordinal position, while also taking into account variance from other sources (e.g., block number). However, our participants demonstrated little variance when their ordinal position effects were estimated by this method, rendering the approach unfeasible.

*Overall response speed*

A reliable effect of by-participant speed revealed that participants who responded slower overall tended to produce longer word durations ($\beta=0.053$, s.e.=0.014, $\chi^2(1)=12.94$, $p<0.001$; fig. 3.1). This result supports the hypothesis that speakers may use some form of top-down control to maintain a consistent speed of processing throughout production.

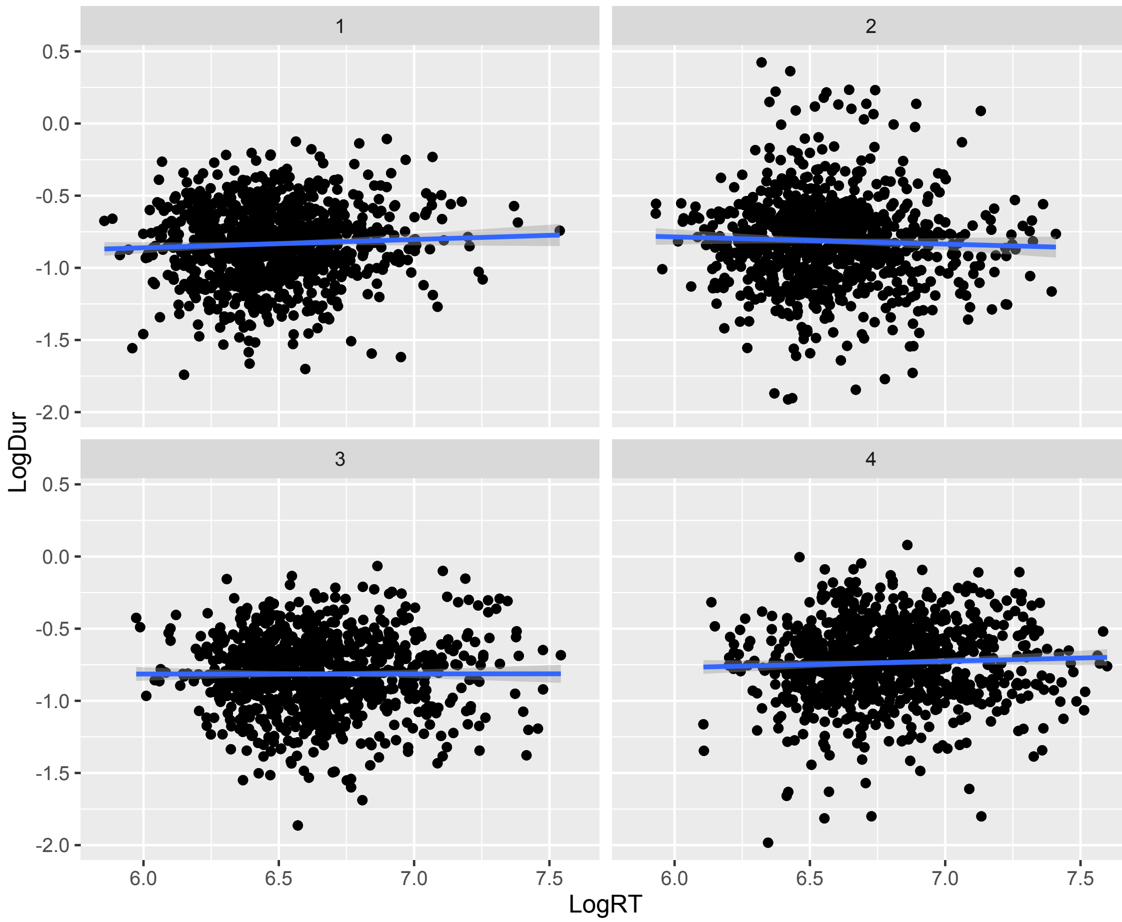



Figure 3.1 Participants divided into 4 quartiles, according to overall speed. Log-transformed durations (y-axis) plotted by log-transformed RTs (x-axis). Slower responders (quartiles 3 & 4) have longer durations than the faster responders (quartiles 1 & 2). No correlation between trial-level RTs and durations (all panels).

*Trial-level RT*

While the overall effect of trial-level RT patterned in a negative direction, such that faster responses had longer durations, it proved too variable to draw conclusions from ($\beta$=-0.019, s.e.=0.016, $\chi^2(1)$=1.36, p=0.24). This null effect does align with certain predictions from the introduction; however, we refrain from interpreting it until later in the study, after we have combined experiments 1 and 2 for further analysis below.

*Response selection difficulty*

There was a significant effect of ordinal position on word durations ($\beta=0.005$, s.e.=0.002, $\chi^2(1)=6.76$, $p<0.01$). As shown in figure 3.2, this result indicates that semantic interference slowed down articulation of target picture names. Because this effect proved reliable with RT accounted for as a covariate, it reflects the influence of lexical co-activation on articulation after response initiation. A marginal interaction of ordinal position and block indicates that this semantic interference in durations emerged over time, particularly in blocks 2 and 3 ($\beta=0.005$, s.e.=0.003, $\chi^2(1)=3.48$, $p=0.062$; fig. 3.3)

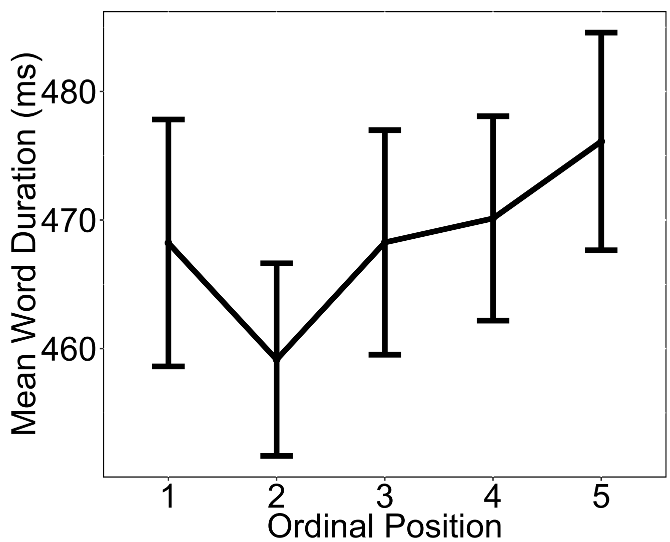



Figure 3.2 Increase in duration across positions in a category shows semantic interference.

The two-way interaction between ordinal position and the size of RT interference had no reliable effect on word durations ($\beta=-0.0001$, s.e.=0.0009, $\chi^2(1)=0.02$, $p=0.88$). The size of semantic interference in durations was apparently consistent across participants, regardless of whether their RTs indicated large neighbor effects prior to response initiation. This finding suggests that during continuous naming, disruptions during planning and disruptions during articulation may only be loosely linked. Alternatively, this null result may stem from the lack of variation in RT interference size across participants (Fink & Goldrick, in prep), which can prevent detection of a relationship that would emerge in a more diverse sample.

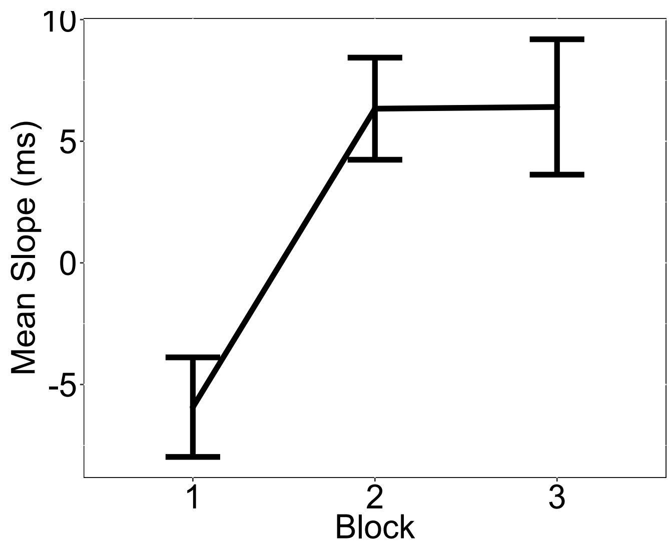



Figure 3.3 Interference (illustrated here using beta coefficients from simple regressions of participants' duration data) emerged in the later two blocks of continuous picture naming.

*Interactions with response selection difficulty*

Similarly, the two-way interaction of ordinal position and trial-level RT had a non-significant effect on durations ($\beta$=0.004, s.e.=0.007, $\chi^2(1)$=0.27, p=0.61). This result suggests that participants adjusted their response decision criteria to allow more time for the resolution of lexical co-activation. If they had not adapted, we would expect to observe greater semantic interference in word durations on faster responses, indicating that some amount of lexical co-activation had to be resolved after response initiation.

*Control variables*

A marginal effect of block showed that durations grew longer across blocks ($\beta$=0.011, s.e.=0.007, $\chi^2(1)=2.80$, $p=0.094$). All remaining effects failed to reach significance ($\chi^2(1)\leq 1.43$, $p\geq 0.23$)

**3.2.3 Summary**

Experiment 1 demonstrated simultaneous effects of response initiation timing and response selection difficulty on articulation during continuous picture naming. In terms of response initiation, we found that the global measure of response timing predicted word durations, while the trial-level measure did not. This suggests a greater role for top-down than bottom-up mechanisms in regulating the timing of response initiation, at least during continuous naming. In terms of response selection difficulty, we found that the ordinal position manipulation not only generates cumulative semantic interference in RTs, but also in word durations. In general, this result aligns with recent evidence of long distance interactions in the production system that extend from lexical selection to articulatory processing (e.g., Goldrick et al, 2011; Kello et al., 2000; Kello, 2004; McMillan et al., 2009). Furthermore, it corroborates recent work showing lexical effects on articulation even when trial-level RT is treated as a covariate (Buz & Jaeger, 2015), suggesting that lexical co-activation can influence articulation after response initiation.

### 3.3 Experiment 2: Continuous Picture Classification

In experiment 2, we explore the relationship between response initiation and articulation when semantic neighbor co-activation supports target selection rather than hindering it. Namely, we conduct the same analyses from experiment 1 on word duration data from a continuous classification task, where semantic facilitation is observed instead of interference. If the RT predictors have similar effects on word durations to those observed in experiment 1, then we might infer the existence of some default mode of coordination between planning and articulation. On the other hand, if we observe different RT effects on durations in the context of semantic facilitation, then we might begin constraining the relationship between response initiation and articulation under different processing conditions.

#### 3.3.1 Methods

*Participants*

Ninety participants were recruited at NU, and they received either course credit or $10 compensation for their time. None had learned a language other than English before age 5, nor reported any history of language or vision impairment.

*Materials and Design*

Similar to Belke's (2013) design, we utilized the same stimuli—in fact, the same experimental scripts—across the picture naming and classification tasks. For experiment 2, we simply changed the instructions, asking participants to verbally classify the pictured objects as natural vs. manmade. This parallel design eliminated concerns about item- and category-specific differences between the two tasks. As in the original study (Fink &

Goldrick, in prep), we note that the fixed 2-lag design made responses predictable within a given block of classification (e.g., a sequence of birds, fruits, and vehicles would receive the responses NATURAL - NATURAL - MANMADE - NATURAL - NATURAL - MANMADE). However, because the three blocks within pretest had different category configurations and therefore different response patterns, participants would not succeed if they simply repeated the same sequence throughout the entire task.

### 3.3.2 Results

Once again, for details of our initial RT trimming and outlier removal, we refer to the original study (Fink & Goldrick, in prep). Subsequent removal of word duration outliers (i.e., durations more than 3 standard deviations from participants' log-transformed means) eliminated an additional 0.7% of the pretest data.

*Model structure*

As in experiment 1, a single mixed-effects model was constructed to test for effects of response timing and response selection difficulty—or in this case, ease—on word durations during continuous picture classification. Note that here, we modeled variation in the duration of only two response candidates, "natural" and "manmade." The fixed effects structure utilized here was identical to experiment 1; the only difference is that we expected a negative effect of ordinal position within a category (1-5), indicating semantic facilitation. Random effects included intercepts for participants and items; by-participant slopes for ordinal position, block, trial-level RT, class and the ordinal position by block interaction; a by-item slope for RT interference size; plus correlations.

*Overall response speed*

In line with experiment 1, an effect of overall speed on durations showed that participants who were generally slower to respond also tended to produce longer responses ($\beta$=0.052, s.e.=0.016, $\chi^2(1)$=9.35, p<0.01; fig. 3.4). This pattern suggests that participants may coordinate production processes in a top-down fashion to achieve a steady processing pace from planning to articulation.

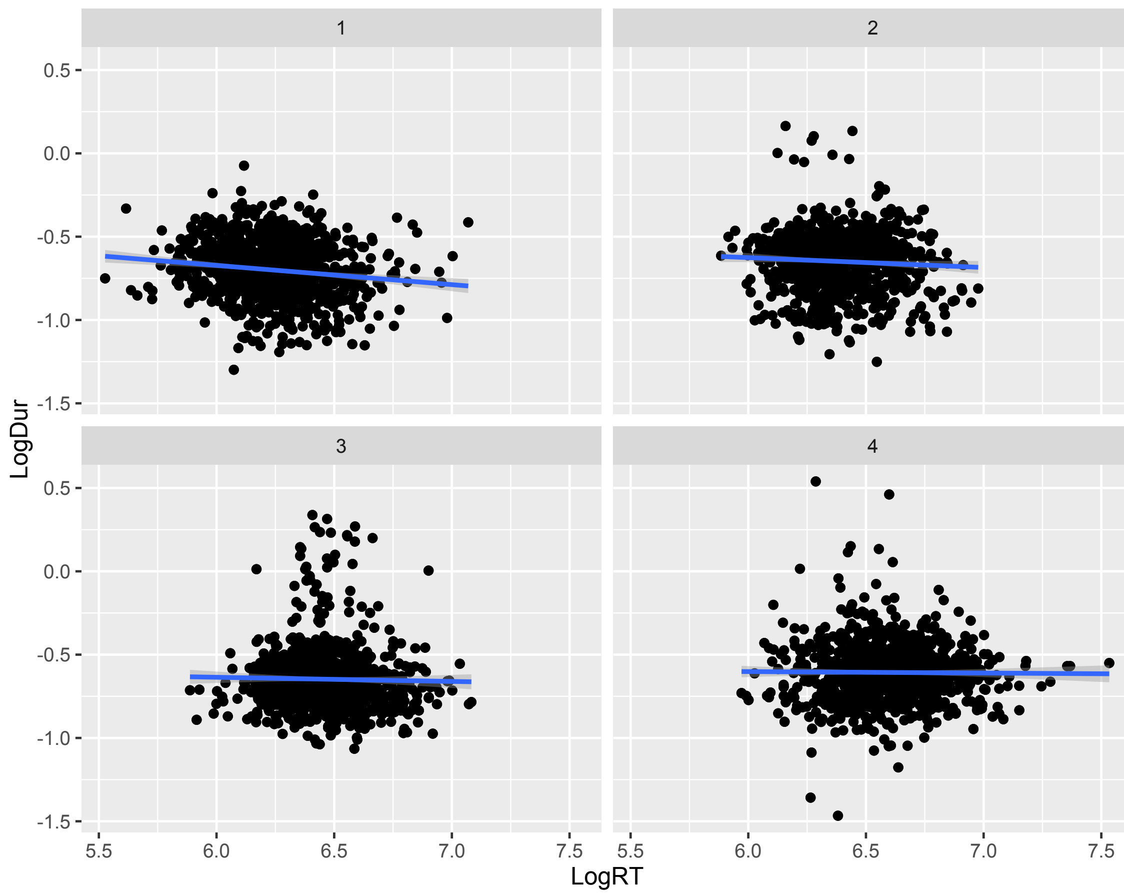



Figure 3.4 Participants divided into 4 quartiles, according to overall speed. Log-transformed durations (y-axis) plotted by log-transformed RTs (x-axis). The slowest

responders (quartile 4; bottom right) have longer durations than the faster responders (quartile 1; top left). Within trials, RTs and durations are negatively correlated (all panels).

*Trial-level RT*

In addition, a negative effect of trial-level RT revealed that faster responses were accompanied by longer word durations ($\beta$=-0.052, s.e.=0.012, $\chi^2(1)$=17.25, p<0.001; fig. 3.4). This finding is consistent with the non-significant pattern observed in experiment 1. According to our predictions, the direction of this effect supports the hypothesis that participants maintain a stable criterion for response initiation, allowing more time to meet that criterion when necessary, even if it detracts from articulation.

*Response selection difficulty*

A null effect of ordinal position within a semantic category showed that semantic facilitation did not impact word durations following response initiation ($\beta$=0.002, s.e.=0.002, $\chi^2(1)$=1.03, p=0.31). While participants responded faster across positions, exhibiting cumulative semantic facilitation in RTs (Fink & Goldrick, in prep), neighbor co-activation provided no processing benefit after target articulation had begun. Figure 3.5 shows that, if anything, durations patterned in the opposite direction than expected, becoming slightly longer in later positions. However, a marginal interaction of ordinal position and block suggests that any mild articulatory interference dissipated in later blocks ($\beta$=-0.004, s.e.=0.002, $\chi^2(1)$=3.70, p=0.054; fig. 3.6).

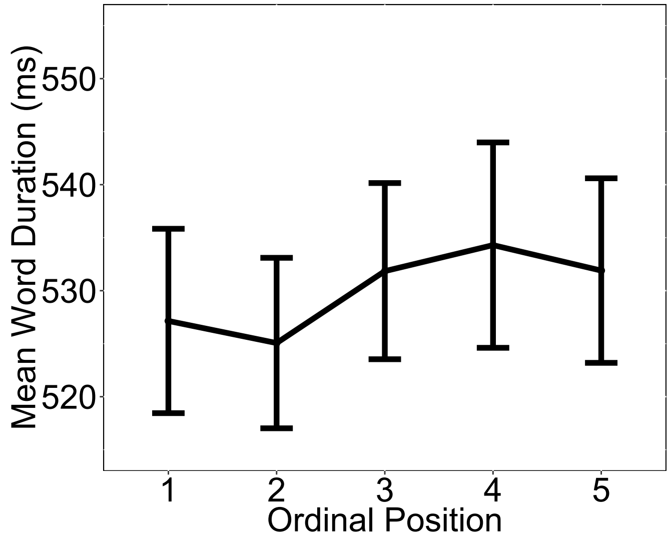



Figure 3.5 Durations did not change across ordinal positions within a category.

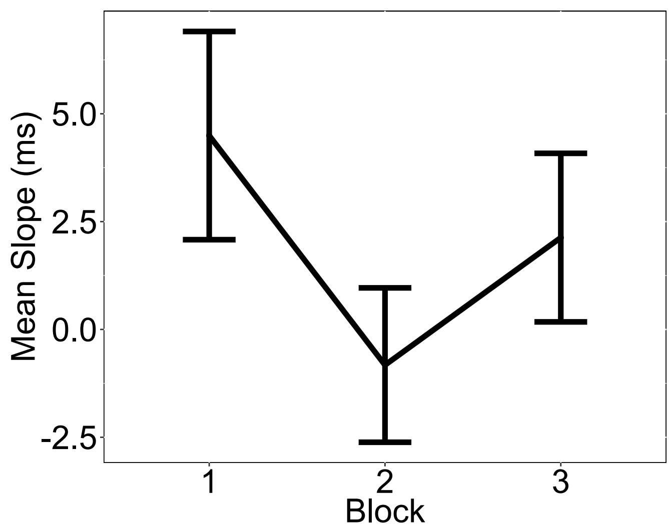



Figure 3.6 Positive slope in block 1 shows slight interference effect in durations. Near-zero slopes in later blocks confirm that this dissipates over time.

*Interactions with response selection difficulty*

Similar to experiment 1, the two-way interaction of ordinal position and the size of RT facilitation had no effect on word durations ($\beta$=-0.0004, s.e.=0.001, $\chi^2(1)$=0.20, p=0.66). This result indicates that the null effect of ordinal position was consistent across participants, regardless of how much facilitation they showed in RTs. The two-way interaction of ordinal position and trial-level RT also failed to reach significance, indicating that within trials, response timing did not reliably impact the ordinal position effect on durations ($\beta$=0.009, s.e.=0.006, $\chi^2(1)$=2.23, p=0.14).

*Control variables*

Finally, a marginal effect of class indicates a trend toward longer durations on "manmade" compared to "natural" responses, which is easily explained as a word length effect ($\beta$=-0.018, s.e.=0.010, $\chi^2(1)$=3.15, p=0.076). All remaining variables and interactions had non-significant effects on word durations ($\chi^2(1) \leq 1.77$, $p \geq 0.18$).

### 3.3.3 Comparison of experiment 1 and 2

*Analysis*

To contrast performance across production tasks, we conducted a combined analysis of experiments 1 and 2. Fixed effects were identical to those experiments, except that a contrast-coded predictor for production task (naming = 0.5, classification = -0.5) was included. This predictor was allowed to interact with each variable that had a significant impact on either of the previous models: by-participant speed, trial-level RT, ordinal position, class, block, and the ordinal position by block interaction. Random effects included intercepts for participants and items; by-participant slopes for ordinal position, block, trial-level RT, and class; by-item slopes for task and RT interference size; plus correlations between these terms (excluding the by-item slope for task, which was problematically correlated with the item intercept).

*Results*

For the sake of brevity, only the primary effect and interactions of the production task predictor are reported here. First, there was an overall effect of production task, such that durations were longer during classification than naming ($\beta$=-0.15, s.e.=0.03, $\chi^2(1)=28.03$, $p<0.001$). This effect may reflect a difference in the average segmental length of responses across tasks; the average response length was 5.21 phonemes in naming, but 6 phonemes in classification. Second, a marginal interaction of production task and trial-level RT confirmed that the negative RT effect was larger during classification than naming ($\beta$=0.038, s.e.=0.021, $\chi^2(1)=3.16$, $p=0.076$). In other words, early response initiation led to slower articulation in both tasks (trial-level RT: $\beta$=-0.035, s.e.=0.011, $\chi^2(1)=10.37$, $p<0.01$), but this pattern was more robust during classification. Finally, a significant three-way interaction between ordinal position, block, and task reflected how the ordinal position effect developed differently across tasks ($\beta$=0.009, s.e.=0.003, t=3.08, $p<0.05$[7]). Semantic interference grew across picture naming blocks in experiment 1, whereas initial signs of articulatory interference were reduced in later classification blocks in experiment 2.

**3.3.4 Summary**

Similar to experiment 1, experiment 2 (and the combined analysis) revealed that the timing of response initiation influenced articulatory processes in multiple ways during continuous picture classification. These results are particularly compelling, because they they patterned in opposite directions: while overall RT had a positive relationship with

[7] The nested model testing the significance of this interaction failed to converge. As a result, we adopt an alternate criterion for significance testing: a t-value over 2.0.

word durations, trial-level RT had a negative relationship. This stark contrast between global vs. within-trial effects of response timing implies that multiple mechanisms may help regulate a participant's decision criterion for response initiation.

Unlike experiment 1, the results of experiment 2 provide no evidence that the dynamics of response selection influenced articulation after response initiation. Based on our predictions, this is rather unsurprising. In the introduction, we argued that semantic neighbor co-activation might directly influence articulation after response initiation if additional time was needed to resolve it. During classification, no such resolution is required, because co-activation facilitates planning rather than interfering with it. In other words, these results confirm that co-activation is more likely to affect articulatory processing after response initiation if it hinders lexical selection (experiment 1) than if it supports it (experiment 2).

### 3.4 Experiment 3: Blocked Cyclic Picture Naming

Our final experiment utilized the blocked-cyclic naming paradigm (Belke et al., 2005) to investigate the relationship between response selection, response initiation, and articulation. Because this block-wise design is more explicit than the continuous one (experiment 1), participants may be able to engage top-down control to resist semantic interference (Belke & Stielow, 2013). Within any block, participants repeatedly name a limited set of items across several cycles, and this constrained response set may allow them to inhibit non-target responses. As a result, experiment 3 allows us to test the coordination of planning and articulatory processes that are subject not only to semantic interference, but also to control processes intended to moderate that interference.

### 3.4.1 Methods

*Participants*

96 undergraduate students were recruited from the Psychology Department subject pool at the University of California San Diego. All reported to be native speakers of English, with no history of language or psychological disorder, and normal or corrected-to-normal vision and hearing. 2 participants were excluded from the analyses below due to a technical error in their response recordings.

*Materials and Procedure*

Similar to experiments 1 and 2, the data reported here were originally collected for a separate study of semantic interference effects on RTs (Oppenheim, in prep). The experiment began with a familiarization phase, where participants were presented with a randomized series of 96 line drawings for naming, including 72 critical items (6 members each of 12 categories) and 24 fillers. Items were randomly organized into blocks of 24 trials, each followed by a self-paced break.

In the test phase of the experiment, participants encountered the same 96 line drawings, this time organized into mixed vs. blocked semantic contexts. Mixed contexts were composed of 6 items belonging to different semantic categories (e.g., *dog*, *arm*); blocked contexts contained 6 category-related items (e.g., *dog*, *goat*). Each block began with a randomly selected filler item, followed by six critical items that were repeated in different orders across 4 cycles (24 test trials per block). Trials were self-paced with a time-out function, and self-paced breaks were offered between blocks.

The test phase was divided into two halves, each containing 12 blocks of picture naming. Participants were randomly assigned to one of two experimental formats, which differed in the composition of these halves. In format 1 (experiment 2, Oppenheim, in prep), mixed and blocked contexts alternated throughout both halves. In format 2 (experiment 3, Oppenheim, in prep), each context type was grouped together in the same half of the experiment.

*Data Pre-Processing*

RTs were collected using a voice-key (Oppenheim, in prep). Word durations were automatically extracted from stereo recordings using the procedure outlined in experiment 1, except that boundaries were not hand-corrected due to the large scale of the data set (over fifty thousand observations).

### 3.4.2 Results

An initial data cleaning process removed incorrect responses, recording errors, and RT outliers (Oppenheim, in prep). We then excluded durations below 100 milliseconds and above 1000 milliseconds and log transformed the data to compensate for positive skew. Outlier trimming removed data points more than 3 standard deviations away from each subject's mean log duration, eliminating an additional 1.0% of trials.

*Model structure*

Fixed effects of interest included by-participant overall speed, trial-level RT, semantic context (blocked=-0.5 vs. mixed=0.5), by-participant RT interference size, and the two-way interactions of semantic context by trial-level RT and semantic context by RT interference. If semantic interference impacted articulation in this task, we expect longer durations in semantically blocked contexts compared to mixed contexts. Additional fixed effects were added as control variables: experimental format (alternating blocks=0.5 vs. blocked halves=-0.5), experimental block (1-12), cycle of naming trials within each block (1-4), the number of trials since the picture was last named (repetition lag, 1-9), and whether an item was named in a previous block. The two-way interactions of semantic context with block, cycle, and repetition lag were also allowed. Random effects included intercepts for participants and items; by-participant slopes for block, cycle, whether an item was previously named, trial-level RT, and semantic context; and by-item slopes for experimental format, block, trial-level RT, and by-participant speed.

Unlike in experiments 1 and 2, the current by-participant predictors were generated using BLUPs from the full mixed effects model of participants' RT data, rather than beta coefficients from simple regressions. As noted above, BLUPs from mixed effects models surpass other regression techniques because each individual estimate is made in light of the entire data set; the random effects are assumed to reflect samples from a larger population (Baayen, 2008). This approach was possible in current experiment because there was more variation across participants in the target semantic effect.

*Overall response speed*

Consistent with experiments 1 and 2, we observed an effect of by-participant speed, such that generally slower responders produced longer durations ($\beta$=0.50, s.e.=0.024, $\chi^2$=4.42, $p<0.05$; fig. 3.7). Apparently this global effect of response timing on articulation holds across a variety of production tasks, including those involving semantic interference (experiments 1 and 3), semantic facilitation (experiment 2), and the potential for top-down cognitive control (experiment 3).

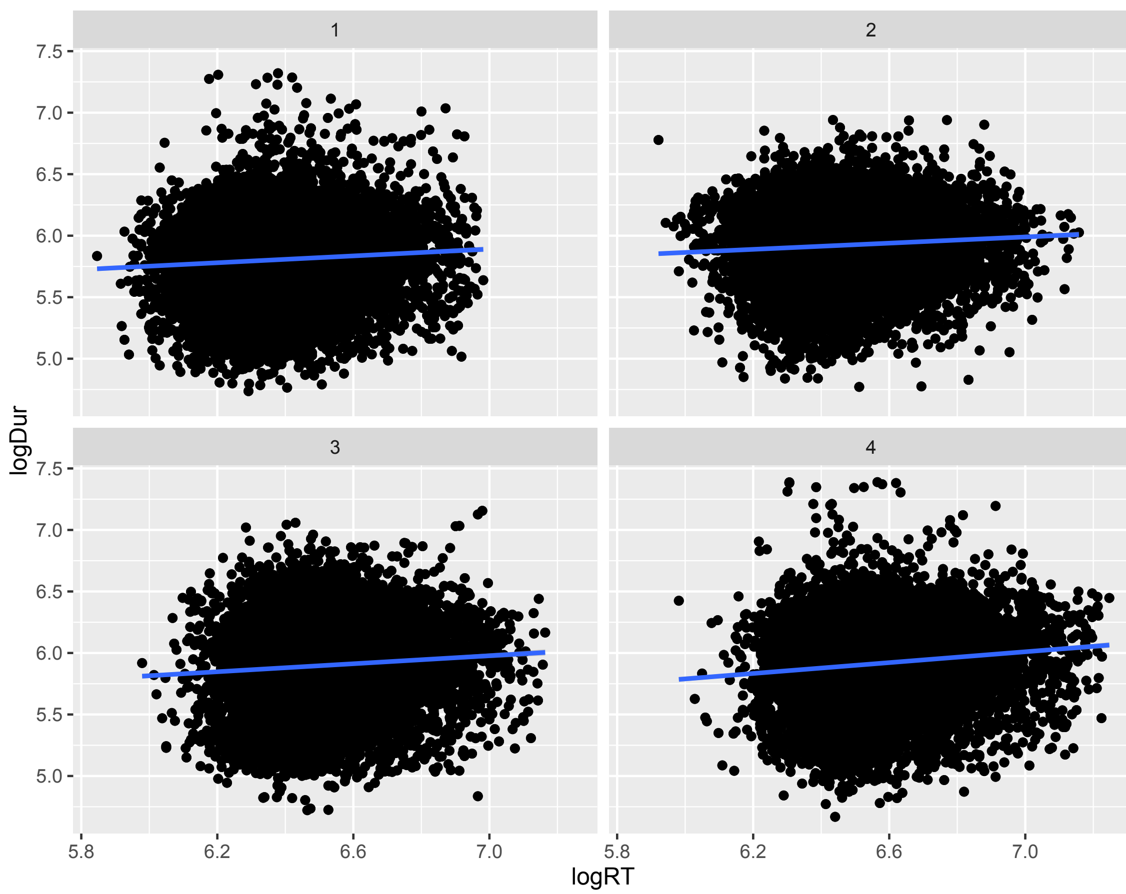



Figure 3.7 Participants divided into 4 quartiles, according to overall speed. Log-transformed durations (y-axis) plotted by log-transformed RTs (x-axis). Slower responders (quartiles 3 & 4) have longer durations than the faster responders (quartiles 1 & 2). Within trials, RTs and durations are positively correlated.

*Trial-level RT*

Intriguingly, trial-level RT also had a significant effect on durations, but it patterned in the opposite direction from the previous two experiments: within trials, faster responses tended to exhibit shorter word durations (β=0.055, s.e.=0.014, $\chi^2$=14.23, p<0.001; fig. 3.7). According to our predictions, this result supports the hypothesis that participants apply similar real-time adjustments to the pace of both planning and articulation, depending on the relative ease or difficulty of target selection. Under this account, both RTs and durations are longer when lexical selection is difficult, because participants allow time for more cautious response selection and execution.

*Response selection difficulty*

We found no overall effect of semantic context on word durations (β=0.0001, s.e.=0.003, $\chi^2$=0.003, p=0.96). Regardless of whether pictures were grouped by semantic category or intermixed, durations did not vary. This finding creates an interesting contrast between experiments 1 and 3: while cumulative semantic interference had a group-level effect on durations, semantic interference during blocked cyclic naming did not. We examine this contrast in the General Discussion.

*Interactions with response selection difficulty*

Three different variables modulated the null effect of semantic context. First, there was an interaction between semantic context and RT interference size (β=0.53, s.e.=0.16, $\chi^2$=10.47, p<0.01). As shown in figure 3.8, participants who were more sensitive to the context manipulation, as indexed by the size of their RT interference effects, were more likely to show similar interference during articulation. In other words, if lexical co-activation was especially intense during planning, it was more likely to continue affecting processing after response initiation.

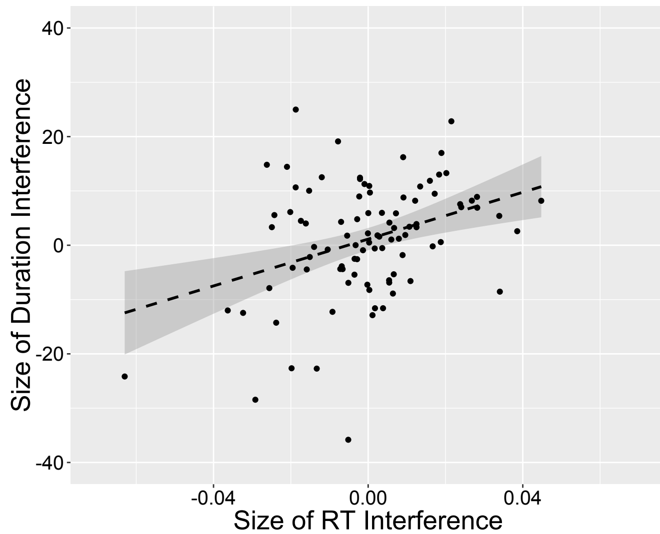



Figure 3.8 Best unbiased linear predictors (BLUPs) were extracted from a model of RT data to estimate participants' semantic interference effects during planning. These reliably predicted the presence of semantic interference in durations.

In addition, trial-level RT weakly interacted with semantic context, such that the context effect on durations was larger on slow responses (β= 0.018, s.e.=0.011, $\chi^2$=2.70, p=0.10). This pattern suggests that on trial with strong lexical selection difficulties (reflected by longer RTs), those difficulties are more likely to continue affecting processing after response initiation than on trials with easy lexical selection. Even if participants adjust their response decision criteria to allow more time for the resolution of neighbor co-activation, that co-activation may continue after articulation has begun.

Finally, an interaction emerged between semantic context and whether or not an item had been named in a previous block (β=-0.008, s.e.=0.003, $\chi^2$=5.22, p<0.05). This interaction appears to be driven by a small effect of semantic context in the first half of the experiment, when targets have not been named in a previous block (half 1: mean context effect =3.09 ms, s.e.= 3.57 vs. half 2: mean effect =-0.75 ms, s.e.=3.4).

*Control variables*

One control variable influenced word durations. Consistent with previous work (e.g., Bell et al., 2009), an effect of repetition lag indicated the presence of repetition priming—naming *cat* reduced the duration of subsequent namings of that *cat*—which dissipated over time ($\beta$=0.0013, s.e.=0.0004, $\chi^2(1)$=11.79, $p<0.001$). All remaining effects failed to reach significance ($\chi^2 \leq 2.44$, $p \geq 0.12$).

**3.4.3 Summary**

In terms of response timing predictors, experiment 3 was only partially consistent with experiments 1 and 2. As above, the global measure of response timing had a positive effect on articulation; however, we observed a negative effect of trial-level RT, which diverged from our earlier results. In the General Discussion, we offer a methodological explanation for these contrastive results. Interestingly, this primary effect of trial-level RT was accompanied by an weak interaction of semantic context, such that context effect was larger on slow responses. This finding suggests at least a partial dependency between planning and articulatory disruptions in this task.

The results pertaining to response selection difficulty also diverged from the previous experiments. Unlike in experiment 1, which involved a similar but distinct form of semantic interference, we observed no overall effect of semantic context on articulation. This contrastive result may reflect engagement of top-down cognitive control in the current blocked naming task (Belke & Stielow, 2013), which could dampen the influence of co-active neighbors and prevent them from influencing processing after response initiation. However, an interaction between semantic context and RT interference size revealed that a subset of participants did exhibit interference in durations. Our ability to detect this interaction most likely stems from the presence of substantial variability in the context effect size; retention of the random by-participant slope for semantic context during our model selection procedure supports this interpretation. Indeed, the difference in variability across experiments 1 and 3 may have contributed to their divergent effects of response selection difficult on articulation. Regardless, this result demonstrates how consideration of individual variation can enhance our ability to detect long distance interactive effects. In addition, it provides support for the flexible cascade hypothesis, which proposes that the degree of functional interactivity in the production system is not fixed, but may vary across individuals (as well as tasks and trials) depending on the properties of the cognitive processes underway (Kello et al., 2000; Kello, 2004).

## 3.5 General Discussion

The overarching goal of this study was to examine whether the timing and coordination of planning and articulatory processes generates variation in the articulation of target productions (as assessed via word durations). This question was motivated by the flexible cascade hypothesis, which assumes not only that articulation may begin before planning is complete, but also that the relative timing of response initiation is dynamically determined by several possible factors, including response selection difficulty. Over the course of three experiments, we used semantic neighbor co-activation to manipulate the difficulty of lexical selection, and the results revealed a range of evidence supporting and constraining this hypothesis. The key findings and their implications for theories of spoken production are presented below.

### 3.5.1 The timing of response initiation matters

Our most robust finding, which appeared in all three experiments, was a positive relationship between participants' overall response speeds and their word durations. Participants who were generally slower to initiate their responses also articulated their responses more slowly. As stated in the discussions above, this pattern suggests that speakers maintain a fairly steady, global speed of processing throughout spoken production. Regardless of whether lexical selection is relatively difficult (experiments 1 and 3) or easy (experiment 2), this parallelism between planning and articulation remains.

However, if we take a finer-grained approach and examine the relationship between response initiation and articulation within trials, a more complex story emerges. In experiments 1 and 2, we discovered a negative relationship between response timing and articulation, such that faster trial-level RTs were coupled with longer word durations. This pattern indicates a trade-off between the processing time invested in planning vs. articulation. In other words, it seems that speakers partitioned the available response window between speech preparation and execution. According to our predictions, this trade-off should be influenced by the relative difficulty of lexical selection, such that planning takes priority when selection is more difficult. This assumption is somewhat validated, given that different effect sizes emerged across tasks evoking semantic interference (experiment 1; $\beta$=-0.019) vs. semantic facilitation (experiment 2; $\beta$=-0.052).

However, the observation of a negative trial-level RT effect in both tasks suggests that the timing of response initiation is also somewhat independent of lexical selection difficulty.

Importantly, the negative effect in experiments 1 and 2 contrasts with the results of experiment 3, where blocked-cyclic naming participants showed a positive effect of trial-level RT on durations. Trials with faster responses, reflecting earlier response initiation, yielded shorter word durations in this task. According to our predictions, which again assumed that the timing of response initiation is contingent on response selection difficulty, this result indicates that speakers adjust the timing of both planning and articulation to accommodate lexical difficulties in the same way. However, as we explain below, additional factors may have influenced this response strategy as well.

Specifically, we propose that a post hoc methodological explanation unites the surprisingly similar results across experiments 1 and 2 with the contrasting results in experiment 3: namely, trial composition. In experiments 1 and 2, participants encountered a fixed response window of 2 seconds, which essentially forced them toward the strategy of partitioning that window between planning and articulatory processes. On the other hand, experiment 3 provided self-paced trial advancement (with a 3 sec deadline), allowing participants the flexibility to speed through relatively easy trials and spend more time on difficult ones. Such a design is more conducive to the sort of sliding scale strategy observed among experiment 3 participants, where they adjusted both planning and articulatory time in the same way, depending on trial difficulty.

Additionally, the overall design contrast between the continuous and blocked naming tasks might have contributed to their contrasting trial-level RT results. As we highlighted earlier, the blocked-cyclic naming task in experiment 3 allows participants the opportunity to apply top-down cognitive control, inhibiting non-targets from the restricted response set in each block (Belke & Stielow, 2013). Presumably, the application of inhibition to co-active lexical representations not only supports planning processes, but also articulation, due to cascade. This shared benefit could contribute to the syncing of RTs and durations observed in experiment 3.

Taken together, the results in this section have provided great insight into the coordination of planning and articulation during speech production. The finding of independent effects of overall RT and trial-level RT suggests that multiple mechanisms may contribute to a speaker's decision about when to respond. This multi-mechanism proposal is further strengthened by the results of experiments 1 and 2, where the global and local RT effects patterned in opposite directions. Such an argument parallels recent developments in the field of domain-general cognitive control, where a dual mechanism framework postulates the existence of a sustained, proactive control mechanism alongside one that operates in a transient, reactive fashion (Braver, 2012). Future production research should bear in mind that the decision to initiate speech can be influenced by numerous factors; only some of these may be theoretically important, but all of them are combined into the unitary RT measure.

### 3.5.2 Individual variation can mask direct lexical effects on articulation

In addition to testing the role of response initiation in acoustic variation, the current study also investigated whether the difficulty of lexical selection had any direct impact on articulation. By definition, the flexible cascade hypothesis argues that disruptions during lexical selection can be transmitted to subsequent stages of processing by cascading activation. This could manifest not only as effects on the initiation of articulation (such that all lexical effects are mediated by the relationship between RT and articulation), but might also have effects after response initiation (yielding independent effects of lexical access difficulty on duration).

Our results follow previous work in their inconsistency (Kello et al., 2000, and Kello, 2004 vs. Damian, 2003). In experiment 1, we found an overall effect of semantic interference on word durations, such that words became longer across positions within a category. Because this effect emerged in a model with several RT covariates (overall speed, trial-level RT, and RT interference size), it constitutes an unmediated effect of lexical selection difficulty on articulation. In contrast, we found no overall effect of semantic interference during blocked naming (experiment 3), nor an effect of semantic facilitation on durations during classification (experiment 2). Based only on first order effects, these two experiments provide no evidence of lexical effects on articulation following response initiation.

However, a significant interaction in experiment 3 revealed that individual variation in experimental sensitivity masked the presence of semantic interference in durations. Specifically, an interaction of RT interference size and semantic context showed that semantic context did impact word durations, but only among the subset of participants with large RT interference effects. In short, this result demonstrates that subtle interactive effects of lexical disruptions on articulation may only be detected when those disruptions are fairly robust to begin with. This interpretation is reinforced by the absence of any lexical effects on articulation in experiment 2, where neighbor co-activation facilitated lexical selection rather than disrupting it. On the whole, these findings suggest that other literatures fraught with inconsistent evidence and/or null results may benefit from consideration of individual variation.

### 3.6 Conclusion

The current study explored a classic question among theories of spoken production: what are the sources of acoustic variation in spoken word forms? Previous work has established that the representation and/or processing of linguistic representations can generate such variation, as can the conversational context in which a target is uttered. Using three semantic neighbor manipulations, we demonstrated a third source of variation: the timing and coordination of production processes. In three experiments, both overall and trial-level effects of RTs on word durations revealed that the relative timing of response initiation has significant consequences for subsequent articulation. Moreover, the independence of these global and local effects suggests that multiple mechanisms may contribute to a speaker's decision to begin articulating an utterance. This study also demonstrated direct effects of semantic interference on word durations, which could not be explained by RT variation alone. Although these effects were small (experiment 1) or detectable only after considering individual variation (experiment 3), they confirmed that lexical properties can continue influencing articulation after a response has been initiated. Additional research that simultaneously analyzes RT and articulatory/acoustic effects is needed to help constrain the complex empirical landscape of interactive effects, and also to clarify the role of methodological choices in their generation.

CHAPTER 4

## 4.1 Introduction to Study 3

Spoken production necessarily begins with conceptualization of the message a speaker wishes to convey. For instance, a pet owner intending to name their favorite type of pet—dog—would first activate a representation of the intended concept in semantic memory (e.g., a set of features like <furry>, <four-legged>, <canine>, and <pet>, or a holistic representation of meaning; Alario & del Prado Martin, 2010; Belke, 2013; Howard, Nickels, Coltheart, Cole-Virtue, 2006; Oppenheim, Dell, & Schwarz, 2010). Following conceptualization, these semantic representations feed activation to the production system, triggering retrieval of the lexical representation for DOG. From there, the sound structure of the selected word is retrieved during phonological (/d/-/ɔ/-/g/) and then phonetic ([dɔg]) encoding, providing a fully specified speech plan for articulation. Despite the central importance of conceptualization as the starting point of speech, this early process and the representations it operates over are generally treated as peripheral to theories of spoken production. Either simplifying assumptions are made about the structure of semantic representations (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999), or multiple structures are shown to be equally viable, leaving the theory agnostic (e.g., Howard et al., 2006; Oppenheim et al., 2010). The current study questions these assumptions, asking what is the structure of the semantic memory representations that feed lexical access?

Such questions have only grown in importance with the rise of semantic relatedness manipulations as a tool for probing the architecture and mechanisms of the production system. Prominent paradigms include semantic priming (Carr, McCauley, Sperber, and Parmelee, 1982; Wheeldon & Monsell, 1994), picture-word interference (Rosinski, Golinkoff, & Kukish, 1975), semantically blocked naming (Kroll & Stewart, 1994; Damian, Vigliocco, Levelt, 2001), blocked cyclic picture naming (Belke, Meyer, & Damian, 2005), and continuous picture naming (Howard et al., 2006). Indeed, Alario and del Prado Martin (2010) utilized one of these paradigms—continuous picture naming—to investigate the structure of semantic memory representations. The current study seeks to replicate and extend their work.

### 4.1.1 Supracategory effects in continuous picture naming

Specifically, Alario and del Prado Martin (2010) reanalyzed the data from Howard et al.'s (2006) original study featuring the continuous picture naming paradigm, which has gained popularity among spoken production researchers for its simple design and implicit semantic manipulation. In this paradigm, participants are asked to name of series of pictured objects as quickly and accurately as possible. Unbeknownst to them, the objects are drawn from a range of semantic categories, whose members are staggered through the stimulus list. A number of studies have demonstrated that this subtle manipulation generates cumulative semantic interference, or a linear increase in response times (RTs) across ordinal positions within a given category (e.g., Belke, 2013; Belke & Stielow, 2013; Howard, et al., 2006; Navarrete, Mahon, Caramazza, 2010; Oppenheim et al., 2010; Schnur, 2014). Furthermore, if the same stimuli are utilized in a continuous classification task, the reverse pattern is observed: classification responses become monotonically faster across positions within a category (Belke, 2013; Fink & Goldrick, in prep).

Alario and del Prado Martin utilized linear mixed effects modeling to reanalyze Howard et al.'s original continuous naming data. This modeling technique captured not only the overall effect of semantic interference (a fixed effect of ordinal position within the set of items in a semantic category), but also variation in this effect across randomly selected variables like subjects and items (random effects). This variation might be due to intrinsic differences between different participants or items, or might reflect the influence of other variables that the fixed effects do not account for. Alario and del Prado Martin pursued this latter possibility, hypothesizing that the semantic categories artificially constructed for the continuous naming task are represented in memory as part of a hierarchical representational structure. Specifically, they proposed that some of the categories might be semantically related to one another (similar to the relationship among members of any given category), forming higher level supracategories. This account predicts that semantic interference might accumulate *across* co-categories from the same supracategory, not just within them. Concretely, it predicts independent effects of an item's ordinal position within a category and its ordinal position within the supracategory.

The authors tested this prediction in two ways. First, they built random effects for semantic category into their regression model and tested for a significant improvement in model fit. This included a random intercept, allowing mean RT to vary across categories, and a random slope for the ordinal position effect, allowing the size of semantic interference to vary as well. They expected that if a supracategory structure was latently present in the stimuli set, then the random effects would help capture variance that had previously been unaccounted for. Their results confirmed this prediction, showing that both random category effects improved their model of participants' continuous naming performance. To confirm that this variation is not simply due to intrinsic differences between semantic categories, Alario and del Prado Martin subjectively identified pairs of potential co-categories in the stimuli set; they extracted that subset of the data, then explicitly built the supracategory structure into a regression model. Namely, they created a contrast-coded predictor for whether the current category's co-category had previously appeared. The authors expected to observe a positive effect of previous co-category naming, reflecting the carryover of previously built up interference. The results revealed independent effects of ordinal position, reflecting semantic interference, and previous co-category naming, suggesting the spread of interference across related co-categories. Alario and del Prado Martin interpreted these findings as evidence for hierarchical semantic representations, rather than a one-dimensional layer of (featural or non-decompositional) representations.

The remainder of this paper is structured as follows. First, we present a selection of recent studies that induced lexical co-activation using semantic neighbor manipulations, to demonstrate the gradiency and diversity of the underlying semantic relationships. Next, the methods section reviews the experimental design of our original continuous naming and classification tasks (Fink and Goldrick, in prep), which differed in important ways from Howard et al.'s (2006) paradigm. Finally, we walk through the results of two analyses. The first explored variation across categories using random effects, while the second used fixed effects to test for an impact of previous co-category naming. We discuss the implications of these results for theories of spoken word production.

### 4.1.2 The structure of representations in semantic memory

As noted above, numerous production studies have employed semantic relatedness manipulations to investigate diverse issues in the field. A few of these studies are particularly relevant to the current investigation, because they demonstrate gradiency and variation in the effects resulting from different kinds of semantic relationships. Such data provide evidence that not all semantic relationships are equal, suggesting that the architecture of semantic memory must somehow encode that diversity. We argue that this complex empirical landscape of relatedness effects might be better understood if explanatory models more carefully considered the memory structures giving rise to them.

One variable modulating semantic neighbor effects is the semantic distance between target items/categories. Vigliocco, Vinson, Damian and Levelt (2002) demonstrated this phenomenon during a blocked picture naming experiment. They selected three semantic categories for their task; two of the categories were near one another in semantic similarity space (*clothing* and *body parts*), while the third was far away from the other two (*vehicles*). The authors then constructed three types of blocks. In “same field” blocks, participants named items from a single semantic category. In “near field” blocks, they named items from the two categories in a near semantic relationship, while in “far field” blocks, they named items from two categories in a far semantic relationship. The results revealed graded effects of semantic distance, such that RTs were slowest in same field blocks, intermediate in near field blocks, and fastest in far field blocks. These data indicate that semantic relatedness is a continuous rather than binary variable, which modulates the intensity of semantic neighbor co-activation and therefore the strength of neighbor effects. Given the observation of similar effects in other production tasks (e.g., picture-word interference: Mahon et al., 2007; Vigliocco, Vinson, Lewis & Garrett, 2004; lexical decision & action naming: Vigliocco et al., 2004), such gradiency seems like a general attribute of the semantic representations that support production.

Another factor influencing semantic effects during spoken production is the type of semantic relationship present between items. The current study focuses on the behavioral consequences of co-activating semantic neighbors, i.e., items that belong to the same (basic or other level) semantic category. However, other related items, like the semantic associates of a target, can also influence processing. For example, Alario, Segui, and Ferrand (2000) used a semantic priming paradigm to demonstrate that prime words in categorical vs. associate relationships have contrasting effects on picture naming performance, as a function of stimulus-onset asynchrony (SOA). With a short SOA between the prime word and target picture, category-related primes caused an interference effect (slower RTs with related compared to unrelated primes), but associates had no effect. With a longer SOA, a different pattern emerged: category-related primes had no impact, while associates triggered facilitation (faster RTs with related compared to unrelated primes). Alario et al. concluded that these results reflect the existence of two distinct types of meaning-based relationships.

This short review demonstrates that semantic relatedness is gradient and that the effects of different types of semantic relatedness are heterogeneous in nature. Given the under-specification of the lexico-semantic network in most production theories (e.g., Dell, 1986; Howard et al., 2006; Levelt et al., 1999; Oppenheim et al., 2010), they do not actively predict these empirical findings.

In the current study, we explore Alario and del Prado Martin's (2010) proposal that the incorporation of hierarchical structure within semantic memory representations might improve the predictive power of contemporary theories. We explore this proposal by testing whether the semantic neighbor effects we observed during a previous study might reveal evidence for multi-dimensional semantic representations if we revisit them with more detailed analyses. In particular, we follow Alario and del Prado Martin (2010) in performing post hoc analyses on a set of continuous naming—and also continuous classification—data gathered for a separate investigation of lexical selection processes. These analyses test for supracategory effects above and beyond the primary effects of cumulative semantic interference and facilitation. Similar to Alario and del Prado Martin, our stimuli were not designed to address the structure of semantic representations. Nonetheless, we incrementally considered random and then fixed effects for our artificial semantic categories, seeking a deeper understanding of the representations underlying the neighbor effects reported in our original study (Fink & Goldrick, in prep). The details of our experimental materials and post hoc analyses are presented in the next two sections.

## 4.2 Methods

*Participants*

For each of two experiments, 90 participants (180 total) were recruited from the Linguistics Department subject pool at Northwestern University or from on-campus flyers.

They received course credit or $10 compensation for their time, respectively. Participants were native English speakers who reported no history of cognitive impairment.

*Materials and design*

*Continuous picture naming.* In the first experiment, a continuous picture naming task was inserted into the negative transfer paradigm, designed to test whether a target cognitive processes is shared across two tasks (Persson, Welsh, Jonides, & Reuter-Lorenz, 2007; Nelson, Reuter-Lorenz, Persson, Sylvester, & Jonides, 2009). This paradigm uses a pre- vs. posttest design (without item repetition): baseline performance in task A is assessed at pretest, task B is intensively practiced during treatment, and then posttest performance on task A is analyzed for treatment effects. In this experiment, reported by Fink and Goldrick (in prep), participants performed continuous naming at pre- and posttest, being randomly assigned to one of two cognitive control treatment tasks. To avoid contamination of the naming data by the treatment tasks, only pretest data are analyzed here. Nonetheless, we describe the construction of both pre- and posttest materials, in order to explain the limitations of the current data set.

90 colorized line drawings were selected from Rossion and Pourtois' (2004) database. These images represented 5 items each from 18 semantic categories (see Appendix), which had an average word frequency of 65.3 words per million (SUBTLEX database; Brysbaert & New, 2009). A master stimulus list was then constructed, placing 9 categories in pretest and 9 in posttest to avoid item repetition within participants. Within each test, the 9 target categories were divided into 3 experimental blocks. Within each block, items were randomly drawn from the 3 semantic categories in rotation and without

fillers. For example, a block containing birds, fruits, and vehicles might begin with the items OWL, APPLE, CAR, PEACOCK, ORANGE, and PLANE. As a result, items from the same semantic category were presented with a consistent lag of 2 intervening items between them. This consistent short-lag design deviates from Howard et al.'s study, where lag varied from 2-8 trials, but recent work has shown that it produces comparable semantic interference effects (Runnqvist, Alario, Strijkers, & Costa, 2012; Schnur, 2014). This lack of fillers may have made the semantic category manipulation more explicit than in the standard paradigm. However, anecdotal evidence from participant debriefings indicated that they were not particularly sensitive to the manipulation.

Next, the master stimulus list was rearranged to create 9 additional versions, allowing us to counterbalance the assignment of items across tests and ordinal positions within a category. Between participants, every item appeared once in each ordinal position at both pre- and posttest. Participants were then randomly distributed across the 10 stimulus lists until 3 participants had completed each version.

Within trials, our design exactly followed Howard et al.'s (2006). A fixation cross appeared in the center of the screen for 500 ms, followed by a blank interval of 250 ms before the target appeared onscreen. The target image remained visible throughout a 2000 ms response interval, during which the participant named the item aloud into a head-mounted microphone. The screen then blanked for a 500 ms inter-trial interval before advancing automatically to the next trial. Pre- and posttest each took just over 3 minutes to complete.

*Continuous classification.* Following Belke (2013), we reused the same stimuli—in fact, the same experimental scripts—for the continuous classification task. This task is known to generate cumulative semantic facilitation, rather than interference, such that RTs decrease linearly across ordinal positions within a semantic category (Belke, 2013). Given this parallel design, the only difference between the production tasks (besides the contrasting neighbor effects they elicit) was a change of instructions, which now required participants to verbally classify the pictured objects as natural vs. manmade rather than naming them. This design allowed us to rule out item- and category-specific differences between the two production tasks.

We note that our fixed 2-lag design had greater ramifications during classification than naming. This structure meant that responses were predictable within each block, such that the aforementioned sequence of birds, fruits, and vegetables would elicit the responses NATURAL-NATURAL-MANMADE-NATURAL-NATURAL-MANMADE. As a result, participants may have developed a more automatic response strategy, repeating this sequence across the five ordinal positions. They could not simply repeat the same sequence throughout the pre- or posttest, because each block had a difference category configuration and therefore response pattern. Nonetheless, this strategy may have impacted the target production processes, particularly on trials in the first ordinal position (i.e., at the start of an experimental block), where any learned response sequence would need to be reset.

### 4.3 Analysis 1: Is there Significant Variation across Semantic Categories?

As a first step towards detecting any hierarchical structure in our stimuli set, we built random effects for semantic category into baseline models of the pretest RT data from each production task. Random effects provide individual adjustments to a fixed effect, allowing the model to capture variation in the target effect size across participants, items, or in this case, categories. First, we added a random intercept for category, which allowed the model to detect variation in mean RT across categories. Then, we incorporated a random slope for the semantic neighbor effect (interference or facilitation), allowing the size of the effect to change across categories. Finally, we included random correlations between those two terms, allowing the relationship between mean RT and semantic effect size to vary across categories. Every time we added a new parameter to the model, we tested for a significant improvement in model fit using nested model comparisons.

### 4.3.1 Continuous picture naming

*Baseline RT model*

Following the model selection procedure outlined by Bates, Kliegl, Vasishth, and Baayen (2015), we arrived at a baseline model of the RT data from the continuous naming pretest. This model included ordinal position (1-5), block (1-3), and their two-way interaction as fixed effects. Random effects included intercepts for participants and items, a by-participant slope for block, and a by-item slope for ordinal position.

The pretest model confirmed that participants experienced cumulative semantic interference, such that RTs increased linearly with each ordinal position in a semantic category ($\beta=0.024$, s.e.=0.003, $\chi^2(1)=34.78$, $p<0.001$). This effect shrank across the three pretest blocks (ordinal position x block: $\beta=-0.012$, s.e.=0.004, $\chi^2(1)=12.01$, $p<0.001$). The overall effect of block was not significant ($\chi^2(1)=1.17$, $p=0.28$).

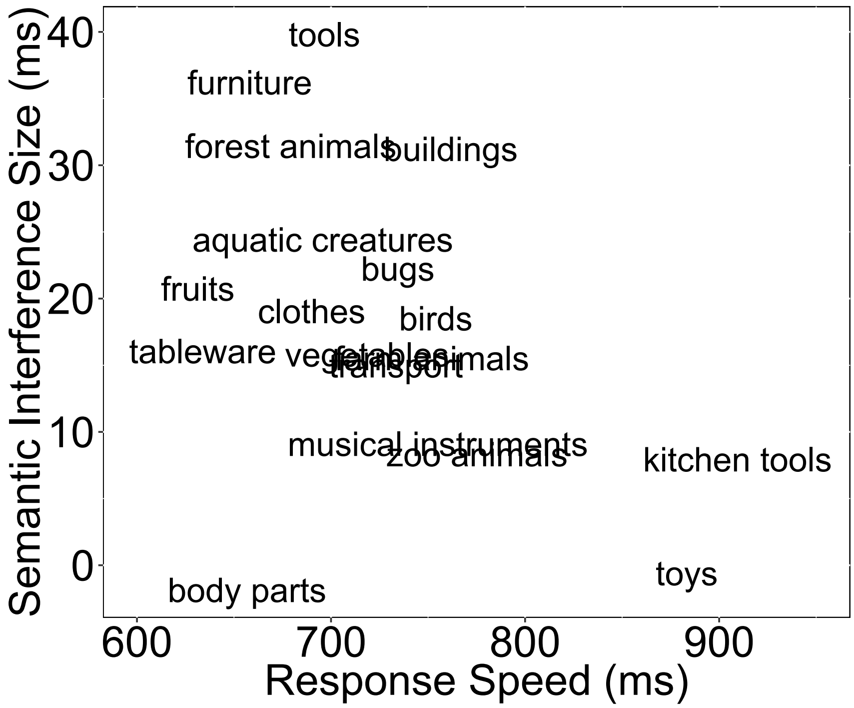



Figure 4.1 Mean RT and semantic interference size, estimated for each category using simple linear regressions. Significant variation in speed, but not interference size, is observed.

*Categories vary in speed but not semantic interference*

Inclusion of a random intercept for semantic category significantly improved the RT model's fit ($\chi^2(1)=4.62$, $p<0.05$). As illustrated in figure 4.1, the mean RTs for different categories covered a range of more than 200 ms. However, addition of a random slope for ordinal position had no impact on the model fit ($\chi^2(1)=0$, $p=1$). Figure 4.1 demonstrates how the size of semantic interference spanned a much smaller range of variation across categories, from an essentially null effect (*body parts*, *toys*) to a 40 ms effect (*tools*). This result parallels our finding of negligible variation in semantic interference across participants; the by-participant slope for interference was similarly attributed with 0 variance. Finally, allowing a correlation between the by-category intercept and ordinal position slope also had a non-significant impact on model fit ($\chi^2(1)=0.37$, $p=0.55$).

### 4.3.2 Continuous picture classification

*Baseline RT model*

The fixed effects structure of the classification model was identical to the naming model. Random effects included intercepts for participants and items, by-participant slopes for all fixed effects, a by-item slope for ordinal position, and correlations among these terms.

The results of the baseline model confirmed the presence of cumulative semantic facilitation, such that participants became faster each time they classified a member of the same semantic category ($\beta=-0.026$, s.e.=0.003, $\chi^2(1)=55.11$, $p<0.001$). A marginal interaction between ordinal position and block suggests that this facilitation effect was somewhat reduced in later blocks ($\beta=0.005$, s.e.=0.003, $\chi^2(1)=2.76$, $p=0.097$). There was also a marginal effect of block, indicating a trend toward slower responses in later blocks ($\beta=0.011$, s.e.=0.007, $\chi^2(1)=2.74$, $p=0.098$).

*Categories vary in speed but not semantic facilitation*

Comparable to the naming data, inclusion of a random intercept for category significantly improved the classification model's fit ($\chi^2(1)=23.44$, $p<0.001$). Figure 4.2 illustrates the wide range of mean RTs observed across categories, spanning more than 200 ms. On the other hand, addition of by-category adjustments to the ordinal position effect did not reliably improve model fit ($\chi^2(1)=0.36$, $p=0.55$). Variation in the semantic facilitation effect was fairly restricted, similar to semantic interference; it ranged from a null effect (*forest animals*) to an approximately 35 ms benefit (*body parts, birds*). Interestingly, the classification results diverge from the naming results in one respect: inclusion of a correlation term between the by-category intercept and ordinal position slope did significantly improve the model ($\chi^2(1)=5.86$, $p<0.05$). Despite the limited variation observed in the semantic facilitation effect, the linear relationship between these variables is apparent in figure 4.2.

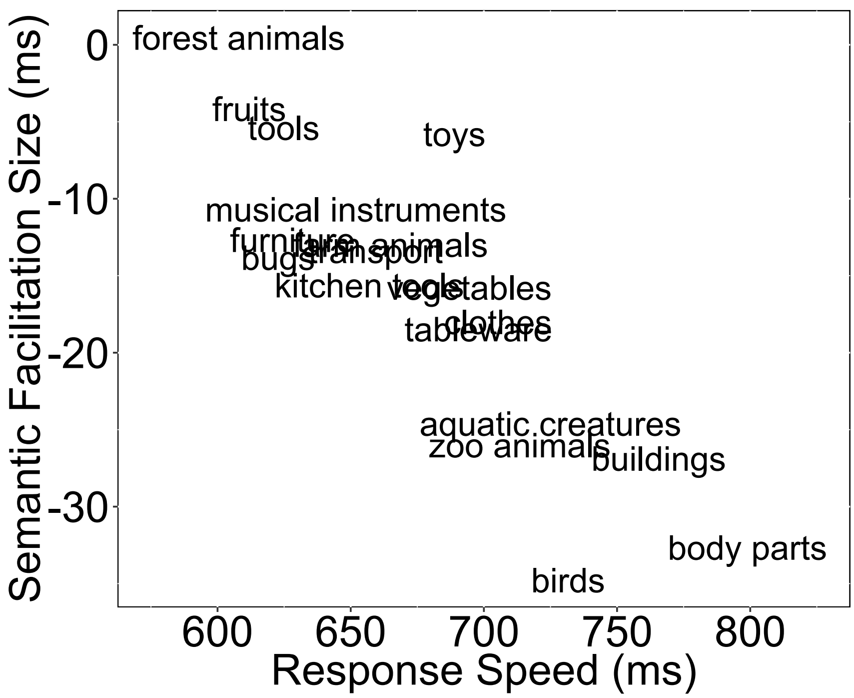



Figure 4.2 Mean RT and semantic facilitation size, estimated for each category using simple linear regressions. Significant variation in speed, but not facilitation size, is observed. These two aspects of classification performance are significantly correlated.

**4.3.3 Comparison of production tasks**

We then examined the correlations between the random by-category effects from each task. First, we tested the relationship between the random category intercepts from continuous naming and classification. If these intercepts improved their respective model fits for the same reason, then we expect a positive correlation across the two tasks, which share identical stimuli. On the other hand, if the intercepts improved the model fits for different reasons, then we expect no correlation to emerge. To test these predictions, we examined the by-category adjustments to the intercept from our mixed effects models. Calculation of Pearson's *r* revealed a near-zero correlation between categories' response speeds during naming and classification; a one-tailed test confirmed that this relationship was not significant ($r(16) = -0.02$, $p = 0.53$). This result indicates that the category intercepts accounted for different sources of variance across the two tasks.

Next, we tested the relationship between the categories' semantic neighbor effect sizes in each task. Although the by-participant slopes for ordinal position had no significant impact on our model fits, we might still expect them to correlate across tasks. If neighbor co-activation in the underlying semantic network triggers both semantic interference and facilitation, then we expect a positive association to emerge between those effects. However, if the origin of these effects is distinct, we expect to observe no relationship. To test these predictions, we could not extract by-category adjustments to the ordinal position slope from our mixed effects models, because the adjustments were estimated as 0 in the naming data. Instead, for each task, we constructed simple linear regressions for each semantic category, predicting RTs by ordinal position within a category. We then extracted the beta coefficients from those regressions and compared them across tasks. Pearson's *r* revealed a mildly positive relationship, but a one-tailed test confirmed that it failed to reach significance ($r(16)=0.27$, $p=0.14$). At most, this trend provides weak support for the idea that similar neighbor co-activation contributes the semantic effects in both tasks.

### 4.3.4 Discussion

Across both continuous picture naming and classification, we observed little to no variation in the size of semantic neighbor effects across categories. This result does not provide support for hierarchical structure in the memory representations of those categories. If there were any supracategory structure linking particular categories, then we would expect to observe some remaining, unexplained variance in the neighbor effect. The presence of a weak correlation between the size of interference and correlation suggests that these phenomena may share some common origin, regardless of other processing differences between the tasks.

While the size of neighbor effects did not vary across categories, we did observe significant variation in their mean naming and classification RTs. These findings suggest that the representations of different semantic categories differ in accessibility. There are many possible explanations for this effect, including differences in categories' frequency of use and age of acquisition (e.g., Morrison, Ellis, & Quinlan, 1992). Interestingly, mean naming and classification RTs were not correlated across tasks, suggesting that the target categories varied in accessibility for different reasons in each. This finding may arise from the tasks' differential reliance on conceptual vs. lexical processing. While conceptualization feeds production processing during naming, it constitutes the primary process underlying classification. Given this processing difference, mean RTs might reflect the lexical accessibility of the semantic categories during naming and the conceptual accessibility of those categories in classification.

Finally, a significant correlation emerged between categories' mean RT and the size of their semantic neighbor effects during classification only. Our ability to detect this relationship is particularly striking, given the lack of variation in semantic facilitation described above. This finding suggests that it takes time for semantic facilitation to influence processing. Specifically, when items/categories take longer to retrieve, i.e., when their semantic representations are less accessible, there is more time available for co-active semantic neighbors to provide a benefit during response selection. The absence of a similar correlation between the category intercept and slope in continuous naming may shed light on the generation of semantic facilitation. We unpack this finding further in the general discussion section.

### 4.4 Analysis 2: An Intuitive Approach to Identifying Supracategory Structure

In the remaining analysis, we followed Alario and del Prado Martin (2010) in attempting to more directly index supracategory effects. Following their work, we took an intuitive approach to clustering semantic categories, subjectively identifying co-categories in the stimuli set (see table 4.1). We then created a binary predictor that coded, for each category, whether any of its co-categories had appeared before. This predictor was valued as -0.5 if a category was the first in its supracategory to appear, 0.5 if one or more of its co-categories had appeared earlier, and 0 if it had no obvious co-categories in the set. That is, within a particular supracategory, all items in the first appearing co-category were marked as -0.5, while all items in subsequent co-categories were marked as 0.5[8]. This predictor was built into the pretest RT models as a fixed effect; it was also allowed to interact with ordinal position within a category.

We predicted that if our artificially constructed semantic categories possessed a hierarchical structure, such that co-categories actually formed part of a larger supracategory, we might see that neighbor effects grow not just across ordinal positions, but also across co-categories. If interference or facilitation from previously encountered co-categories carried over to later co-categories in a straightforward fashion, then we expect to see a main effect of the contrast-coded predictor. Such results would replicate Alario and del Prado Martin (2010), indicating that that semantic representations are stored in a hierarchical fashion, and that additive neighbor effects can result from this structure.

| Supracategory | Co-categories |
|---|---|
| Animals | Aquatic creatures, birds, bugs, farm animals, forest animals, zoo animals |
| Edibles | Fruits, vegetables |
| Implements | Kitchen tools, tableware, tools |
| Leisure | Musical instruments, toys |
| (None) | Body parts, buildings, clothes, furniture, vehicles |

Table 4.1 Subjectively assigned supracategory membership.

[8] A more precise approach might have coded each item's ordinal position within its supracategory (e.g, positions 1-15 among animals appearing at pretest). However, this coding scheme was prohibited by its high correlation with the temporal measure, block.

### 4.4.1 Continuous naming

*Model structure*

The fixed effects structure included ordinal position (1-5), block (1-3), whether a co-category had been previously named (no=-0.5 vs. yes=0.5), and the two-way interactions of ordinal position by block and ordinal position by previous co-category. Random effects included intercepts for participants, items, and categories; by-participant slopes for block and previously named co-categories; a by-item slope for ordinal position; and correlations among these terms.

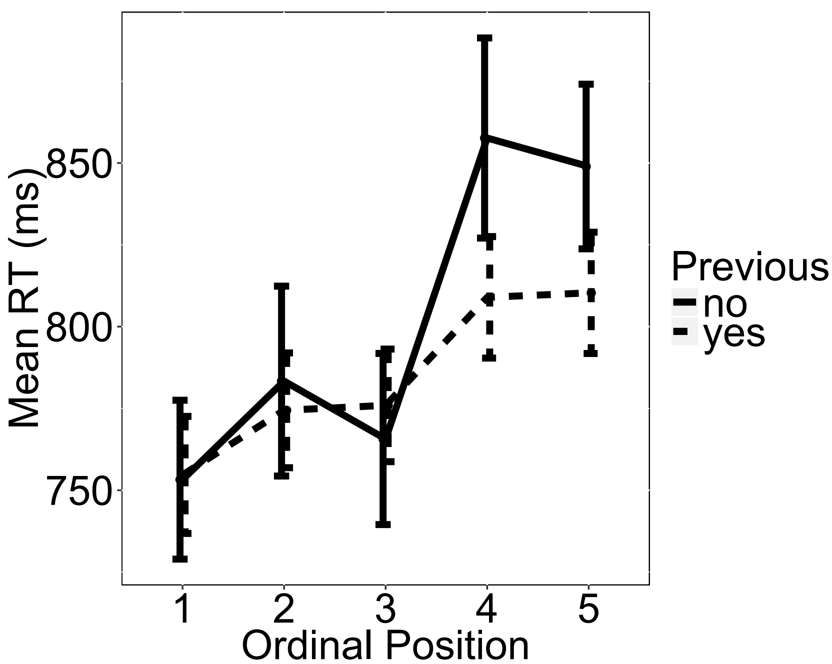



Figure 4.3 Semantic interference (increasing RTs across ordinal positions) is slightly larger when co-categories have not been previously named ("no") compared to when co-categories have been named before ("yes").

*Results*

As above, the results confirmed the presence of semantic interference (ordinal position; $\beta=0.023$, s.e.=0.003, $\chi^2(1)=38.56$, $p<0.001$), which shrank across blocks within the experiment (ordinal position x block; $\beta=-0.010$, s.e.=0.004, $\chi^2(1)=6.09$, $p<0.05$). Crucially, there was no overall effect of previous co-category naming ($\chi^2(1)=0.27$, $p=0.60$). A marginal two-way interaction emerged between ordinal position and whether a co-category had been previously named ($\beta=-0.016$, s.e.=0.009, $\chi^2(1)=3.15$, $p=0.08$). The direction of this effect suggests a trend toward larger semantic interference when a co-category had not been previously named compared to when it has (fig. 4.3)—opposite our prediction for the effect of co-category.

### 4.4.2 Continuous classification

*Model structure*

Fixed effects in the classification model were identical to the naming model. Random effects included intercepts for participants, items, and categories; by-participant slopes for ordinal position, block, whether a co-category was previously named, and the ordinal position by block interaction; and a by-item slope for ordinal position.

*Results*

As above, a negative effect of ordinal position showed semantic facilitation ($\beta=-0.025$, s.e.=0.003, $\chi^2(1)=54.13$, $p<0.05$). There was also an overall effect of block, such that RTs became longer in later blocks ($\beta=0.013$, s.e.=0.006, $\chi^2(1)=4.38$, $p<0.05$). Critically, there was no main effect of whether a co-category had been previously named ($\chi^2(1)=1.99$, $p=0.16$). However, a two-way interaction emerged between ordinal position and previously named co-categories (β=-0.013, s.e.=0.009, $\chi^2(1)=3.16$, $p<0.05$). Once again, this effect patterned in a surprising direction, showing stronger facilitation when co-categories had not been named in earlier blocks compared with when they had. However, figure 4.4 shows that this effect was driven by responses in the first ordinal position. A methodological explanation for this effect, unrelated to the structure of semantic memory representations, is offered below.

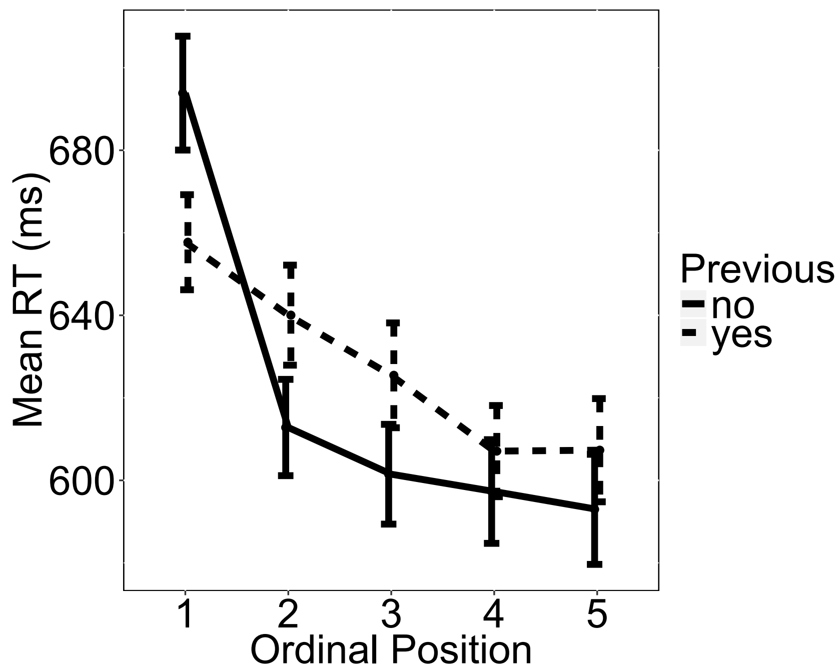



Figure 4.3 Semantic facilitation (decreasing RTs across ordinal positions) is larger when co-categories have not been previously named ("no") compared to when co-categories have been named before ("yes").

### 4.4.3 Discussion

Our second analyses, which intuitively grouped semantic categories into supracategories, provided no strong evidence of a hierarchical structure in semantic memory representations. The critical predictor of whether or not a co-category had previously appeared had no overall effect on RTs. Despite the apparent similarity of some of our constructed semantic categories (see table 4.1), there was no overall dependency among them.

Moreover, previous naming of a co-category did not reliably interact with the ordinal position effect in the continuous naming task. This finding shows that semantic interference was comparable whether or not a co-category had appeared earlier during pretest. Previous co-category naming did interact with ordinal position in continuous classification, in the opposite direction than expected: semantic facilitation was larger when a previous co-category had not been named compared to when it had. However, the interaction was driven by trials at the beginning of each experimental block (i.e., the trials in ordinal position 1). This finding most likely arose from the design of the classification task, where a consistent lag of 2 between category-related items created a repetitive response pattern within each block (see Methods section for more detail). As participants became accustomed to that pattern, they may have anticipated or automatized responses, getting faster over time until the first trial of the next block, which prompted them to slow down and reset their response strategies. Thus, the interaction depicted in figure 4.4 can be attributed to strategic processes, rather than the underlying structure of semantic memory.

## 4.5 General Discussion

Over the course of two linear mixed effects regression analyses, we explored the structure of the semantic memory representations that feed lexical access during spoken word production. Production theories have historically made simplifying assumptions or remained agnostic about this structure; however, the complex empirical landscape of semantic relatedness effects invites further scrutiny of the issue. We therefore followed Alario and del Prado Martin (2010) in searching for evidence of hierarchical semantic representations, using previously collected data from continuous picture naming and classification tasks. We generally failed to replicate their findings.

In the first analysis, we did not observe any significant variation in semantic interference or facilitation across semantic categories. If a subset of the categories belonged to higher level supracategories, then we expected to find unexplained variance in the size of their semantic neighbor effects. However, building that potential variance into the baseline models of participants' naming and classification performance had no effect: inclusion of by-category slopes for the ordinal position effect did not improve model fits.

Furthermore, the second analysis revealed no evidence for the accumulation of semantic neighbor effects across co-categories. Unlike Alario and del Prado Martin (2010), we did not find independent effects of item's ordinal position within a category and a category's ordinal position within a supracategory. Only the former, canonical effect was observed. From these data, we can only infer a single level of semantic representation, rather than a more complex hierarchical structure. In the remaining sections, we consider why our results diverged from Alario and del Prado Martin's and what implications they have for production theories nonetheless.

### 4.5.1 The devil is in the detail of semantic manipulations (and analyses)

The explanation for our divergent results probably lies in one or more of the methodological differences between our study and Alario and del Prado Martin's (2010). First, our studies varied at a very basic level, because they utilized different stimuli. While

Howard et al. (2006) and therefore Alario and del Prado Martin used black and white drawings from Snodgrass and Vanderwart's (1980) database, we used colorized drawings from Rossion and Pourtois' (2004) database. Moreover, there was only partial overlap between the items depicted. These differences could certainly have affected participants' picture naming and classification performance, either because of differences in visual processing or because of the retrieval of different representations from semantic memory.

In addition, the design of our stimulus lists may have also impacted the results. Howard et al. (2006) allowed the lag between members of the same semantic category to range from 2 to 8 trials, and they utilized both targets and filler to achieve this design. In contrast, we elicited semantic neighbor effects with a consistent lag of 2 intervening items between members of the same category, and those intervening items were always targets from other categories (i.e., there were no fillers). This set-up created a blocked structure within our test(s), such that each block rotated through the items of 3 semantic categories. If the delineation between blocks was salient to participants, it might have reduced the chances of semantic neighbor effects spreading throughout a supracategory, given that related co-categories occurred in different blocks. That is, the demarcation of separate blocks might have triggered some sort of reset in the network. For example, if neighbor effects are generated by incremental adjustments to the weighted connections between semantic representations and other representations in the network (Belke, 2013; Oppenheim et al., 2006), then the weights might revert towards their original settings upon detection of a block boundary. This could prevent supracategory relationships spanning those boundaries from influencing processing.

A final difference between our two studies was the approach we took to creating a binary supracategory predictor. On one hand, Alario and del Prado Martin (2010) identified pairs of co-categories in Howard et al.'s (2006) stimuli, which were mostly balanced across participants in terms of order of appearance. They then extracted the subset of data including only those pairs and coded whether a co-category appeared first or second. In this way, their contrast-coded predictor was truly binary. On the other hand, we were unable to select and subset pairs in the same way, because of the pre- vs. posttest design of our original study (Fink & Goldrick, in prep). The counterbalancing of that study ensured that across participants, all items appeared in all ordinal positions at both pre- and posttest, but it did not counterbalance their appearance across blocks within a test. This design means that within pretest, co-categories are not balanced in terms of order of appearance. Moreover, examination of pretest data only meant that we could not easily identify pairs of co-categories: there were no pairs that co-occurred at pretest in all 10 versions of the stimulus list (see Methods). Instead, we identified clusters of co-categories (table 4.1), creating binary contrasts between the first co-category to appear in a stimulus list and all its co-categories appearing later. Given these limitations, our post hoc analyzes may have been biased against finding evidence of supracategory structure.

**4.5.2 Category accessibility and lexical contributions to semantic facilitation**

Nonetheless, the results of analysis 1 have interesting implications for theories of spoken production. For both continuous picture naming and classification, inclusion of a random intercept for category significantly improved model fit. This result, reflecting differences in mean RT across categories, suggests that the categories varied in terms of accessibility in memory. However, because the categories' mean RTs did not correlate across production tasks, we infer that the source of this variation in accessibility differed across naming and classification. This difference most likely relates to the fact that continuous naming depends primarily on lexical selection, while continuous classification depends primarily on conceptualization (a distinction emphasized in the account detailed below). As a result, mean RTs in naming probably reflect variation in the lexical accessibility of semantic categories, whereas mean RTs in classification reflect their semantic accessibility. As noted earlier, variation in either conceptualization of lexicalization might arise from factors like lexical frequency or age of acquisition of the items in a category (e.g., Morrison et al., 1992), which might make varying contributions to different levels of processing.

In addition, the second finding from analysis 1 makes an important contribution to the literature exploring continuous picture naming and classification. Recall that the random category intercept and ordinal position slope were significantly correlated during classification, but not naming. This finding reveals the time sensitivity of semantic facilitation: semantic neighbor co-activation provides more of a benefit (i.e., facilitation is larger) when the target category takes longer to retrieve (i.e., when its mean RT is slower). This correlation—and its absence during naming—provides insight into the generation of semantic neighbor effects.

We follow Oppenheim et al. (2010) in assuming an incremental learning account of such effects. The original account, developed to explain cumulative semantic interference only, argues that every instance of picture naming causes incremental changes to the weighted connections between semantic and lexical representations. For instance, when DOG is named, the connections between the semantic features <furry>, <four-legged>, and <pet> and the word DOG are strengthened, while the connections between those features and related words like CAT are weakened. As a result, subsequent naming of CAT proves more difficult than usual, because the connections from its semantic inputs are less robust. A similar explanation might be constructed to explain cumulative semantic facilitation (see Belke, 2013, for discussion of this account and alternatives). When a dog is classified as “natural,” the connections between <furry>, <four-legged>, and <pet> and the response node NATURAL are strengthened, while the connections between those features and MANMADE are weakened. Subsequent classification of a cat should benefit from these changes, because the connections between some its features and the appropriate response are more robust.

From the finding that semantic facilitation is time-sensitive, we conclude that lexical activation may also contribute to that effect. While not strictly necessary for the classification process, the lexical representations of semantic neighbors may become co-activated during this task. For example, when conceptualization of DOG activates <furry>, <four-legged>, and <pet>, activation may automatically spread not only to the response node NATURAL, but also to the lexical nodes for DOG and its neighbor CAT. These co-active lexical neighbors may then feed activation back to the conceptual level, providing an additional boost to the target classification response, NATURAL. However, the spread of activation to the lexical level and back up to the conceptual level takes time. When an item/category is already quite accessible, the additional lexical boost may not be achieved, because response selection occurs before any benefit is reaped from lexical activation.

Our observation of a significant correlation between categories' mean RT and their semantic facilitation effects supports this account. Furthermore, the absence of a similar effect in continuous naming validates Oppenheim et al.'s (2010) interference account. If interference arises solely from incremental changes to the links between semantic and lexical representations, then the spread of activation over time should have less influence on that effect. Namely, if interference relies exclusively on the feed forward spread of activation along these incrementally adjusted connections, then any potential feedback of that activation over time will not impact its effect size. Thus, the current analyses shed new light on the origins of these popularly studied behavioral phenomena.

## 4.6 Conclusion

The goal of this study was to investigate the structure of semantic representations in memory. Specifically, we sought to identify whether such representations are hierarchical in structure or occupy a one-dimensional space. Over the course of two analyses, which revisited the data from previous continuous picture naming and classification tasks (Fink & Goldrick, in prep), we found no evidence to suggest the existence of hierarchical semantic representations. These results constitute a failure to replicate previous work (Alario & del Prado Martin, 2010). Follow-up studies that explicitly manipulate the hierarchical structure of constructed semantic categories are needed to systematically investigate the organization of the semantic representations supporting spoken word production.

# CHAPTER 5

In the final chapter of this dissertation, we attempt to synthesize the results of the empirical studies presented within. Each subsection begins with a quick summary of our major findings regarding the proposed interactions of the production system with executive functions, articulation, and conceptualization, respectively. We then discuss the high level implications of these findings for theories of spoken word production. Finally, we conclude by presenting recommendations for future work in the field.

## 5.1 Speech may be Special, but it can Benefit from Domain-General Processes

In study 1, four negative transfer experiments revealed that domain-general executive functions can play a role in the resolution of linguistic conflict, particularly during lexical selection. Specifically, the transfer observed in experiment 2 implicates response inhibition in resolving lexical co-activation. This finding aligns with other recent behavioral studies (Crowther & Martin, 2014; Shao, Meyer, & Roelofs, 2013; Shao Roelofs, & Meyer, 2012). On the other hand, the null results in experiment 1 provide no evidence that (proactive) interference resolution modulates lexical co-activation. To our knowledge, few studies have explored the role of this inhibitory executive function during lexical selection. The current result is consistent with one such study, which reported no correlation between interference resolution ability and the growth of semantic interference; instead, the authors found a link between this executive function and repetition priming during production (Crowther & Martin, 2014). Thus, experiments 1 and 2 reveal that  executive function engagement can reduce lexical conflict, but only when that it involves the correct executive function—response inhibition.

Furthermore, the results of experiment 3 and 4 showed that engagement of executive functions can have unexpected consequences on production performance. Across both experiments, we observed that the treatment-induced enhancement of inhibitory control actually caused a performance decrement, because it reduced the effect of semantic facilitation. When co-active semantic neighbors are suppressed by the addition of response inhibition or interference resolution to production processing, they provide less of a benefit to response selection during classification.

On the whole, this suite of experiments provides support for a conflict adaptation account of executive function engagement (e.g., Verguts, Notebaert, Kunde, & Wuehr, 2011). Given the highly practiced nature of spoken production, speakers undoubtedly have production-internal mechanisms for managing conflict between co-active linguistic representations. Nonetheless, this study adds to a growing body of evidence demonstrating that speakers can receive a processing benefit from the strategic recruitment of domain-general executive functions (experiment 2; Hsu & Novick, 2016; Shell, Linck, & Slevc, 2015). In other words, it illustrates the adaptive nature of executive function engagement. This interpretation holds despite the fact that executive function enhancement can also have seemingly maladaptive effects on behavior, when it is carried over to non-target processes (experiments 3 and 4).

### 5.1.2 Implications

The results of study 1 have fairly wide ranging of implications. Most narrowly, they call into question the original framing of the negative transfer paradigm, which developed from a resource depletion perspective (e.g., Baumeister, Bratslavsky, Muraven, & Tice,

1998) of executive functions (Persson, Larsson, & Reuter-Lorenz, 2013; Persson, Welsh, Jonides, & Reuter-Lorenz, 2007). This perspective argues that executive functions rely on a limited pool of cognitive resources that can be drained by continuous usage. Under this account, executive function treatments during a negative transfer paradigm should always produce performance decrements; however, the results of experiments 1 and 2 clearly contradict that prediction. This contradictory evidence—and others like it (Hsu & Novick, 2016; Shell et al., 2015)—strengthen the case of opponents of the resource depletion framework, who question its biological and empirical foundations (e.g., Kurzban, Duckworth, Kable, & Myers, 2013).

With this conclusion, a new question arises: how can we account for Persson et al.'s (2007) original results if not with the resource depletion framework? Their negative transfer study utilized the verb generation task (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997), which presents participants with pictured objects and asks them to generate corresponding verbs. During this task, participants are slower to provide verbs on "many" trials (e.g., BALL could elicit KICK, BOUNCE, or THROW) compared to "few" trials (e.g., SCISSORS is very likely to elicit CUT). Under a conflict adaptation account, we might predict an inhibitory executive function treatment to reduce the many vs. few effect at posttest, if up-regulation of the target executive function biased activation towards a single response on "many" trials. However, Persson et al. found that an interference resolution treatment increased this many vs. few effect at posttest.

One possible explanation for this finding, which is problematic for the conflict adaptation account, is that Persson et al. (2007) targeted the wrong executive function. As in our experiments 1 and 3, their interference resolution treatment task repeatedly exposed participants to proactive interference, such that information from previous trials interfered with response selection on subsequent ones. Such proactive interference is not identical with the underdetermined response conflict present on "many" trials during verb generation. In fact, it is possible that up-regulation of interference resolution may have accidentally strengthened that underdetermined response conflict. Specifically, suppression of potentially interfering information from previous trials may have diverted additional activation towards the co-active response candidates. In other words, application of interference resolution to the verb generation task might have inadvertently intensified the competition among potential verb responses. In this light, Persson et al.'s results become comparable to the unexpected performance decrements observed in experiments 3 and 4, and the conflict adaptation framework holds.

Interestingly, even if the resource depletion account of negative transfer effects is debunked, there may still be value in drawing a link between executive functions and self-regulatory behaviors, as Persson et al. (2007) did. As noted by Hofman, Schmeichel, and Baddeley (2012), these two literatures share some compelling empirical and theoretical common ground. Of particular interest, the link between these fields helps draw our attention to the potential role of motivation in driving executive function engagement.

Indeed, a general implication of study 1 is to encourage consideration of the factors and mechanism(s) regulating executive function engagement. Under a conflict adaptation account, the experience of conflict can trigger up-regulation of executive functions (Verguts et al., 2011). As stated, this theory leaves open the questions of (a) what other factors might also trigger such adaptive behavior (e.g., motivation) and (b) what mechanism(s) monitor for those triggers and implement the resulting adjustments to cognitive control. At least two proposals have recently come to light. Braver (2012) argues for a dual system of executive function regulation, including a strategic system that proactively applies top-down control during difficult tasks and an automatic system that reactively applies it in response to conflict. A second proposal comes from Kurzban et al. (2013), who argue that cost-benefit computations weighing the relative effort vs. payoff of executive function engagement may drive their regulation. This proposal is not mutually exclusive with Braver's (2012); instead, such cost-benefit computations provide one possible implementation of the strategic component of control.

**5.1.3 Next steps**

It is tempting for us to recommend that future work continue implementing the negative transfer paradigm, in order to confirm and clarify the interpretation of our results in isolation and vis-à-vis Persson et al.'s (2007). However, the intermediate timeframe of transfer effects in this task seems non-optimal for exploring executive function engagement, given recent evidence that it can fluctuate on a trial-to-trial basis (Hsu & Novick, 2016, Shell et al., 2015). Because the negative transfer paradigm places the trigger for executive function up-regulation (i.e., the treatment task) and the domain-specific transfer task in separate blocks, we cannot carefully track changes in executive function engagement over time. We note that the block-wise design of the current study provides some evidence of such changes, given the decline and growth of semantic interference and facilitation across posttest blocks. Nonetheless, future trial-level manipulations would allow for finer-grained analyses, enabling us to explore the speed of conflict adaptation in the limit.

If we set aside the question of conflict adaptation and its dynamics, then a second approach emerges as an excellent candidate for deepening our understanding of the role that executive function can play during language processing: training studies. As reviewed by Hussey and Novick (2012), this long-term approach is particularly effective for assessing the causal links between executive control abilities and diverse memory and language phenomena (Hussey, Harbison, Teubner-Rhodes, Mishler, Velnoskey, & Novick, 2016; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014). For example, Novick et al., (2014) demonstrated that twenty hours of training with non-linguistic executive function tasks led to significant improvements in participants' comprehension of garden path sentence structures. While such training tasks obviously require a substantial investment of time and funding, the compelling evidence that results may prove the investment worthwhile.

## 5.2 Articulation is not so Peripheral after all

Across three experiments, study 2 revealed fairly extensive interactions between lexical selection and articulation, which were indexed by response times (RTs) and word durations, respectively. Importantly, observation of these interactions was subject to between- and within-participant variation in the timing of response initiation and the difficulty of response selection. These results help constrain interactive theories of spoken word production.

First, we found a reliable effect of by-participant response speed in all experiments, such that generally slower responders tended to produce longer word durations (over and above any effects of reaction time on a particular trial). This finding suggests that speakers set a steady pace of processing for all stages of production—including articulation. Next, a reliable effect of trial-level RT also emerged in all three experiments. Apparently, a speaker's decision about when to initiate speech has systematic consequences on the articulation of the utterance. Finally, we observed two direct lexical effects on articulation, even with RT accounted for as a covariate. In experiment 1, semantic interference had an overall effect, leading to longer word durations; in experiment 3, a similar effect appeared, but only among participants whose RT interference was sufficiently large.

### 5.2.1 Implications

From our view, this study has three primary implications for theories of spoken production. First, it provides support for the flexible cascade hypothesis (Kello, 2004; Kello MacWhinney, & Plaut, 2000). According to this hypothesis, the architecture of the production system generally supports interactive processing, but the behavioral manifestation of that interactivity depends in part on the degree of temporal overlap between stages of processing. The effects of overall response speed and trial-level RT on word durations support this proposal, showing that the global and local timing of response initiation have systematic consequences for articulatory processing. Furthermore, the finding of semantic interference in durations (experiment 1), even with RT as a co-variate, indicates that temporal overlap between processes allowed lexical co-activation to impact articulation after response initiation. Such direct interactions are subject to between participant variation in sensitivity to lexical disruptions, as shown by the interaction of RT and duration effect sizes in experiment 3.

Study 2 also has a second implication for the field: namely, that the search for minimal planning units (MPUs) may be futile. As reviewed in study 2, one well-known line of inquiry investigates the MPU that must be fully specified before articulation can begin (see Kawamoto, Liu, & Kello, 2015, for a review). Several different MPUs have been proposed, including the phonological word (i.e., an intonational phrase bounded by pauses; Levelt, Roelofs, & Meyer, 1999), the syllable (Meyer, Roelofs, & Levelt, 2003), and the phoneme (Kawamoto et al., 2015). However, counterevidence has also been presented, suggesting that articulation is continuous rather than unit-based (e.g., Pluymaekers, Ernestus, & Baayen, 2005). The current work adds to that counterevidence. Specifically, the finding of opposite trial-level RT effects on word durations in experiments 1 and 2 vs. experiment 3 suggests that the relationship between planning and articulation can flexibly change in response to different task demands.

Lastly, that same piece of evidence—opposite effects of trial-level RT across our experiments—offers a final lesson, albeit methodological rather than theoretical. This result reminds us that seemingly minor methodological choices can have far-reaching consequences on experimental results. In this case, it appears that the choice between automatically timed vs. self-paced trials influenced the relationship between participants' planning and articulatory processes. When the response window was fixed, they strategically divided the time available between planning and articulatory processes, whereas a self-paced window led them to scale the timing of both processes up and down simultaneously. This result corroborates others in the literature, e.g., work suggesting that the scope of syntactic planning can flexibly vary according to task demands (Ferreira & Swets, 2002).

### 5.2.2 Next steps

The obvious recommendation that emerges from study 2 is for future production research to continue the trend of simultaneously analyzing response and articulatory timing measures (durations: Buz & Jaeger, 2015; Damian, 2003; Kello, 2004; Kello et al., 2000; Heller & Goldrick, 2014, 2015). As demonstrated in the current study, this approach can illuminate the flexible coordination of production processes and the interactive effects that result. An excellent example of this approach can be found in the work of Kello (2004), who tested for lexical effects on RTs vs. durations under a range of task conditions. Critically, he observed migration of the target effects from RTs to word durations as task parameters forced more overlap between planning and articulatory processes. Thus, simultaneous analysis of a pure planning measure (RTs) and a potentially blended planning/articulatory measure (durations) ensures that we detect the target effects, no matter where they manifest.

In addition, future work might explore what other mechanisms besides temporal overlap between planning and articulatory processes could contribute to direct lexical effects on articulation. One potential candidate is an internal speech monitor Levelt (1983), which could track the difficulty of planning processes and send a signal for slower articulation when additional planning time is needed. An alternate possibility, suggested by Buz and Jaeger (2015), is that integration of multiple levels of linguistic representation into a single memory structure (e.g., Pierrehumbert, 2002) might give rise to such effects. This account assumes that articulatory variation is stored as part of the target words' representations. We leave it to future work to review and test the predictions of these different accounts.

## 5.3 Considering Semantic Structure is Worth the FUSS

In study 3, we failed to replicate Alario and del Prado Martin (2010), finding no significant effects of supracategory effects in our continuous naming and classification tasks. In other words, we found no corroborating evidence that hierarchical semantic representations underlie these semantic neighbor effects. These divergent results are most likely due to methodological differences between our study and Alario and del Prado Martin's.

Nonetheless, our mixed effects models did detect significant variation in mean RTs across semantic categories, suggesting differences in their accessibility in memory. Furthermore, in the classification task only, the results revealed a significant correlation between categories' mean RTs and the size of their semantic facilitation effects. This relationship, which reflects larger facilitation among slower named categories, indicates that semantic facilitation is somewhat time sensitive. This finding sheds new light on the generation of semantic facilitation, suggesting that even if its origins are conceptual (Belke, 2013), it receives contributions from lexical co-activation and feedback as well. Therefore, the post hoc analyses presented in this study provide new evidence for theories of cumulative semantic neighbor effects.

### 5.3.1 Implications and next steps

Indeed, the primary implication of this study is that it demonstrates the potential utility of fine-grained semantic analyses for illuminating theories of production, particularly those intended to capture semantic relatedness effects. Production researchers continue to vigorously debate the origins of particular semantic effects (see Spalek, Damian, & Boelte, 2013, for a review). As a result, any new evidence that can be brought to bear on them holds promise. This is particularly true of evidence concerning the structure of the conceptual and/or semantic representations underlying neighbor effects like those investigated here.

To that end, future work might follow the example of Vigliocco, Vinson, Lewis, and Garrett (2004). They developed a computational model—the Featural and Unitary Semantic Space (FUSS) model—of the interface between conceptual and lexical processing. Specifically, they proposed that semantic memory includes two types of representations: amodal conceptual features (e.g., <mammal>) and non-decompositional, lexicalized concepts (e.g., DOG). In a sense, they proposed a hierarchical semantic representation that straddles the interface of domain-general and linguistic conceptualization. With this carefully specified model, Vigliocco et al. predicted gradient semantic relatedness effects across a variety of production tasks, which they confirmed during a series of six empirical experiments. Subsequent work has adopted their framework, strengthening the case for this multi-dimensional framework (Belke, 2013). We suggest that future studies have much to gain from similar consideration of the structures and processes involved in conceptualization, the lead-in process to spoken production.

## 5.4 Conclusion

Across three empirical studies, this dissertation has promoted a very broad view of the production system. We have considered the contributions of seemingly distinct or peripheral cognitive systems—executive functions, articulation, and conceptualization—to the articulatory/acoustic outcomes of spoken word production. Adopting this type of perspective is certainly a challenge, because it requires a working knowledge of diverse literatures. While acknowledging this challenge, we believe that pushing towards more cognitively integrated production theories is worth the effort. As we hope to have demonstrated, this approach holds the potential to unravel diverse empirical mysteries and inconsistencies within the field of spoken word production

# REFERENCES


Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, 53, 741-764.

Alario, F. X., & del Prado Martín, F. M. (2010). On the origin of the "cumulative semantic inhibition" effect. *Memory & Cognition*, 38, 57-66.

Alario, F. X., Ziegler, J. C., Massol, S., & de Cara, B. (2012). Probing the link between cognitive control and lexical selection in monolingual speakers. *Topics in Cognitive Psychology,* 112, 545-559.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47, 31-56.

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. 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 and Language*, 59, 390-412.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In The psychology of learning and motivation (Vol. 8, pp. 47-89). New York: Academic Press.

Badre, D., & Wagner, A. D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cerebral Cortex,* 15, 2003-2012.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24, 527-554.

Baker, R., & Bradlow, A.. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, 52, 319-413.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28,* 14-36.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bates, D., Maechler, M., Bolker, B., & S. Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48. doi:10.18637/jss.v067.i01.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252.

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.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, 58, 667-692.

Belke, E. & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *The Quarterly Journal of Experimental Psychology*, 66, 2135-2160.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60, 92-111.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, 16, 106-113.

Breining, B., Nozari, N., & Rapp, B. (2015). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, 1-7.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one*, 5, e10773.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977-990.

Buz, E., & Jaeger, T. F. (2015). The (in) dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 1-21.

Carr, T. H., McCauley, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: on semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 757.

Chrysikou, E. G., Weber, M. J., & Thompson-Schill, S. L. (2014). A matched filter hypothesis for cognitive control. *Neuropsychologia*, 62, 341-355.

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, 12, 125-131.

Crowther, J. E., & Martin, R. C. (2014). Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. *Frontiers in Human Neuroscience*, 8, 1-20.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 416.

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

Damian, M. F., Vigliocco, G., & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81, B77-B86.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

Desender, K., & Van den Bussche, E. (2012). Is consciousness necessary for conflict adaptation? A state of the art. *Frontiers in Human Neuroscience*, 6.

de Zubicaray, G., McMahon, K., & Howard, D. (2015). Perfusion fMRI evidence for priming of shared feature-to-lexical connections during cumulative semantic interference in spoken word production. *Language, Cognition and Neuroscience*, 30, 261-272.

de Zubicaray, G. I., Wilson, S. J., McMahon, K. L., & Muthiah, S. (2001). The semantic interference effect in the picture-word paradigm: An event-related fMRI study employing overt responses. *Human Brain Mapping*, 14, 218-227.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27-41.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57–84. doi: 10.1006/jmla.2001.2797.

Fink, A., & Goldrick, M. (2015). The Influence of Word Retrieval and Planning on Phonetic Variation: Implications for Exemplar Models. *Linguistics Vanguard*, 1, 215-225.

Fink, A., & Goldrick, M. (in prep). Transfer from Executive Function Tasks to Picture Naming and Classification: Conflict Adaptation across Domains.

Friedman, M. P. & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133, 101-135.

Finkbeiner, M., & Caramazza, A. (2006). Lexical selection is not a competitive process: A reply to La Heij et al.(2006). *Cortex*, 42, 1032-1036.

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, 83, 97-117.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26, 489-504.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84, 474-496.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66, 789-806.

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

Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, 121, 58-72.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21, 649-683.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America*, 134, EL172-EL177.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103, 386-412.

Heller, J. R., & Goldrick, M. (2014). Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin & Review*, 21, 1576-1582.

Heller, J. R., & Goldrick, M., (2015). Corrigendum to "Grammatical constraints on phonological encoding in speech production". *Psychonomic Bulletin & Review*.

Hofman, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive function and self-regulation. *Trends in Cognitive Science*, 16, 174-180.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic interference in picture naming: Experimental and computational studies. *Cognition*, 100, 464-482.

Hsu, N. & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science*, doi:10.1177/0956797615625223.

Hussey, E., Harbison, J. I., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (in press). Memory and language improvements following cognitive control training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Hussey, E. K., & Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Frontiers in Psychology*, 3, 1-14.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, 67, 311-325.

Kahn, J. M., & Arnold, J. E. (2013). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *Language and Cognitive Processes,* DOI: 10.1080/01690965.2013.848989.

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 43-57.

Kawamoto, A. H., Liu, Q., Lee, R. J., & Grebe, P. R. (2014). The segment as the minimal planning unit in speech production: Evidence based on absolute response latencies. *The Quarterly Journal of Experimental Psychology*, *67*, 2340-2359.

Kawamoto, A. H., Liu, Q., & Kello, C. T. (2015). The segment as the minimal planning unit in speech production and reading aloud: evidence and implications. *Frontiers in Psychology*, 6.

Kello, C.T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 942-955.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General*, 129, 340–360.

Kirov, C., & Wilson, C. (2013). Bayesian speech production: Evidence from latency and hyperarticulation. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 788-793).

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, 36, 661-679.

Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 829.

Levelt, W.  J. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user's guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), Inhibitory Processes in Attention, Memory, and Language (pp. 189–239). San Diego, CA: Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91, 295.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, 109, 163.

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.

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117, 243-260.

McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, 24, 44-66.

Meyer, A. S., Roelofs, A., & Levelt, W. J. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, 48, 131-147.

Miozzo, M., & Caramazza, A. (2003). When more is less: a counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, 132, 228.

Miyake, A., Friedman, M. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and the contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, 10, 465-501.

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, 20, 705-714.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, 47, 1048.

Navarrete, E., Del Prato, P., & Mahon, B. Z. (2012). Factors determining semantic facilitation and interference in the cyclic naming paradigm. *Frontiers in Psychology*, 3, 1-15.

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

Nelson, J. K., Reuter-Lorenz, P. A., Persson, J., Sylvester, C. Y. C., & Jonides, J. (2009). Mapping interference resolution across task domains: A shared control process in left inferior frontal gyrus. *Brain Research*, 1256, 92-100.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29, 186-217.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4, 906-924.

Oppenheim, G. M. (in prep). Persistent semantic interference in picture naming: Test of an incremental learning account.

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.

Persson, J., Welsh, K. M, Jonides, J., Reuter-Lorenz, P. A. (2007). Cognitive fatigue of executive processes: Interaction between interference resolution tasks. *Neuropsychologia*, 45, 1571-1579.

Persson, J., Larsson, A., & Reuter-Lorenz, P. A. (2013). Imaging fatigue of interference control reveals the neural basis of executive resource depletion. *Journal of Cognitive Neuroscience*, 25, 338-351.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology*, 7, 101-139.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118, 2561-2569.

Riès, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T., & Dronkers, N. F. (2015). Specifying the role of the left prefrontal cortex in word selection. *Brain and Language*, 149, 135-147.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, 247-253.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33, 217-236.

Runnqvist, E., Strijkers, K., Alario, F. X., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, 66, 850-869.

Scarborough, R. (2004). Coarticulation and the structure of the lexicon. Doctoral dissertation, University of California Los Angeles.

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In by C. Fougeron, B. Kuhnert, M. D'Imperio and N. Vallee (Eds.) Laboratory Phonology 10 (pp. 557-586) Berlin: Mouton de Gruyter.

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, 75, 27-44.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, 54, 199-227.

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirschorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the functions of Broca's area. *PNAS*, 106, 322-327.

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.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133, 140-155.

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory and Cognition*, 41, 1200-1211.

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Research*, 1586, 130-142.

Shao, Z., Roelofs, A., & Meyer, A.S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The Quarterly Journal of Experimental Psychology,* 65, 1927-1944.

Shell, A. R., Linck, J. A., & Slevc, L. R. (2015). Examining the role of inhibitory control in bilingual language switching. In D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, & P.P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2146-2151). Austin, TX: Cognitive Science Society.

Simon, H. (1990). *Reason in human affairs*. Stanford University Press.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174.

Spalek, K., Damian, M. F., & Bölte, J. (2012). Is lexical selection in spoken word production competitive? Introduction to the Special Issue on lexical competition in language production. *Language and Cognitive Processes*, 28, 597-614.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94, 14792-14797.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language,* 28**,** 127–154.

Verguts, T., Notebaert, W., Kunde, W., & Wühr, P. (2011). Post-conflict slowing: cognitive adaptation after conflict processing. *Psychonomic Bulletin & Review*, 18, 76-82.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, 85, B61-B69.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422-488.

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language*, 33, 332-356.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. J. Local, R. Ogden, & R. Temple (Eds.), Laboratory Phonology VI (pp. 75–87). Cambridge: Cambridge University Press.

# APPENDIX

Semantic categories and items in studies 1-3

| Category | Items |
|---|---|
| Aquatic creatures | Fish, lobster, penguin, seal, seahorse |
| Birds | Duck, eagle, owl, peacock, swan |
| Body parts | Ear, eye, finger, hand, nose |
| Bugs | Ant, beetle, butterfuly, caterpillar, spider |
| Buildings | Barn, church, house, well, windmill |
| Clothes | Coat, dress, shirt, skirt, sock |
| Farm animals | Chicken, donkey, horse, pig, sheep |
| Forest animals | Deer, fox, raccoon, skunk, squirrel |
| Fruits | Apple, banana, lemon, orange, pear |
| Furniture | Bed, chair, desk, stool, table |
| Kitchen tools | Kettle, pan, pitcher, pot, rolling pin |
| Musical instruments | Drum, guitar, piano, trumpet, violin |
| Tableware | Cup, fork, glass, knife, spoon |
| Tools | Axe, hammer, saw, screwdriver, wrench |
| Toys | Ball, doll, kite, sled, top |
| Vegetables | Carrot, celery, lettuce, onion, potato |
| Vehicles | Bus, car, helicopter, plane, train |
| Zoo animals | Elephant, gorilla, monkey, tiger, zebra |