The Probability of Lexical Selection

By

Mark Koranda

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Psychology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 12/15/2020

The dissertation is approved by the following members of the Final Oral Committee:
- Maryellen MacDonald, Professor, Psychology
- Martha Alibali, Professor, Psychology
- Joseph Austerweil, Assistant Professor, Psychology
- Gary Lupyan, Associate Professor, Psychology
- Margarita Kaushanskaya, Professor, Communication Sciences and Disorders

To Mom and Dad,

for whom language has been magnificent,

and yet so often failed.

Contents

Acknowledgements

Language is central to most of the great achievements of humanity. The creation of literature, spread of science, and study of history require words. Language is central to humanity's major shortcomings, too. Literacy stratifies communities, dogmas justify prejudice, and words perpetuate fear and lies. I both love and distrust language. Language's natural power—to simplify and manage humans' complex social behavior—leads to oversights in social equity. Therefore, I can think of no clearer form of expressing my gratitude, than in terms of the essential community actions that permit solo ventures into the unknown (e.g., a thesis).

To Troop 254 and my fellow Marines, who taught me to survive and love the unknown.

To the faculty at the University of St. Thomas, who provided a home for my understanding of human mind and behavior, and whose genuine support framed my early academic curiosity, especially Greg Robinson-Riegler and J. Roxanne Prichard.

To the University of Wisconsin-Madison's Psychology Department, especially my advisor Maryellen C. MacDonald, who provided regular meals and taught me to cook. Thank you to my committee members for snacks and manners. To Erik Raimy and Morton Ann Gernsbacher, for late-night food runs. To the Language and Cognitive Neuroscience lab, a world-class kitchen of world-class chefs I've had the pleasure of training alongside: Yaling Hsaio, Christopher Cox, Lang Chen, Elise Hopman, Matthew Cooper-Borkenhagen, Misty Kabasa, Steven Schwering, Arella Gussow and Matthew Borman. To the Research Assistants for their willingness to learn, experiment and count alongside me: Amanda Timmerman, Ann Meinholz, Arya Liao, Chris Borges, Christine White, Erika Dorow, Jocelyn Marks, Joshua Geiger, McKenna Wisch, Mengyuan Wang, Morgan Wheeler, Najhee Purdy, Sichao Yang, Sydney Polzin, Tali Despins, Zak Kofiro, and Allie Durdall.

To my siblings Martha and Matt, who put the roof on and kept it there. To Mary and Tim, for fresh air, perspective, and sanctuary at the base camp.

To Bill Kenney, who gave me a ride, Inocencio Ortiz-Toro, who gave me a pillow, and Nicholas Michalak, who gave me shoes. To Monica Munoz-Torrez, who gave me a map, Pierce Edmiston who gave me the code, and Natalie Eilbert who gave me a jacket of the world.

Author contributions

Chapter Two was co-authored with Martin Zettersten and Maryellen C. MacDonald and is currently in submission for publication. MZ and MK shared responsibility for development of experimental design and MZ was responsible for the first draft of analysis. MK wrote the first draft except for the analysis. MZ and MCM contributed to revisions of the entire manuscript.

Chapter Three was co-authored with MCM. MK wrote the first draft, and MCM contributed substantial revisions, especially to the introduction and discussion.

This research was supported by NIH T32DC005359 in addition to NSF-GRFP DGE-1256259 awarded to Martin Zettersten and the Wisconsin Alumni Research Fund awarded to Maryellen C. MacDonald.

Abstract

Language production, among other things, is the encoding of meaning into words. It is therefore reasonable that theories have assumed that words’ meanings alone determine which word is selected when expressing a message. Such an emphasis on meaning-based word selection accounts for perfect and erroneous language production but leaves little room for perceived in-between experiences, such as saying something just “good enough” or “almost wrong.” This dissertation establishes two new experimental paradigms to investigate the impact of other effects, like word frequency and form, on the probability of lexical selection.

The first paradigm tests the effect of relatively more prior experience with words on later word choices. In these experiments, participants learned novel words with variable practice across the vocabulary. Then, in a task that rewarded precise word choices (i.e., meaning accuracy), participants preferred to produce high-frequency words, even when other words they knew were more meaningful, suggesting a direct trade-off of word frequency for meaning.

The remaining experiments test the impact of phonological and lexical form on picture naming with natural language. On some trials, participants read aloud printed words that shared either phonological onset or the entire name of an upcoming picture’s possible names. In both cases, the probabilities changed for what participants called the picture. Naming likelihood was reduced by phonological similarity and increased by lexical priming.

All three studies provide strong, converging support that properties of word form influence lexical selection. This work challenges the assumption that word use strictly reflects meaning. People say what is circumstantially convenient, as well as what they mean. Probability in lexical selection has consequences for many aspects of language, including what we might assume about our own and others’ words while communicating.

1. Introduction

Consider a situation in which there are two different mugs on a table, and a person picks one up. Why that mug, and not the other? Perhaps it was the one they wanted, or the one more convenient to grab. Hypotheses about factors that affect mug selection could be investigated by setting up circumstances that pit the factors against each other, such as mug preference or reach distance, and evaluate the distribution of mugs selected. Often, action decisions such as mug preference and distance are a blend of utility and efficiency (Wolpert & Landy, 2012).

Contrast the action of selecting a mug with saying the word *dog*. Why the word *dog*? Unlike many action decisions, investigating this question poses a challenge, because it is often unknown what alternative options were possible to convey the message, and just how practical they were. Whereas mugs share objective affordances in an objective physical world, the properties of a word can vary from individual to individual. Mugs can be tested independent of their practical use, for example weighing them before testing whether weight determines human choice for a mug, but a similar separation is not practical for investigating decisions with words. We rely on the usage of words to infer their inert properties. This might account for why theories about decision-making for words (lexical selection), begin with a strong assumption: people say what they mean.

When it comes to word choice, or *lexical selection*, a tremendous amount of work and debate concerns the mechanism of how the correct word is retrieved from memory among the many thousands of words a producer knows (Levelt, 1999). Theories of production can account for varying difficulty in the retrieval of words, including slower retrieval, occasional fumbles in pronunciation, or outright failure to retrieve the word at all. There is a theoretical blind spot when it comes to explaining the times another, less-than-right word is produced. In fact, there are no words for it at all.

Despite a colloquial sense of the word accuracy, these in-between phenomena cannot be described as inaccurate. In language production, *accuracy* refers to the proportion of times a person says a single, specified word in response to a visual stimulus, usually a picture that unambiguously depicts a familiar object. For example, saying *dog* in response to a picture depicting a dog would be accurate. Saying anything but the word *dog*, to include accidentally saying *dock* or *wolf*, lowers the response accuracy for that stimulus. Saying *dock* is an example of a *disfluency*, a failure to articulate the accurate word. Saying *wolf* is labeled an *alternative* name, in that an alternative name is the accurate response (for whatever unexamined reason), for that individual. For the purposes of understanding lexical selection, this individual's behavior is noise.

The rest of this dissertation is an attempt to investigate and describe an alternative account for alternative words being selected. Rather than *wolf* being that producer's accurate word, it could be that *wolf* was only circumstantially the optimal word for production. Before investigating those circumstances, I provide a brief overview of the process of choosing words, elaborating on the distinctions between current views of production and this alternative.

1.1 Lexical Selection

As a matter of experimental control, this investigation is largely conducted in the narrow scope of producing single words or short phrases (vs sentences or discourses). Specifically, studies with short utterances provide the advantage of controlling for various spillover and planning effects from, e.g., the grammatical encoding of multiple-words (Allum & Wheeldon, 2007). In this scope, a message is typically prompted by the presentation of a simple visual stimulus (e.g., black and white line drawing), and a participant is tasked with naming it, saying a word or short phrase (Johnson et al., 1996).

The process of lexical selection comprises several subprocesses (e.g., Dell & Reich, 1981; Levelt, 1992), the first of which is the activation of semantic properties relevant to the communicative act, given the stimulus. The activated properties which precede any linguistic constraints (c.f. Slobin, 1987), are the *message* (Konopka & Brown-Schmidt, 2014).

A key question pertains which semantic aspects of a stimulus guide selection of a word (Arnold, 2008). Relevant semantic properties are typically operationalized in terms of (semantic differences between) viable alternative words. For example, producers could characterize an event in terms that are aligned with a model of the listener's viewpoint (Arnold, 2008), or their own viewpoint (Engelhardt et al., 2006; Ferreira & Dell, 2000a). For illustration, consider a comprehender facing a producer. From a producer-centered viewpoint, the phrase "I will step to the…" might be completed with "my right", whereas accommodating the listener's perspective might result in a completion of "your left". Both listener and producer-centric criteria for selection concern how best to characterize message-alignment, the relative semantic overlap between a word and the intended message. Regardless of whether message-alignment is listener- or producer-centric, the shared assumption is that semantic overlap determines selection.

This assumption is carried forward with investigations of the mechanism of lexical selection. Here, an unambiguous message is a starting point, defined empirically as the "semantically and syntactically appropriate" selection of a word (Levelt, 1999, p. 223). This manifests in experimental work as stimuli for which one name is the overwhelmingly common response among participants (dominant name). These are called high name agreement stimuli (Johnson et al., 1996; Perret & Bonin, 2018).

The focus of research in lexical selection is to account for the time course of activation of an intended word, alternative words, and the extent to which they interact in the process of selection (Nozari & Hepner, 2018). Most accounts of lexical selection assume spreading activation from message to distributed semantic features (Collins & Loftus, 1975), which then activate the words that share those semantic features (Dell, 1986). Among other things, spreading activation posits that the degree of activation for a word is proportionate to the degree of overlap between the semantics of the word and the semantics of the intended message, which I call message-alignment. The most message-aligned word will tend to be the most activated, and selected most often, given a distribution of production trials. That is to say, if the person saw a *wolf,* saying *dog* would be possible, but likely considered a production error.

After selection of a word are aspects of phonological encoding and articulating. Evidence of separate stages comes from difficulty in speaking (Levelt et al., 1999), such as Tip of the tongue states (Abrams et al., 2003; Reilly & Blumstein, 2014a), during which a producer is often able to describe semantic properties of a selected word, but fails to retrieve its phonological form; naming latencies that are best accounted for by phonological properties such as the relative number of phonologically similar words (e.g., Yates et al., 2004); and production errors induced by phonological priming (e.g., Ferreira & Griffin, 2003). The separation of phonological representations of words and processing of their use, from semantic representations and processes, are a point of significant debate in language production.

Interactive accounts of language production more readily allow for effects of alternative word selection, such as *dog* instead of *wolf* (see Baese-Berk & Goldrick, 2009; Dell, 1986). These models posit that activation can spread across a word's semantics, its form. To take one implication, under circumstances where form properties between alternative words are notably similar, an alternative word might have a slightly higher likelihood of causing a dysfluency (Dell & Reich, 1981). However, in practice, such effects are assumed to be small (Dell & O'Seaghdha, 1991; Goldrick, 2006). Evidence for interactive accounts, and the reasons for their practical limitations on lexical selection of alternatives, are discussed in Chapters Three and Four.

1.2 Lexical Variability

Relevant to this dissertation, production models comprise a key assumption: that the decision about words is made because of activation properties in the semantic mapping. In this way, investigations of referring expressions, and the mechanism of lexical selection alike assume a message-aligned distribution of utterances. This assumption has an important consequence for how we understand the function of communication. It posits a default attribution of any variation in utterances as stemming from variation in *message* (Johnson et al., 1996), or properties of the semantic stages of production.

Such an assumption does not predict, let alone easily account for, the case where an alternative word arises given a fixed message (See Figure 1). While typically dispreferred for reasons of experimental control (Johnson et al., 1996), it is not uncommon for lexical selection data to yield a distribution of names (vs complete name agreement) given a fixed message. When this happens, the interpretation is either that message-alignment mappings, overlap between semantic properties of a message and semantic properties of a word, vary across individuals, due to idiosyncratic differences in producers' experiences with words, or an error occurred in production (Johnson et al., 1996).

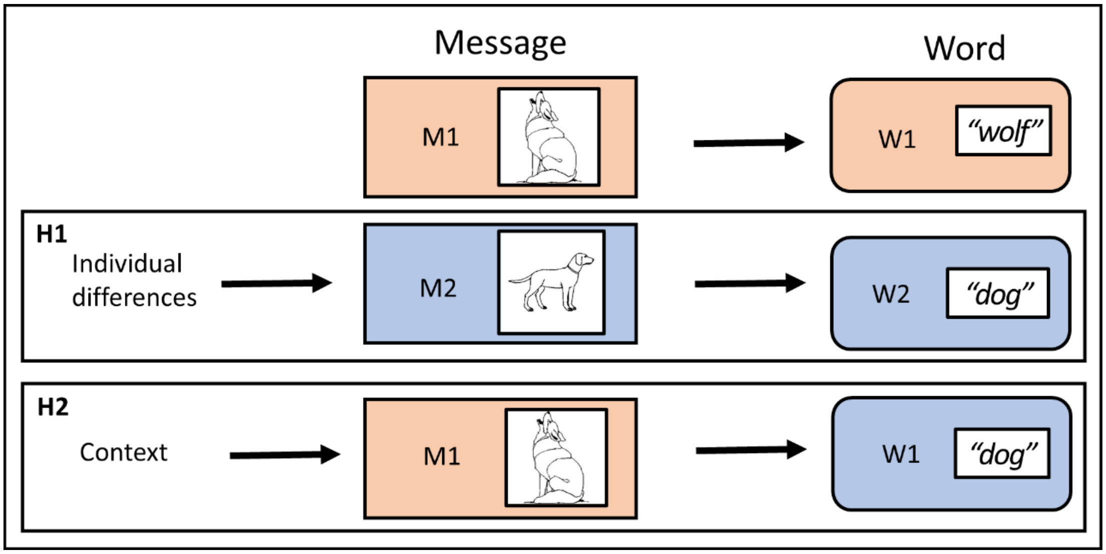

Figure 1. Contrasting hypotheses about messages underlying variation in lexical selection. Under current models of lexical selection (H1), differences in word form (W1, W2) are activated and selected as the result of intending to communicate a correspondingly different message (M1, M2). When a person uses a different word, such as *dog* in response to WOLF (picture) the attribution is to individual differences in meaning-to-word mapping, but not a non-meaning factor, such as a child considering *dog* the correct word for WOLF. An alternative hypothesis (H2) is that, beyond individual differences, context-specific variability can occur such that a word less aligned with the message is selected.

In this dissertation I investigate an alternative account that predicts variability within producers. In this interactive account, activation and processing of word forms can affect what is said. In addition to experience driving baseline differences in individuals' message alignment, experience has a dynamic effect in the immediate processing of words (Figure 1).

In practice, current models of lexical selection, interaction ones included, are set up to account for what Levelt (1999) referred to as the *appropriate* word. Nuances aside (see Chapter Three and Four), none predict extensive variability of production within an individual, given a fixed message[1]. For example, why is it possible for the same presentation talk to be given well one day, and then poorly, another? If both talks occur at the same time of day, in the same room, by the same speaker, with a random sample of audience members, previous models would predict relatively little, if any difference in production. They are ill-equipped to describe variability (not errors) sufficient to change the quality of communication. What is critically missing in these accounts is an investigation of non-message factors systematically influencing lexical selection.

1.3 Approach

Across eight experiments I investigate the potential influence of non-message factors central to language production, namely, properties of language experience, word forms and phonology. I show that within the scope of naming a simple visual stimulus there is clear evidence for changed naming distributions, or *lexical variability*. The approach is to hold the message stimulus constant and vary context-specific properties that are unrelated to message-alignment (*non-message factors)*. These studies take the first critical step to understanding variability in larger production contexts, such as a presentation, by investigating variability in the smallest units of language production: our words.

In Chapter Two, starting with an invented language, I assigned a quantifiable meaning to a set of words for participants to learn and use. Participants learned a new language of eight inter-related meanings. Then I manipulated language experience to investigate whether

---

[1] A caveat about sentence-production data is worth mentioning. On the surface, such variability is well-accounted for in sentence production under the umbrella term *accessibility* (K. Bock, 1987); producers will opt for a sentence structure (Gennari et al., 2012) or word choice (Jaeger et al., 2012b) over an alternative given appropriate cognitive advantages or biases. However, as mentioned, it is impossible to fully tease apart effects of sentence planning from accessibility. More to the point, for the same reasons, the consensus about sentence-level accessibility is in line with the general assumption that we speak *appropriately*.

experience with specific word forms influences the (retrieval, and subsequently,) use of words. Usage of more frequent words can be directly evaluated in terms of a trade-off between meaning precision and efficiency.

This paradigm allowed for the control of individual participants' message-alignment mappings, world- and word- experiences. On this premise, I then indirectly manipulated the strength of connection for each participant, across their eight-word vocabulary, so that half of the words were much easier to retrieve than the other half. In a follow-up experiment, I further control for the potential confound of world exposure with word exposure.

Perhaps the hallmark non-message factor is the phonemes in a words' form. Chapter Three investigates lexical variability in the latter production stage, phonological encoding. The serial model (and the interactive model in practice) predicts that factors of phonological encoding – assembling the phonemes associated with the word-meaning—should implicate speed and accuracy, but not selection. To test lexical variability, I used a priming method, where primed words shared part of the phonology of a target picture name. For example *code* / *coat.* The prime words were semantically unrelated to the target word and occurred on independent production trials, immediately preceding target naming. An effect of lexical variability from phonological patterns across words, would offer an idiosyncratic yet systematic source that interferes with message-aligned selection of a word.

Chapter Four investigates the role of the full lexical form in selection. If variable activation to part of a word's phonological form can affect lexical selection, it raises the question of whether activation of complete word forms similarly affects selection, much like word frequency effects in Chapter One.

To test the impact of word form on selection, I used another form of priming. This time, participants read aloud the second most common (secondary) name for target pictures. Between fifty and a hundred trials later, they saw and named the picture for the first time. If spreading activation and selection are strictly concerned with the overlap of message to the meaning of words being retrieved, then manipulations to the activation of word forms would have no effect on lexical selection. Conversely, an effect of producing words in the recent past, on later, new messages, would be another compelling case of lexical variability.

In the conclusion I discuss the implications. This dissertation characterizes an interactive model of language production more dynamic than previously posited. The various innovations of design license a novel, modest generalization about how lexical selection occurs. Namely, that the words a person says can vary widely, and partly depend on idiosyncrasies of words they have recently said. This generates predictions for what we assume about what we mean by our words and by extension the meaning of words said.

2 Good-enough production

Options for a behavior require a decision process, such as which hand to use to reach for a coffee cup. Recent theories of motor control have hypothesized probabilistic decision making processes that maximize utility of actions in face of uncertainty (e.g., Wolpert & Landy, 2012). Language production is a form of action that offers abundant alternative behaviors to convey a message, namely alternative words and phrases, but measuring the costs and benefits between alternatives has been prohibitive.

Interestingly, theories of word choice, also known as lexical selection, have not favored the probabilistic approach seen in motor control research but have instead hypothesized a more deterministic and encapsulated set of processes (Levelt, 1999). In theories of lexical selection, early grammatical encoding processes settle on words and word order to fit the message, and later phonological encoding processes develop the phonological code for overt production (Levelt et al., 1999). On this view, a speaker’s word choices, such as *cat* vs. *kitten,* are guided solely by which words best align with the intended message. Factors beyond message alignment that might affect the ease of producing words (often called “accessibility”), such as a word’s frequency or length, may affect processing difficulty, but not lexical selection. This deterministic approach thus places theories of language production at odds with the probabilistic processes hypothesized for motor control. In this paper, we investigate whether lexical selection in fact has this deterministic message-driven character or whether there is evidence of more probabilistic decision making. The results could have important implications for the ways in which the uniquely human behavior of language production does and does not differ from other forms of action that are seen across species (MacDonald, 2013).

Several studies cast doubt on whether lexical selection is truly controlled by a single deterministic factor. Ferreira and Griffin (2003) examined errors in picture naming and found that compared to a control condition, speakers misnamed pictures more often when the phonological form of an incorrect competitor word had been primed, such as calling a picture of a priest a *nun* when the homophone *none* had been primed. Ferreira and Griffin termed this result "good enough production," meaning that producers weigh not only message alignment but also the accessibility of phonological forms in lexical selection, so that when the form nun was especially accessible, this word was incorrectly chosen to describe a priest. Similarly, other studies have increased the difficulty of producing certain words via phonological interference manipulations and found that the interference manipulations increase producers' avoidance of difficult words when describing pictures (Jaeger, Furth & Hilliard, 2012; Koranda & MacDonald, 2018). Together, these studies are consistent with a probabilistic model of lexical selection, in which message alignment and accessibility together constrain producers' word choices.

These few studies with fairly restricted materials have had relatively little impact on deterministic accounts of lexical selection, for several reasons. One is methodological: it is very difficult to test empirically the degree to which message alignment and accessibility are independent. For example, Bock (1982) noted that message alignment is part of what makes a word easy to produce, on the view that the semantics of an intended message is a source of activation of candidate words during lexical selection. That is, in everyday language, a highly active word might have been chosen both because it fit the message and was frequent in past experience. Distinguishing these interpretations has proved difficult.

A second methodological challenge lies in controlling the variability across speakers' usage of words. Individual variation in producers' dialects and other experiences may lead to variation in word use, which further amplifies the difficulty of precisely measuring message alignment. Testing the hypothesis that speakers weigh both message alignment and production difficulty requires that we have independent evidence of both of these factors. Relatedly, it has proved difficult to quantify the degree to which some non-message factor might affect word choices: Is lexical selection effectively a deterministic process except in unusual cases such as homophone production, or is probabilistic integration of several factors an intrinsic part of lexical selection, as is hypothesized in other motor behaviors?

For these reasons, we designed a small artificial language that allowed us to precisely manipulate the strength of both message alignment and accessibility, in order to quantify the interplay between these factors. We assigned the novel words equidistantly along a single, continuous semantic space--directions on a compass. We varied the frequency of these words during training, thereby affecting participants' practice with different words and thus their accessibility--the ease with which these words could be retrieved and produced. After the training phase, we assessed lexical selection behavior in a treasure hunting communication game, where participants responded to compass points on screen and produced directions to guide elves to treasure locations.

The task of using a small number of compass terms to describe many different angles resembles a common feature of everyday language. For example, the cities of Detroit and Pittsburgh are located at different precise compass directions from Chicago, but we can describe both of them as "east" of Chicago. Therefore, if production choices are deterministic and driven only by message alignment, then participants should produce directions that best match the message indicated by a compass arrow prompt. However, if lexical selection proceeds via a more probabilistic integration of message alignment and accessibility, then producers' directions should deviate from the most accurate message in some circumstances. Specifically, producers should sometimes produce high-frequency words even when the low-frequency alternative is more aligned with the message. Because our artificial language exactly specified the messages in the compass points in both learning and test, and because we also controlled the relative frequency - and consequently the accessibility - of different words in the language, we can quantify the degree to which message alignment and frequency affects lexical selection in a way that has not been possible to date.

2.1 Experiments 1 and 2: Word frequency modulates lexical selection

We developed a small artificial language containing four novel high-frequency and four low-frequency words, each of which referred to a precise direction on a compass. Our experimental setup, manipulating word meaning, word frequency, and the distance between known words and new locations (messages) that participants needed to describe, allows us to quantify the role of these factors. Experiment 1 tests the degree to which message alignment and word frequency affect participants' use of words in the language, and Experiment 2 replicates and extends our results to a different layout of compass points.

2.1.1 Method

2.1.1.1 Participants

Eighty-three University of Wisconsin-Madison undergraduates participated for course credit (39 in Experiment 1, 44 in Experiment 2, 51 female; mean age: 18.6 years). With one exception in each experiment, participants were native speakers of English.

2.1.1.2 Materials

For each participant, eight novel words were drawn randomly from a set of 18 pseudo-words (*pim, dak, vorg, yeen, grah, skod, gled, veek, blit, peka, sarp, minada, hoon, clate, noobda, gorm, frabda, mog*) developed by Amato and MacDonald (2010). Each participants' set of eight words was randomly assigned to eight equidistant compass directions across the 360-degree face of a compass image: 15°, 60°, 105°, 150°, 195°, 240°, 285°, and 330° (see Figure 1). These compass positions were chosen to avoid translation to standard directions such as “north”.

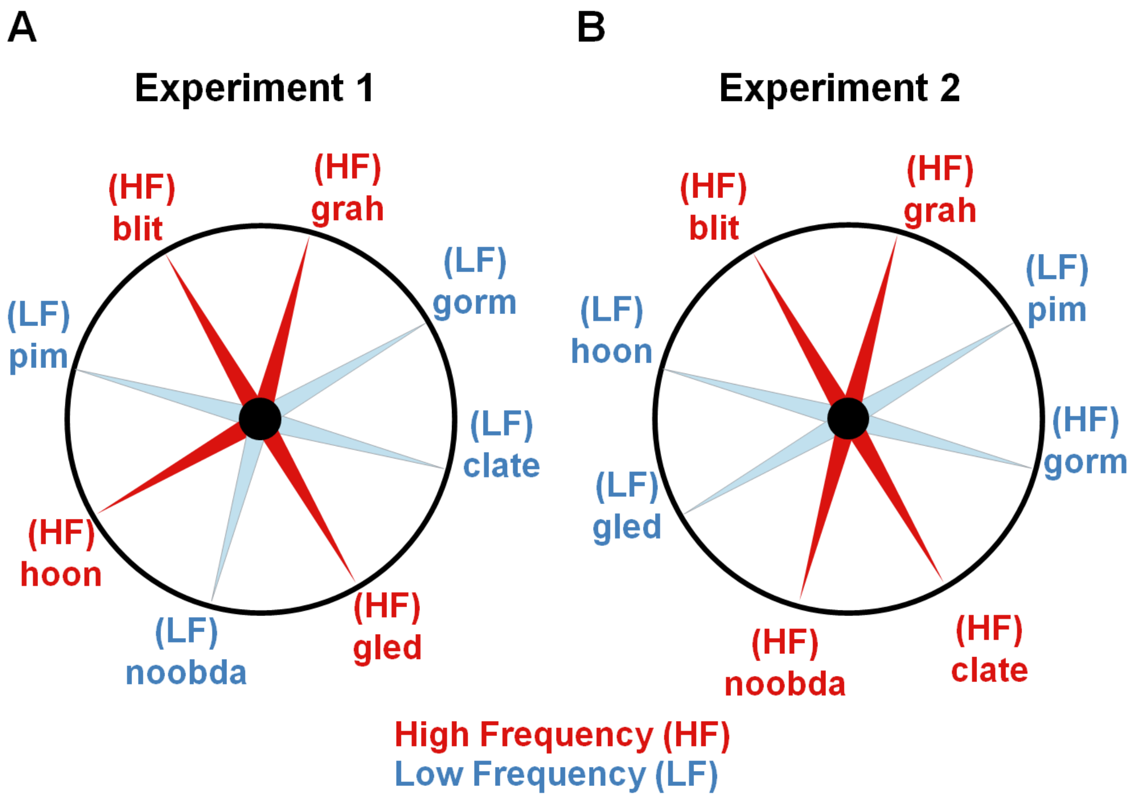

Figure 1. The eight compass directions learned and word frequency assigned to each compass direction (HF = high-frequency; LF = low-frequency) during training in (A) Experiment 1 and (B) Experiment 2. Novel words shown are examples; each participant got a different random assignment of words for the eight compass directions.

Each direction was assigned to a high-frequency or a low-frequency category in one of two counterbalanced compass arrangements. In Experiment 1 (Figure 1A), the arrangement of low-frequency/high-frequency words was designed to maximize the number of compass regions in which a high-frequency word was adjacent to a low-frequency word. The arrangement in

Experiment 2 (Figure 1B) allowed a higher proportion of critical trials in which a low-frequency word was the closest position to a probed compass direction.

2.1.1.3 Procedure

Participants were told that they were playing a game in which their job was to help elves hunt for gold by indicating a search direction for the treasure. The experiment consisted of a Training Phase, in which participants were taught novel words for the 8 compass directions (see Figure 1), and a Treasure Hunt described as a language game, in which participants were tested on angles that varied in distance from the trained compass directions. All instructions and trials were presented on screen, and participants typed all responses. Typing is known to be sensitive to word frequency, including in tasks with nonwords (Baus, Strijkers, & Costa, 2013; Kapatsinski, 2010; Barry & Seymour, 1988).

Training Phase. Participants were first presented with each compass direction and its assigned word, and they typed each of the novel words into a text box. Next, participants completed a Word Learning training in which they were presented with one of the eight compass directions and chose which of two words (a target and foil) matched that direction. Participants typed their response into a text box prompt and received immediate feedback on their answer. Critically, words in the high-frequency condition occurred four times more frequently (as both a target and as a foil) than the low-frequency words. The Word Learning Phase proceeded in blocks of 20 trials presented in random order. Each block contained four presentations of the High-frequency words and one presentation of the Low-frequency words. If participants scored below 80% on a 20-trial block, they were presented with another block of 20 Word Learning trials. Once participants achieved 80% accuracy on a 20-trial block, they proceeded to the Word Recall phase.

In the Word Recall phase, participants' explicit recall of the words for the eight compass directions was tested. Participants were prompted to recall each word via typed responses. If participants made an error, they returned to the Word Learning phase. The Training Phase continued until participants achieved 100% accuracy on all 8 words during the Word Recall trials. Thus, all participants entered the Treasure Hunt having learned the word for each compass direction, but having experienced high-frequency words four times more frequently than low-frequency words.

Treasure Hunt. The Treasure Hunt contained two phases. The first phase was described to participants as a game directing elves hunting for gold. The game was designed to test participants' naming responses to new compass directions. Following the game, the second phase consisted of two test blocks designed to re-check participants' knowledge of the original trained compass directions.

The first phase of the Treasure Hunt contained two trial types: Near Distance trials and Far Distance trials. In Near Distance trials, participants described randomly generated angles that were clearly nearer to one of the 8 compass directions than to others (see Figures 2A & 2C). Each test stimulus direction was 0° to 11° away from a compass direction. During this block, the 4:1 ratio of high-frequency to low-frequency words was maintained. Participants saw a compass direction near each high-frequency word 12 times and a compass direction near each low-frequency word 3 times, for a total of 60 test trials. For each trial, participants were asked to type a direction word into the text box based on the compass to direct a group of elves towards a hidden treasure. Trials timed out after 5s if participants did not begin typing.

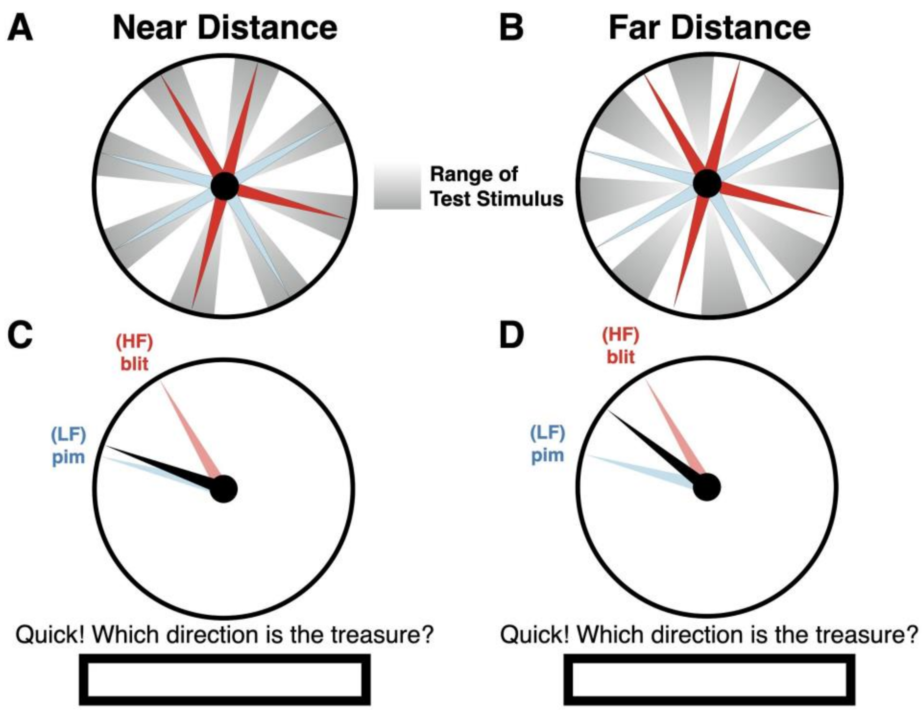

Figure 2. Gray shading indicates directions tested on (A) Near Distance and (B) Far Distance trials in Experiment 1, with examples of a (C) Near Distance and (D) Far Distance trial during the Treasure Hunt. For directions tested, the same distance manipulation was introduced in all experiments. The frequency of the trained compass directions are shown with the configuration used in Experiment 1. Red (dark) compass points were trained four times as often as blue (light) points in the training phase. For experiment trials (C & D), participants saw only the direction in black. The two nearest compass directions and words in light blue (low-frequency) and light red (high-frequency) are added for illustration purposes and were not visible to participants. After entering a word, a gold bag appeared and points were awarded.

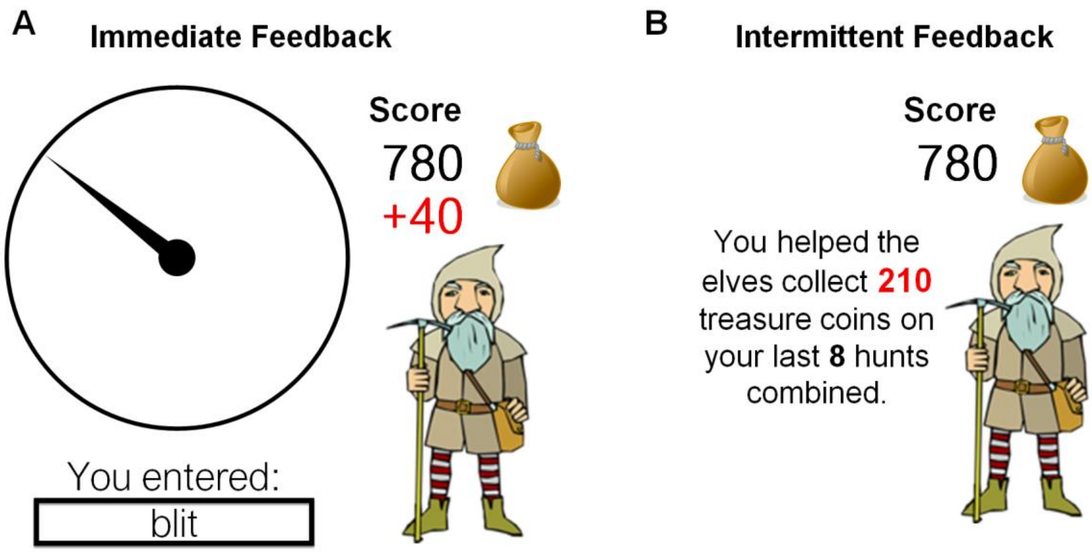

Figure 3. Example of immediate feedback (A) vs. intermittent feedback (B). In Experiments 1 and 2, participants received feedback after each game trial showing their response and gold coins earned. In Experiments 3 and 4, participants received no feedback after each response, and after every eight trials were shown the cumulative amount of gold coins earned during those trials.

In Far Distance trials, participants were tested with randomly generated angles that were close to the midline of two compass directions, between 11° - 22° from each, creating conflict between two words that could guide the elves (see Figures 2B & 2D; n = 64). On critical Far Distance trials, the angle fell between a low-frequency and a high-frequency word (Experiment 1: n = 48; Experiment 2: n = 32), though the compass direction always lay at least two degrees closer to one compass direction than another. The trial design and feedback were otherwise identical to Near Distance trials. In Experiment 1, participants saw the Near Distance trials, followed by the 64 Far Distance trials. In Experiment 2, participants first completed 20 Near Distance trials, to ensure that the task goal was clear to participants during the initial Treasure

Hunt trials. On the remaining trials, Near Distance trials (n = 40) and Far Distance trials (n = 64) were randomly intermixed.

Feedback. To incentivize fast and accurate performance, participants received feedback in the form of a score after each trial (Figure 3A), with points proportional to participant's message alignment (how close the word was to the typed compass direction) and speed (how quickly participants completed typing the word). Participants' base score varied from 0 to 45 points based on the distance of the tested angle from the word entered, with closer labels yielding higher points (45 points = no difference between tested angle and the entered word's compass direction; 0 points = tested angle is 45° or more away from the entered word's compass direction). This base score was then scaled based on the speed of participants' responses. For example, a difference in reaction time of 300ms corresponded to a change in base score by 0-2 points. Thus, while both speed and message alignment were emphasized, the scoring system weighed message alignment much more heavily than speed in assigning points. Participants received a score of 0 if they did not complete typing before the trial timed out or if their response was a word that named a direction more than 45° from the indicated compass direction.

Word retention. In the second phase of the game, participants were re-tested on their knowledge of the eight trained compass directions. The first eight retention trials (8 compass directions, randomized) preserved task demands of the previous trials and from the participants' perspective were simply additional trials in the game. These trials thus provided a covert Timed Retention test. Participants were then introduced to a new block of trials described as being separate from the treasure hunting game. This block served as an Untimed Retention test. Eight trials (the trained compass directions, randomized) appeared, and participants recalled the words without time limits. No feedback was presented.

2.1.2 Results

Word Training Performance

Participants' accuracy across all word learning blocks was high (Experiment 1: M = 95.2%, SD=3.1%; Experiment 2: M = 95.8%, SD=3.3%). On average, participants completed approximately 5 learning blocks (Experiment 1: M = 4.59, SD = 1.93; Experiment 2: M = 4.36, SD = 2.62) before reaching the required perfect performance on the recall test, progressing to the testing portion of the game.

Word Retention

See Table 1 for an overview of participants' recall of high- and low-frequency words in the final two retention tests. On the timed retention, participants showed greater accuracy and were faster to respond for high-frequency as compared to low-frequency words. Participants in both experiments showed high retention of both high-frequency (Experiment 1: M = 97.4%, 95% CI = [95.2%, 99.7%]; Experiment 2: M = 98.9%, 95% CI = [96.6%, 100%]) and low-frequency words (Experiment 1: M = 97.4%, 95% CI = [95.2%, 99.7%]; Experiment 2: M = 94.9%, 95% CI = [92.7%, 97.1%]) on the Untimed Retention. These results suggest that participants maintained accuracy on both high- and low-frequency words at the end of the Treasure Hunt, with an advantage for high-frequency words emerging under time constraints. Table 1 also displays reaction times, measured from the onset of a test prompt to the participant pressing the Enter key following typing the word. Participants also tended to be faster to respond to high-frequency than low-frequency words across Experiments 1 and 2. There was no main effect of experiment version (Experiment 1 vs. Experiment 2) on accuracy and reaction times and no interaction between experiment version and frequency for either block, suggesting the general learning patterns were similar across experiments.

Table 1.

Mean Accuracy and Reaction Times for High-Frequency (HF) and Low-Frequency (LF) Compass Directions in Untimed and Timed Retention in Experiments 1-4

Test Block	HF words	LF words	paired t-test
Experiment 1			
Timed Retention			
Mean Accuracy	91.7% [87.2%, 96.1%]	84.6% [80.2%, 89.0%]	$t(38)=1.99$, $p = .054$
Mean Reaction Time (in ms)	2347ms [2205ms, 2488ms]	2628ms [2486ms, 22770ms]	$t(38)=-2.46$, $p = .018$
Untimed Retention			
Mean Accuracy	97.4% [95.2%, 99.7%]	97.4% [95.2%, 99.7%]	$t(38)=0$, $p = 1$
Mean Reaction Time (in ms)	2755ms [2533ms, 2978ms]	3080ms [2858ms, 3303ms]	$t(38)=-1.81$, $p = .08$
Experiment 2			
Timed Retention			
Mean Accuracy	93.8% [89.5%, 98.0%]	84.7% [80.4%, 88.9%]	$t(43) = 2.63$, $p = .01$
Mean Reaction Time (in ms)	2366ms [2255ms, 2478ms]	2540ms [2429ms, 2653ms]	$t(43) = -1.93$, $p = .06$
Untimed Retention			
Mean Accuracy	98.9% [96.6%, 100%]	94.9% [92.7%, 97.1%]	$t(43) = 2.20$, $p = .03$
Mean Reaction Time (in ms)	2684ms [2473ms, 2895ms]	2982ms [2772ms, 3193ms]	$t(43) = -1.75$, $p = .09$
Experiment 3			
Timed Retention			
Mean Accuracy	91.8% [88.1%, 95.6%]	89.5% [85.8%, 93.3%]	$t(54) = 0.74$, $p = .46$
Mean Reaction Time (in ms)	2311ms [2220ms, 2402ms]	2447ms [2356ms, 2538ms]	$t(54) = -1.84$, $p = .07$
Untimed Retention			
Mean Accuracy	97.7% [95.4%, 100%]	95.0% [92.7%, 97.3%]	$t(54) = 1.43$, $p = .16$
Mean Reaction Time (in ms)	2581ms [2445ms, 2716ms]	2603ms [2468 ms, 2739ms]	$t(54) = -0.20$, $p = .84$
Experiment 4			
Timed Retention			
Mean Accuracy	92.4% [88.7%, 96.2%]	83.1% [79.4%, 86.8%]	$t(42) = 3.10$, $p = .003$
Mean Reaction Time (in ms)	2439ms [2340ms, 2538ms]	2601ms [2501ms, 2700ms]	$t(42) = -2.01$, $p = .05$
Untimed Retention			
Mean Accuracy	97.6% [94.1%, 100%]	91.9% [88.3%, 95.4%]	$t(42) = 2.03$, $p = .049$
Mean Reaction Time (in ms)	2542ms [2361ms, 2723ms]	2901ms [2720 ms, 3082ms]	$t(42) = -2.45$, $p = .018$

Note. Values in square brackets represent 95% within-participants confidence intervals (Morey, 2008).

Test Performance

Our main question was whether word frequency experience during training would increase the likelihood of participants overextending high-frequency words during test (the first phase of the Treasure Hunt), including in situations when a more aligned word (closer on the compass) was available. To investigate participants' tendency to overextend words, we focused specifically on low-frequency/high-frequency trials, in which a compass direction was tested in between a low-frequency and a high-frequency trained direction. We considered participants' likelihood of choosing the word for the nearest trained compass direction, dependent on whether that compass direction was a high- or a low-frequency word, while controlling for the distance from the nearest learned compass direction. As a conservative test, we focused exclusively on trials in which participants chose one of the two principal direction words within 45° of the stimulus direction (Experiment 1: 94.4% of responses; Experiment 2: 93.8% of responses). All of the patterns of findings remain identical if all low-frequency/high-frequency trials are considered.

Experiment 1. We fit a logistic mixed-effects model predicting the likelihood of choosing the nearest word from Word Frequency (centered; High = -0.5 vs. Low = -0.5) and the distance of the stimulus from the nearest compass direction. We included by-subject and by-item random intercepts as well as by-subject random slopes for word frequency and distance. The likelihood of choosing the nearest word decreased with increasing distance from the nearest compass direction, $b = -0.23$, Wald 95% CI = [-0.25, -0.20], $z = -17.00$, $p < .0001$. Crucially, controlling for distance from the nearest principal direction, participants were more likely to use the nearest word when it was a high-frequency word compared to a low-frequency word, $b = 0.71$, Wald 95% CI = [0.30, 1.12], $z = 3.37$, $p < .001$ (Figure 4A). This effect corresponded to an estimated 3.1° shift (95% CI = [1.3°, 4.9°]) in participants' decision boundary toward high-frequency words as compared to low-frequency words.

To ensure that this effect is not an artifact of participants' being slightly more likely to forget the low-frequency labels, we conducted a series of robustness checks. First, to account for the fact that participants varied in their final accuracy for each label in the Untimed Retention test, we fit the same model while controlling for participants' average final accuracy for the two compass directions to either side of the target angle on a given trial. We treated participants' average final accuracy on the two neighboring compass directions as a fixed effect, and also added a random slope for average final accuracy to the main model. The frequency effect in participants' choices remained highly similar (b = 0.71, Wald 95% CI = [0.31, 1.13], z = 3.41, p < .001) even after controlling for participants' average final accuracy on the two neighboring compass directions. In all models, we also fit models controlling for the character length of the nearest compass direction, since labels varied in length. All effects held after controlling for character length, and we found no significant effects of character length in any of the present experiments.

Next, as an even more conservative test of the robustness of the frequency effect, we refit the original logistic mixed-effects model including only participants who successfully named all compass directions correctly in the Untimed Retention at the end of the experiment (n = 34). The effect held even after removing all participants who did not perfectly name all compass directions at the conclusion of the experiment, b = 0.54, Wald 95% CI = [0.16, 0.92], z = 2.76, *p* = .006.

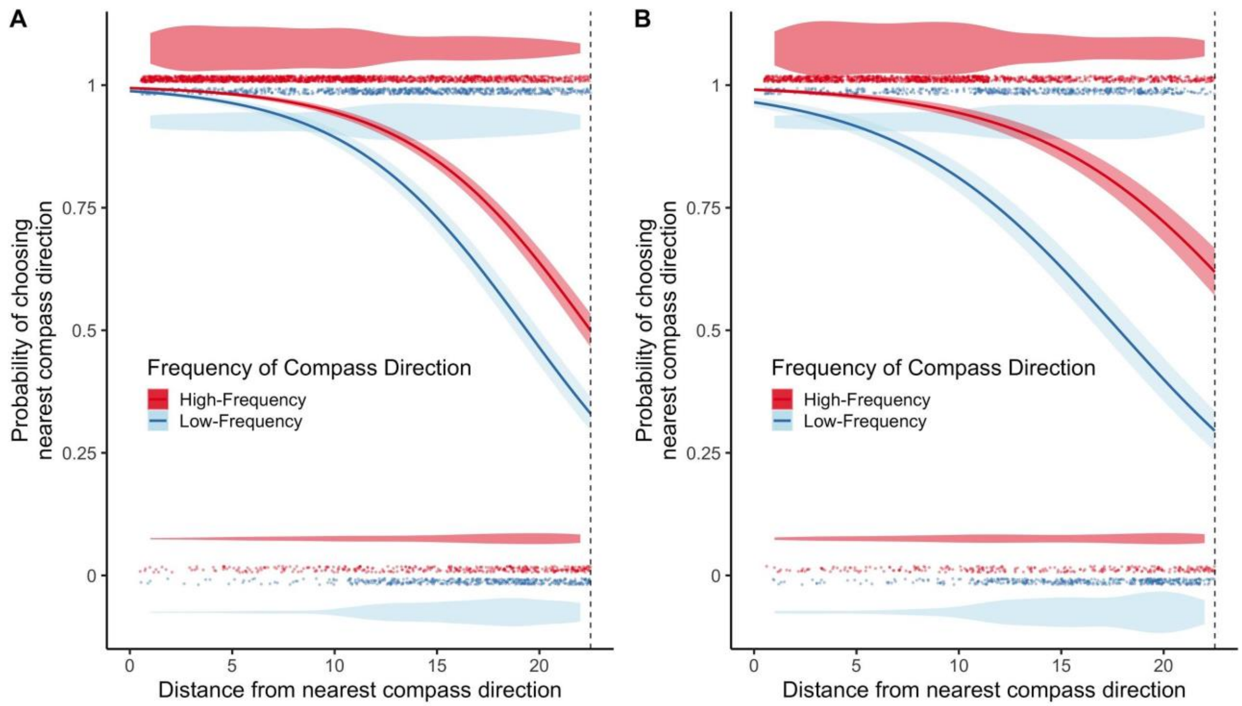

Figure 4. Probability of choosing the nearest compass direction between a low-frequency and high-frequency word in (A) Experiment 1 and (B) Experiment 2. X-axis reflects distance from the nearest compass direction, in degrees, when the nearest direction was a high-frequency word (red) vs low-frequency (blue). Error bands represent +1 / -1 SEs. Dots represent individual participant responses and violin plots show the density of the response distribution, with distributions at the top of the plot corresponding to choices for the word corresponding to the nearest compass direction, and distributions at the bottom of the plot corresponding to selection of the compass direction that is farther away.

Experiment 2. To test the impact of frequency on participants' overextension tendencies, we fit the same model as in Experiment 1. Controlling for angle distance from the nearest compass direction, participants were more likely to use the nearest trained word when it was a high-frequency word as compared to a low-frequency word, b = 1.36, Wald 95% CI = [0.75, 1.97], z = 4.37, p < .001 (Figure 4B). This effect corresponded to an estimated 7.2° degree shift (95% CI = [4.0°, 10.5°]) in participants' decision boundary for high-frequency words as compared to low-frequency words. There was no interaction between Low vs. Near Distance trials. The effect held after controlling for participants' average final accuracy on the two neighboring compass directions (b = 1.36, Wald 95% CI = [0.75, 1.97], z = 4.38, p < .001) and when including only participants (n = 35) with perfect accuracy in the Untimed Retention test at the end of the experiment (b = 1.19, Wald 95% CI = [0.66, 1.73], z = 4.42, p < .001).

2.2 Experiment 3

Experiments 1-2 show that participants' word use was affected by the frequency of potential responses. One concern with our findings is that they may have been driven by the explicit feedback given on every trial, as such consistent feedback is not a regular feature of natural language use. If explicit feedback is a key explanation for our frequency effect, then intermittent feedback should reduce or eliminate the effect. In Experiment 3, we greatly reduced the frequency of feedback, and obscured its directness, by providing only multi-trial, aggregated updates on scores.

2.2.1 Method

2.2.1.1 Participants

A new group of University of Wisconsin-Madison psychology undergraduate students (n = 55; 38 female; mean age: 18.9 years, *SD* = 0.88; 54 native speakers of English) participated for course credit. Four additional participants were excluded because they did not complete the study. The larger sample size was due to unintentional over-collection of data.

2.2.1.2 Design and Procedure

The experimental design and procedure was identical to Experiment 2, with the following differences. First, to prevent excessive perseverance on the learning block, a participant was limited to 10 learning test blocks before automatically advancing to the Treasure Hunt. Second, unlike in previous experiments, participants were instructed that they would receive intermittent feedback on their performance during the Treasure Hunt. A single, cumulative score was displayed every eighth trial (Figure 3B).

2.2.2 Results

Word Training Performance

Participants' accuracy across all pair learning blocks was high (M = 94.4%, SD=4.2%). On average, participants completed around 5 pair learning blocks (M = 4.64, SD = 1.99) before progressing to the Treasure Hunt.

2.2.2.1 Word Retention

Participants' accuracy and response times were similar for high-frequency and low-frequency words on both timed and untimed trials at the end of the Treasure Hunt, though performance for high-frequency words was numerically slightly higher (Table 1).

Test Performance

To test the impact of frequency on participants' overextension tendencies, we fit the same model as in Experiments 1 and 2. Controlling for angle distance from the nearest compass direction, participants were more likely to use the nearest word when it was a high-frequency word as compared to a low-frequency word (Figure 5), $b = 1.17$, Wald 95% CI = [0.58, 1.77], $z = 3.89$, $p < .001$. This effect corresponded to an estimated 6.0° degree shift (95% CI = [3.0°, 9.0°]) in participants' decision boundary for high-frequency words as compared to low-frequency words. When a high-frequency word was within this range, participants reliably selected it over the message-aligned, low-frequency word. The effect held after controlling for participants' average final accuracy on the two neighboring compass directions (b = 1.21, Wald 95% CI = [0.62, 1.80], z = 4.04, p < .001) and when including only participants (n = 42) with perfect accuracy in the Untimed Retention test at the end of the experiment (b = 0.98, Wald 95% CI = [0.34, 1.62], z = 2.98, *p* = .002).

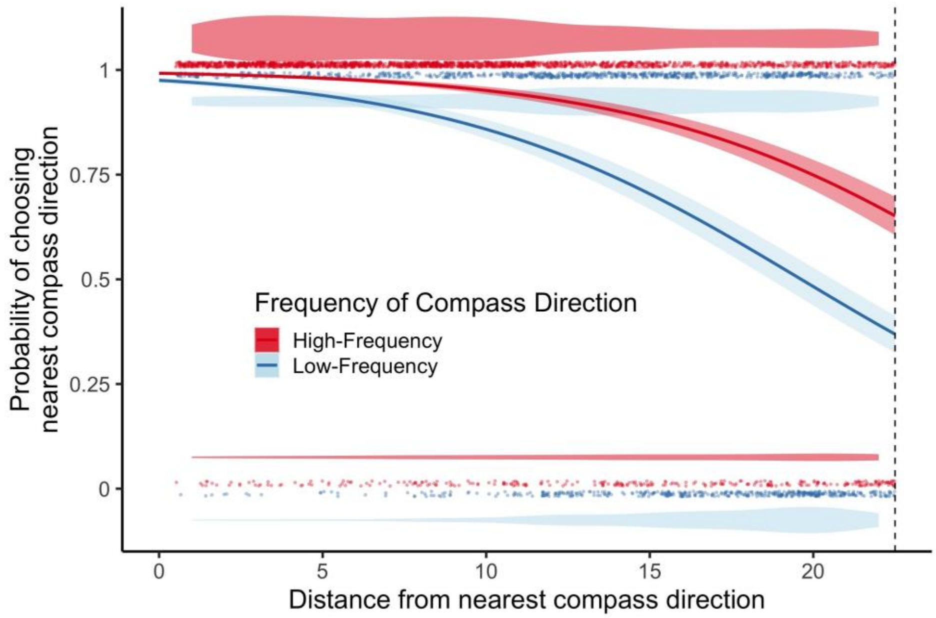

Figure 5. Probability of choosing the nearest compass direction between a low-frequency and high-frequency word in Experiment 3. X-axis reflects distance from the nearest compass direction, in degrees, when the nearest direction was a high-frequency word (red) vs low-frequency (blue). Error bands represent +1/ -1 SEs. Dots represent individual participant responses and violin plots show the density of the response distribution, with distributions at the top of the plot corresponding to choices for the word corresponding to the nearest compass direction, and distributions at the bottom of the plot corresponding to selection of the compass direction that is farther away.

2.3 Experiment 4

A concern about Experiments 1-3 is that familiarity of the trained compass direction is confounded with frequency of producing a word, because every presentation of a compass direction was accompanied by production of that direction's name. Thus, it is possible that the frequency effects in Experiments 1-3 are driven by familiarity with the visual stimuli rather than by word frequency. Therefore we added a new compass direction task to Experiment 4 in order to unconfound the frequency of visual stimuli and associated words.

2.3.1 Method

2.3.1.1 Participants

A new group of University of Wisconsin-Madison psychology undergraduate students (n = 43; 24 female; mean age: 18.7 years, *SD* = 0.88; all native speakers of English) participated for course credit.

2.3.1.2 Design & Procedure

The experiment design and procedure was identical to Experiment 3, with two main adjustments to the training phase described below.

Compass Practice Block. This new block preceded word learning and contained no words. On each trial, a compass circle was displayed and one of the eight compass directions appeared for 500 ms before disappearing. Next, a second randomly generated compass direction appeared from among the remaining seven compass directions. The participant was then instructed to adjust the angle to match the previous orientation by rotating the computer mouse click wheel. To ensure that participants only received exposure to the eight compass directions, the angle moved in 45° increments. Once the participant was satisfied with the angle position, they left-clicked with the mouse to end the trial.. Feedback then appeared on screen informing them of recall accuracy (correct or incorrect). Participants completed 100 trials, in random order.

In order to unconfound compass direction exposure and word frequency, the compass directions for which a low-frequency name would later be assigned appeared four times more often than the compass directions for which a high-frequency name would be assigned. This was true of both the targets displayed for 500 ms and also the starting position for participants' response. Thus after participants completed this Compass Practice block and the Word learning block, they had encountered each of the eight compass directions the same number of times. By contrast, the associated words for each compass direction were presented at either high- or low-frequencies, as in the previous experiments.

Word Learning Block. In order to match compass direction experience across high and low-frequency words, it was necessary to fix the number of learning trial blocks for participants. Across Experiments 1 - 3, the modal number of Learning Trial Blocks completed before learning all words and advancing to the Treasure Hunt was five. To match learning exposure to all compass directions, all participants were automatically advanced to the Treasure Hunt after five Learning Blocks.

2.3.2 Results

Compass Performance

Participants were highly accurate in their memory for compass directions (M = 97.7%, SD = 3.8%).

Word Training Performance

Accuracy across all pair learning blocks was high (M = 93.0%, SD = 7.7%).

2.3.2.1 Word Retention

Unlike in Experiments 1-3, high-frequency words were remembered reliably more accurately and identified more quickly in both Timed Retention and Untimed Retention tasks (see Table 1).

Test Performance

To test the impact of frequency on participants' overextension tendencies, we fit the same model as in Experiments 1 - 3. Controlling for angle distance from the nearest compass direction, participants were more likely to use the nearest word when it was a high-frequency word as compared to a low-frequency word (Figure 6), b = 0.70, Wald 95% CI = [0.10, 1.31], z = 2.27, *p* = .023. This effect corresponded to an estimated 3.4° degree shift (95% CI = [0.5°, 6.4°]) in participants' decision boundary for high-frequency words as compared to low-frequency words. The effect held after controlling for participants' average final accuracy on the two neighboring compass directions (b = 0.78, Wald 95% CI = [0.17, 1.39], z = 2.51, *p* = .01). However, the effect of frequency was not significant when including only participants (n = 33) with perfect accuracy in the Untimed Retention test at the end of the experiment (b = 0.42, Wald 95% CI = [-0.18, 1.04], z = 1.35, *p* = .18), indicating that the effect was more dependent on the inclusion of participants with imperfect final retention of all compass directions in Experiment 4.

Unlike in Experiments 1-3, participants could advance to the Test phase prior to achieving 100% accuracy on the eight compass directions during the Word Learning phase, since the number of Word Learning blocks was fixed at five. We therefore additionally investigated whether the effect depended on the inclusion of participants who had not yet learned all compass labels perfectly. The effect of frequency remained similar even after removing all participants who did not correctly label all words at the end of the Training phase (n = 29), b = 0.73, Wald 95% CI = [0.02, 1.45], z = 2.00, *p* = .045.

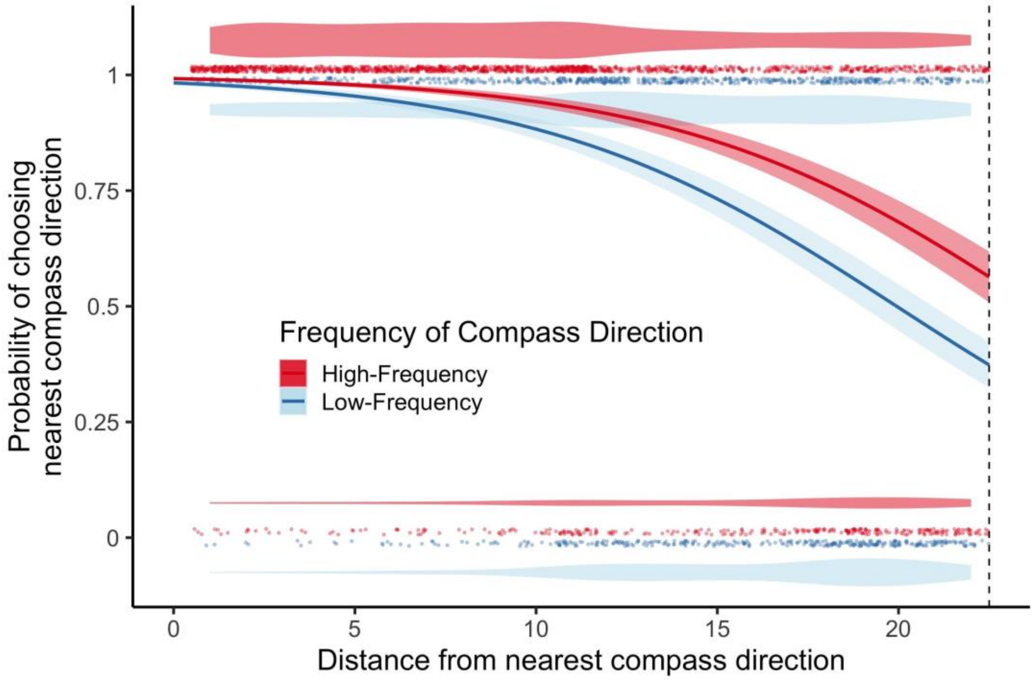

Figure 6. Probability of choosing the nearest compass direction on low-frequency/ high-frequency trials in Experiment 4. Error bands represent +1/ -1 SEs. Dots represent individual participant responses and violin plots show the density of the response distribution, with distributions at the top of the plot corresponding to choices for the word corresponding to the nearest compass direction, and distributions at the bottom of the plot corresponding to selection of the compass direction that is farther away.

2.4 General Discussion

We developed a novel language and communication game allowing us to quantify, for the first time, the degree to which language producers engage in probabilistic lexical selection and weigh both word accessibility and alignment with a message. In critical conditions across four studies, high-frequency words were favored over more precise low-frequency alternatives. This trade-off emerged even when participants knew low-frequency words well, as evidenced by performance on a post-test, and despite the fact that the point system in the communication game always rewarded message alignment more than speed. These results suggest that lexical selection can be characterized as “good enough” (Ferreira & Griffin, 2003), via probabilistic decision making that weighs message alignment and accessibility, broadly consistent with other accounts of action (Wolpert & Landy, 2012).

These results are consistent with prior evidence that in limited circumstances, phonologically-based accessibility factors influence the choice of words in language production (Ferreira & Griffin, 2003; Jaeger, Furth & Hilliard, 2012; Koranda & MacDonald, 2018). Our work extends to effects of word frequency and furthermore quantifies deviations from message alignment, owing to unique design features of our paradigm. The key factor was assigning words and communication goals to precise angles along a single, continuous semantic dimension, the compass directions. Then, by manipulating the relative frequency of the trained words, feedback from the language environment and experience with the visual stimuli, we were able to show that participants’ increased selection of high-frequency words also compromised message alignment.

These findings are related to several other language production phenomena, though more work is needed to determine whether similarities reflect similar underlying processes. For example, this work may provide insight into some types of speech errors. When speakers make word-substitution errors, such as saying *salt* when *pepper* is intended, a higher-frequency word tends to replace a lower-frequency intended word (Harley & MacAndrew, 2001). This outcome might reflect the same probabilistic decision making that we advocate here, in which a more accessible word is chosen over a more accurate but less accessible alternative. If so, a similar process may sometimes underlie speakers' use of high-frequency words such as *cat*, when an alternative, message-aligned word, such as *kitten*, is less accessible.

Our results have several implications for theories of language production. First, the finding that implicit production choices balance message alignment and accessibility are not predicted in models of language production in which lexical selection is influenced only by message alignment and not by linguistic form (e.g., Levelt et al., 1999). Other accounts suggest that word form has only a limited effect on lexical selection because selection is generally completed before word-form computations have gotten underway (e.g., Dell & O'Seaghdha, 1991; Goldrick, 2006). The consideration of`lexical selection as a form of probabilistic decision making may be broadly consistent with this view, but our findings that lexical selection shifts away from message alignment suggests that weighing of multiple factors may be more pervasive than these theories have posited.

Second, this work places constraints on the extent to which language production accommodates the listener's perspective (Arnold, 2008). Our results are an important counter to the view that producers routinely strive to benefit listeners, because selecting a word that poorly aligns with a message instead of a more precise word would pose problems for the comprehender. We did not study comprehension in this study, but points awarded in the communication game rewarded message alignment, yet participants still valued frequent words. Morgan et al. (2020) investigated the production of ungrammatical phrases such as "a word that I don't know what it means" and the consequences for comprehension. They found that these utterances increase comprehension difficulty, and their production appears to make the producer's task easier.

This work may also have implications for how language production processes modulate language change over time (MacDonald, 2013). Probabilistic decisions favoring more accessible, high-frequency words could account for some instances of diachronic change, where the meaning of accessible words changes over generations of use (Bybee, 2014). Recently, Harmon and Kapatsinksi (2017) showed that when two equally viable options could describe a novel meaning, the high-frequency one was more reliably extended. Our results suggest such extensions may occur even when high frequency words are initially less precise, predicting a more robust influence of frequency on diachronic change.

In summary, we provide an empirical demonstration that lexical selection reflects a trade-off between utility and efficiency, continuous with other motor behaviors across species. This builds on other work that investigates language as a case of more general action planning (e.g., Koranda et al., 2020). In contrast to claims in other theories of language production, lexical selection is good enough--a probabilistic compromise between message alignment and efficiency.

3 Phonological Overlap

When we have something to say, our message often may be realized through a variety of lexical, syntactic, and prosodic forms. For example, Brennan (1990) illustrated how a single request could be conveyed via 50 different sentences, each with a different mix of words and syntax. Because language affords so many alternatives, a key question in language production research concerns the factors and mechanisms that allow the producer to settle on one utterance over alternatives. A common view in language production research holds that the planning level called grammatical encoding is responsible for implicit decisions about words (lexical selection, such as settling on *sofa* vs. *couch*) and word orders and sentence structure (e.g., settling on the active *The cat scratched the sofa* vs. the passive form *The sofa got scratched by the cat*) (see Ferreira et al., 2018, for review).

Phonological processes controlling the form of words and the prosody of utterances are generally thought to play little role in word and syntax decisions in grammatical encoding, either because the processes of grammatical and phonological encoding are thought to occur in isolated stages (Levelt et al., 1991) or because feedback from phonological levels to grammatical encoding is weak because of the their timing: the process of mapping from message to words and word order is largely completed before phonological encoding gets underway (Dell & O'Seaghdha, 1991, 1992). A number of studies have explored the temporal dynamics of grammatical encoding processes, but with few exceptions (e.g., Jaeger et al., 2012b), researchers have not examined how phonological form affects lexical selection. This paucity of studies is unfortunate, because this work could be highly informative about the architecture and timing of lexico/grammatical and phonological encoding processes in language production. Here, we investigate this question via a novel word production paradigm that combines methodological elements of both word production studies and sentence production studies. We first review key findings and methodological choices in each of these literatures.

3.1.1 Phonological Effects in Word Production Studies

Much of the evidence for a distinction between stages of production comes from picture naming tasks (Johnson et al., 1996), in which participants view simple pictures and name them with a single word or short phrase, e.g. *boat, red boat*. Naming latencies, disfluencies, and errors reflect the ease of picture naming under various conditions. Common manipulations include presenting a word or another picture that participants are to ignore while naming the target picture. Phonological overlap between the target picture and the to-be-ignored stimulus sometimes increases naming latencies and sometimes decreases naming latencies for the target, depending at least in part on the relative timing of the appearance of target and distractor word onscreen (Jescheniak et al., 2020; Schriefers et al., 1990; Spalek et al., 2013). Phonological effects in this task are generally thought to have a locus in phonological encoding, where overlapping phonological codes from distractor stimuli typically affect the process of phonologically encoding the word(s) to be produced.

Convergent evidence for an effect of phonological factors on the ease of phonological encoding include interference effects in producing tongue twisters, sequences with many overlapping phonological patterns such as *She sells sea shells by the sea shore*. Greater disfluencies in tongue twisters compared to control conditions are observed even when participants are reading written phrases aloud, a task that requires phonological encoding but not lexical selection (Acheson & MacDonald, 2009). Importantly for our purposes, Sevald and Dell (1994) investigated different kinds of repetition patterns when speakers produced four-word sequences (e.g. repeating rhymes, consonant onsets, and others) and found that onset overlap (e.g. *pick, pun…*) increased articulation difficulty more than repeating rhymes (*pick, tick…*).

They attribute these differences in the effects of phonological overlap to the sequential nature of phonological planning and the different statistical structure of word onsets (more variable) and rhymes (more constrained) in English.

However, evidence also suggests that phonological encoding affects retrieval speed. Initiation latencies vary with a word's phonological neighborhood (Fox et al., 2015). For example, the word *ball* has a large amount of phonological overlap ("neighbors") with other words in English, including *bought, bond, balloon, tall, hall*, etc., while other words such as *broom* have fewer phonologically overlapping neighbors (Yates et al., 2004). Other factors being equal, picture naming is slower and more error-prone when the picture's name is in a denser phonological neighborhood than a sparser one. This increased difficulty is thought to stem from interference from other common phonological forms during the phonological encoding process (Baese-Berk & Goldrick, 2009; Dell, 1986; Fox et al., 2015).

With clear evidence for effects of phonological interference (and occasionally facilitation) on phonological encoding, a key question becomes whether phonological form also affects grammatical encoding. Here reaction time measures such as initiation latency may be less useful. V. Ferreira and Griffin (2003) used a different approach, examining rates of lexical selection errors in picture naming under different phonological conditions. Their technique mixed sentence fragments to be read silently and pictures named aloud. In critical trials, the sentence fragment strongly biased expectations for a certain word (a homophone in critical conditions), but in place of the expected word, a semantically unrelated picture appeared. For example, one sentence led to the expectation of the word *none*, and a picture of a priest was shown. Ferreira and Griffin investigated whether expectation of the word *none*, a homophone of the word *nun*, would interfere with correctly naming the picture as *priest*. Indeed, participants were more likely mislabel the picture as *nun* compared to a control condition. These results suggest that the phonological form of the recently primed homophone was accessible during the lexical selection process for a semantically unrelated picture. Thus at least in the case of phonological activation of homophones, phonological information appears to affect lexical selection.

Another approach to studying phonological influences on lexical selection has been to study situations in which lexical retrieval is difficult, as when someone is in a tip-of-the tongue (TOT) state. In the laboratory, TOTs are typically elicited in picture naming tasks or naming words in response to definitions (Abrams et al., 2003; Reilly & Blumstein, 2014a). The pictures or definitions are usually fairly obscure, so that participants typically know the person or word but may have difficulty in retrieving it in response to the prompt. In these situations, phonological cues to the intended word tend to improve word retrieval, including in paradigms in which participants simply read prime words before a definition prompt is presented (Abrams et al., 2003; Reilly & Blumstein, 2014a). The benefit from phonological overlap in the prime words supports a role for phonological information in word retrieval, at least in situations when retrieval may be delayed and difficult.

A notable feature of these TOT studies and Ferreira & Griffin's (2003) homophone interference studies is the focus on the content of participants' productions to ground the effects at grammatical encoding processes. This same focus on utterance form is a hallmark of studies grammatical encoding processes in sentence production, which we review next.

3.1.2 Accessibility in Sentence Planning

Investigations of implicit decisions in sentence production consider the proportion of utterances among viable alternatives. Some studies address the processes that settle on word order and syntactic, such as when someone describes two pets as *a dog and a cat* vs. *a cat and a* *dog* or an active vs. passive sentence such as *The dog chased the cat* vs. *The cat was chased by the dog*. Word order and syntactic structure appear to be strongly shaped by the ease with which words can be retrieved from long term memory, often termed *accessibility* (Bock, 1987). There appear to be a range of factors that affect accessibility; studies in a number of languages have demonstrated that words reflecting more salient concepts for the producer, more frequent, animate, and given in the discourse tend to be produced earlier and/or in more syntactically prominent positions such as grammatical subject (Bock & Warren, 1985; Bock, 1987; Christianson & F. Ferreira, 2005; Gennari et al., 2012; Prat-Sala & Branigan, 2000). This work has been interpreted to support incremental sentence planning processes, in which words that are more ready to incorporate in the utterance plan are prioritized over others.

Some sentence production studies have explored the extent to which phonological information influences these grammatical ordering processes, providing mixed results. For example, using a sentence recall paradigm, McDonald et al. (1993) investigated whether word length and syllable stress—features of phonological form—influenced word ordering both in conjoined noun phrases like “the key and the manager” and in whole sentences. They found robust effects of noun animacy on word order but little evidence that word order is affected by properties of a word’s phonological form. McDonald et al. speculated that their recall task might tend to dampen subtle effects of phonological form on word order in their studies, and indeed, some other tasks have yielded different results.

Bock (1987) combined word priming and picture description paradigms to investigate whether the phonological form of a word that is read aloud influenced the word order used to describe a picture presented immediately afterwards. For example, participants read aloud a printed word and then saw a picture of a bee on an unhappy man’s arm, which could be described in an active form *A bee is stinging a man,* or a passive form *A man is getting stung by a bee.* The printed word was semantically unrelated to the picture, but its onset was related to one of the pictured entities, either the man (phonologically overlapping word *mat*) or the bee (overlapping word *beet*). Bock found that picture descriptions tended have a word order in which the phonologically overlapping word was placed later in the utterance. She argued that the word onset overlap between the prime word and the name of one of the elements in the target picture increased the difficulty of phonological encoding of the overlapping word, which in turn delayed the inclusion of that word in the utterance plan (see also Bock & Irwin, 1980; Jaeger et al., 2012a). This result suggests that word order choices during grammatical encoding can be affected by a word's phonological form (though see Levelt & Maasen, 1981).

There are a small number of studies that have investigated the role of phonological information on lexical selection in fluent language production. Rapp and Samuel (2002) investigated phonological overlap in a written sentence completion task and found that a rhyme prime word earlier in the sentence increased the likelihood that participants completed the sentence with a rhyming word compared to a control. Jaeger et al. (2012b) investigated the effects of phonological form in a task in which participants produced an entire sentence. They developed animated videos in which one character transferred an object to another character. The videos were designed to manipulate the degree of phonological overlap between the subject character's name, the verb, and the direct object; for example, *Patty, passed, and pan* contain identical consonant onsets and vowel. The actions in the videos could be described with several different verbs (*gave, handed, passed*), and Jaeger et al. investigated whether participants' choice of verb in video descriptions varied with the phonological form of the subject character name and direct object. They found no effects of phonological overlap with the direct object on the preceding verb form, but they did find an effect of overlap with the name of the grammatical subject, which immediately preceded the verb. Unlike Rapp and Samuel's facilitative effect of rhyme overlap on selection, Jaeger et al. found an interference effect of overlap of word onsets: participants tended to avoid verb forms that overlapped with the character's name, such as *Patty passed* and *Hannah handed*, instead favoring sequences without such overlap, such as *Hannah passed* and *Patty handed*. Together, these effects open the possibility of an effect of phonological overlap on lexical selection, with interference from onsets and facilitation from rhymes, aligned with the patterns that Sevald and Dell (1994) observed with latency measures.

3.1.3 Phonological Influences on Lexical Selection

Considering all the above work, there appears to be some evidence for phonological effects on lexical selection. The studies that offer the clearest evidence have two key features. First, choice of utterance form (that is, the word or word order that is produced) is the primary dependent variable, and most studies contain materials for which several different utterance forms are valid responses. For example, Bock (1987) used pictures that could be described with two different word orders or sentence structures and investigated how the patterns of word order changed as a function of a prime word that overlapped phonologically with one of the pictured elements. Other studies examined lexical selection, either via measuring errors as a function of phonological primes (V. Ferreira & Griffin, 2003) or in situations in which several different words provided a valid response (Jaeger et al., 2012b; Rapp & Samuel, 2002). Developing materials and task demands that allow multiple viable responses creates a potentially more sensitive arena to see effects of subtle influences on those responses. This strategy has been used extensively in psychological research, including in studies of word production (Peterson & Savoy, 1998) and grammatical agreement production (Haskell & MacDonald, 2005; Lorimor et al., 2018; Mirković & MacDonald, 2013).

The second feature, at least of the studies of lexical selection to date, is that the materials in some studies are fairly limited, which in turn may limit the generality of the effects. Studies of phonological interference on word order (e.g., Bock, 1987) afford a wide range of stimuli, but studies of lexical selection have been constrained in their range of items, for example by the number of available homophones (Ferreira & Griffin, 2003), or by the number of available verb synonyms, which must be repeated across many trials (Jaeger et al., 2012b). Here we aim to develop a paradigm to investigate phonological influences on lexical selection that retains the choice of utterance form as the primary dependent measure, but with both a different and a wider range of stimulus materials than has been available to date.

Our method focuses on single word picture naming, but rather than presenting pictures with a unique preferred name (that is, having "high name agreement," as it is termed in this literature, Alario et al., 2004; Johnson et al., 1996), we present pictures for which two or more single-word names are acceptable (i.e., pictures with low name agreement, Vitkovitch & Tyrrell, 1995). Experiment 1 is a large norming study in which we identify a set of pictures that have this character, for example a picture for which both *pail* and *bucket* are commonly given names, and a picture that elicits both *cup* and *glass* names. Our subsequent experiments then place these pictures in different kinds of contexts manipulating phonological overlap. Building on phonological interference priming studies in which word onsets overlap (Bock, 1987, Jaeger et al., 2012b), we present pictures preceded by a semantically unrelated word, which participants read aloud. In critical conditions, this prime word shares the consonant onset and typically more phonemes (but not the rhyme) with one of the names of the picture that follows it. For example, for the glass/cup picture, the prime word *glare* overlaps phonologically with the word *glass,* and the word *cast* has consonant onset overlap with the word cup.

Bock (1987) found that onset overlap between the prime word and one entity in a picture affected word ordering in description the picture, with the onset-overlapping pictured entity tending to appear later in the description. If phonological information also affects lexical selection in our similar paradigm, then we would expect to see changes in the naming distribution for low name agreement pictures as a function of phonological prime words. For example, we predict that the prime word *cast* will reduce the rate at which participants name the glass/cup picture as *cup* relative to a baseline condition or a condition in which the pre-picture word has phonological overlap with the name *glass*. If so, this result would suggest that lexical selection processes are affected by the phonological content of recently uttered words, biasing selection away from words with phonological overlap with recent utterances. That result would support a production architecture in which the difficulty of phonological encoding could influence grammatical encoding processes (Dell & O'Seaghdha, 1991, 1992).

3.2 Experiment 1: Baseline Name Agreement

In order to investigate whether phonological overlap changes speakers' naming patterns, we must first have a set of pictures that afford several different names, and we must also know the baseline naming patterns for these items, so as to identify whether experimental manipulations shift the naming patterns from this baseline. To meet these goals, we collected Dominant and Secondary name agreements for a wide range of pictures. Prior studies of picture naming exist (Snodgrass & Vanderwart, 1980; Szekely et al., 2005), but we conducted our own study with participants from the same population as in our subsequent experiments. This is an important step, because patterns of object naming may vary over time and across regional dialects. Because many of our items overlapped with items in another study of name agreement (Szekely et al., 2005), we report some comparisons to previous results.

3.2.1 Method

3.2.1.1 Participants

Native English speakers (n = 45) from the University of Wisconsin-Madison completed the experiment for course credit or pay. Data from one additional participant was excluded because of technical difficulties.

3.2.1.2 Materials and Procedure

Stimuli were black and white line drawings (n = 83) used in previous studies (Snodgrass & Vanderwart, 1980; Szekely et al., 2005) or from work freely available on the internet (3). Pictures were assembled into four different lists, each with a different random order of pictures.

Participants sat in front of a computer monitor and a microphone. Instructions and pictures were presented using E-Prime 2.0 (2012). Participants were instructed that they would see pictures on screen one at a time, and that they should name each picture as quickly and accurately as possible. Each trial began with a center fixation cross which appeared for 500ms, followed by a picture centered on screen (620 x 420 pixels). The picture remained on screen until the participant named it, at which point an experimenter, who sat next to the participant, manually advanced the trial.

3.2.1.3 Coding

Blind coders transcribed verbal responses to pictures. They were instructed to transcribe the first intelligible attempt to name a picture and annotate errors, such as hesitations and articles (um/uh/a/an), dysfluencies (articulation errors and false starts—a partial word uttered first then corrected), delays (more than 3s), and corrections (if a full word was uttered first then corrected), or no utterance. Anomalous transcriptions were identified by the first author across all experiments and evaluated for typos, or misheard names, or transcriptions using the wrong audio file (e.g. “balloon” for a picture of a glass, total fixed = 28).

Criteria from Szekely et al. (2005) were used to establish name agreement. Unique tokens (i.e., variation) were defined as distinct base word forms, such as *bricks* vs *wall*, plural morphemes (*bricks* vs. *brick*) or abbreviations or expansions (*wall* vs *brick wall*). Full corrections in which the participant first gave an incorrect name and then corrected it, were treated as correct trials and included in all analyses. Error trials (141), trials with no utterance, (10), or and trials where utterance onset was less than 400ms or greater than 4000ms (54), were all excluded.

3.2.2 Results and Discussion

In this study we aimed to identify pictures with at least two viable alternatives. Productions for each picture were tallied and ranked to identify the two most common names (Dominant and Secondary names), and proportions of these two responses (see Figure 1).

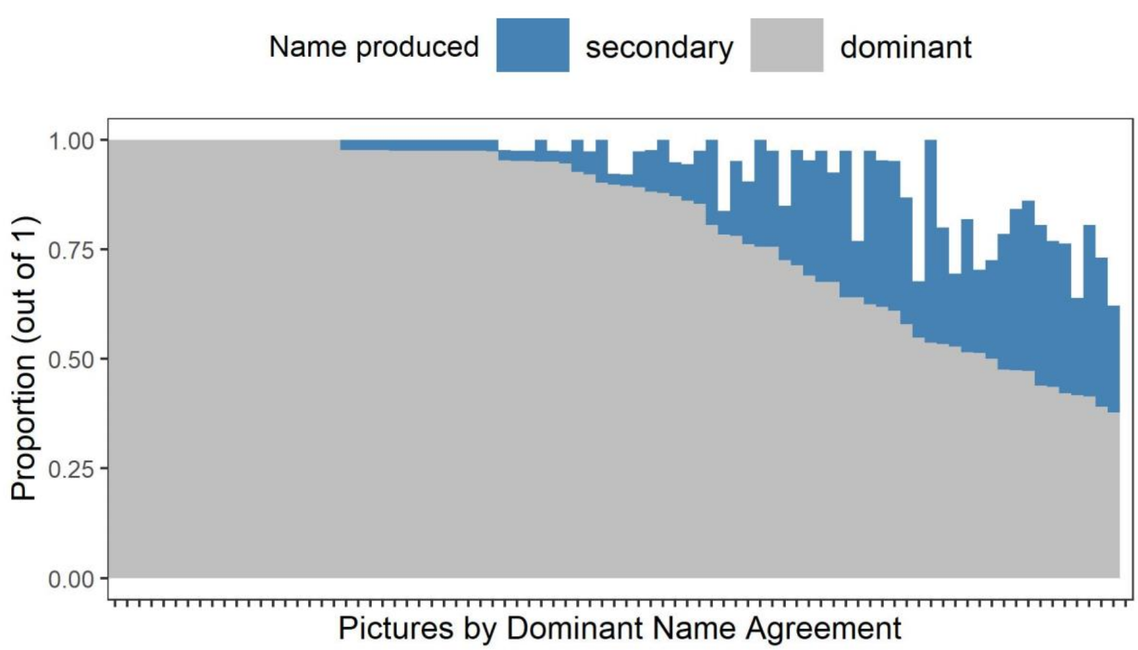

Figure 1. Proportion of dominant (gray) and secondary (blue) names for the 86 pictures tested in Experiment 1. Pictures are ordered on the x axis by rate of dominant name agreement.

In addition to the identified Dominant Name, we calculated naming distributions with a metric that accounts for unique responses and degree of variability (H-statistic; Snodgrass & Vanderwart, 1980), where k refers to unique responses and p_i is the proportion of participants giving that response.

$$H = \sum_{i=1}^{k} p_i \log_2(1/p_i)$$

Word usage has been known to depend on factors such as region and age cohort (Labov, 2011). To compare our population characteristics with others', we compared item-level means and H-statistics with that of Szekely et al. (2005). Correlations with the Szekely et al. norms were high for both probability of the Dominant name (0.83, $p < .01$) and Naming Distribution (0.85, $p < .01$). Dominant names from our data matched the dominant names identified in Szekely et al. for all but four pictures: *brickwall* (Szekely et al.—*bricks*), *ice cream* (*ice cream cone*), *skate* (*roller skate*), *cube* (*block*). Of these, only *cube* was used as a target picture in subsequent experiments.

These baseline naming proportions provide independent criteria for identifying targets in Experiments 2 and 3, with more than one common response and an empirical basis to the manipulation categories of Dominant and Secondary name overlap.

3.3 Experiment 2: Dominant vs. Secondary Phonological Overlap

To investigate the role of the effect of the phonological form of prior utterances on lexical selection in subsequent utterances, Experiment 2 combined reading aloud trials with picture naming trials, using a subset of the picture stimuli from Experiment 1 as target low name agreement pictures. The key question was whether the word form, specifically the phonological onset of a word read aloud, would affect naming distributions for a picture on the subsequent trial. If so, then this result would argue for a phonological influence on lexical selection.

3.3.1 Method

3.3.1.1 Participants

Thirty-five native English speakers from the University of Wisconsin-Madison completed the experiment for course credit or pay. Data for additional two participants were excluded because of failure to follow instructions (1) and technical difficulties (1).

3.3.1.2 Materials

Stimuli contained a mix of pictures and printed words, creating targets for naming, phonological overlap items (primes) hypothesized to influence target picture naming, and filler items to separate critical trials in the study and obscure any prime-target phonological relationships.

*Target Pictures and name properties***.** Twenty target pictures were selected for having two strong alternative names (Dominant and Secondary names) that began with different word onsets, as determined in Experiment 1 (see Appendix). Constrained by these criteria, this set allows ten observations per condition in the study; this number of observations is not enormous but affords more power and a different set of items compared to some previous investigations. Seventeen of these pictures were also normed by Szekely et al. (2005) and three were new, normed in Experiment 1.

All dominant and secondary names for the target pictures were one or two syllables in length, and syllable length was roughly equally distributed across dominant and secondary names. The semantic relationship between dominant and secondary names for each picture varied across items; for example, some were near synonyms (*glass/cup*) and others category level alternatives (*parrot/bird*).

Word primes. For each target picture, two words were selected to serve as primes. One word overlapped in form with the target picture's dominant name and the other with its secondary name. Overlap always included the initial phoneme (onset overlap), and often extended to additional phonemes, in order to create as robust an overlap manipulation wherever possible, similar to the degree phonological overlap in Jaeger et al. (2012b). For example, for the glass/cup picture, which elicited *glass* as the dominant name and *cup* as the secondary name in Experiment 1, the word *glare* had phonological overlap with the dominant name (i.e., Dominant Overlap) and the Secondary Overlap word was *cast.* Similar to target names, the prime words were all 1-2 syllables in length and nearly equally matched on length across the Dominant and Secondary Overlap conditions. No attempt was made to match prime words for frequency or other factors, because these items are the source of phonological codes and are not themselves under investigation.

Filler trials. Forty pictures and twenty words were selected for filler trials. Pictures were other line drawings from Experiment 1 with a range of name agreement. Words were selected for having relatively similar frequency, concreteness and orthographic complexity, using a word generator designed for psycholinguistic research (Friendly, 2018).

Design and Trial sequence. All participants saw the same 20 target pictures. The trial sequence is illustrated in Figure 2. It began with four filler trials selected randomly without replacement from the list of word and picture fillers. Next came a printed word followed by a target picture. The figure shows the Dominant (*glare*) and Secondary (*cast*) Overlap words for the target picture shown; participants saw only one of these words. For each participant, ten target pictures were preceded by a word with Dominant overlap, and ten by with Secondary overlap, counterbalanced across participants. Participants completed one of four counterbalanced lists that distributed alternatives of Dominant vs Secondary overlap for each target picture and order of filler and target trials.

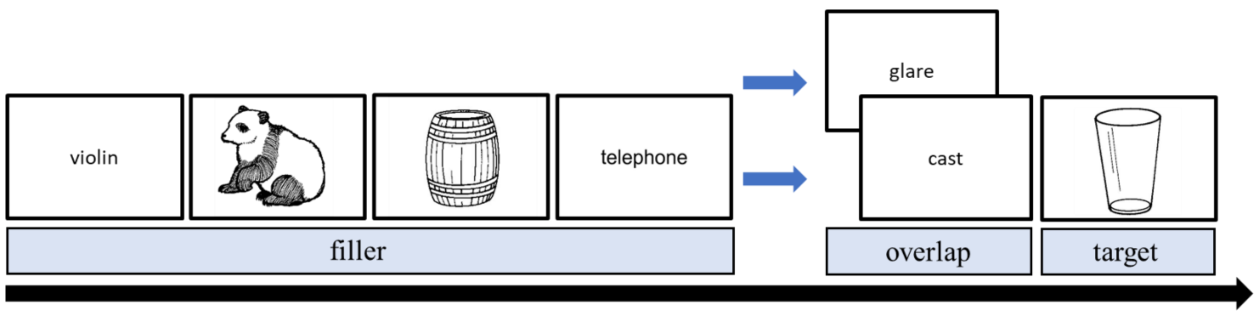

Figure 2. Example of trial order in Experiment 2. Order of trials were blocked such that a random mixture of four word and picture filler trials preceded a prime word (identified with arrows) and a target picture. In the example target shown, *cup* and *glass* are the Dominant and Secondary names for the target picture, and the phonologically overlapping words *cast* and *glare*, Dominant and Secondary overlap, respectively.

3.3.1.3 Procedure

The experiment set up and procedure was identical to Experiment 1, with the addition of instructions for reading printed words on a screen. Participants were asked to read aloud words or produce names for pictures as quickly and as accurately as possible. After completion of all trials, the experimenter asked participants what they thought the study was about, then debriefed them on the nature of the study. The study was completed in about 20 minutes.

3.3.1.4 Coding and Analysis

All target utterances were transcribed and checked using the same method as Experiment 1. Two trials were excluded for partial or disfluent utterances, and two for incorrectly reading aloud the preceding overlap word. Utterances that did not correspond to normed Dominant or Secondary names (n = 91) were excluded from analyses. Inclusion of these observations did not change the pattern of results reported here.

A few participants' post-experiment responses to questions indicated either implicit or explicit awareness of target pictures having alternative names (n = 4), or that neighboring trials had phonological overlap with some picture names (n = 4). None reported any awareness of the link between the two, or unnatural strategies for naming.

3.3.2 Results

Linear mixed logit models (Jaeger, 2008) were used for all analyses in order to predict participants' lexical selection in each task. Models started with maximal random-effects structures (by-item and by-subject random intercept and random slopes for predictor variables). In cases where a model failed to converge, planned steps were taken to achieve convergence (Barr et al., 2013). To assess the effect of phonological overlap on selection decision, productions of Dominant (1) and Secondary (0) names were estimated as a function of the phonological overlap manipulation, controlling for trial order, and the interaction of trial order and overlap manipulation. Not surprisingly, dominant picture names were produced overall more often than secondary names, and trial order did not significantly predict the rate of dominant vs. secondary picture name choices (see Table 1).

These analyses also address our question of interest concerning whether phonologically overlapping words presented prior to picture naming affect participants' choices of name. Table 1 shows that the Dominant vs. Secondary Overlap factor was a reliable predictor of picture naming. As Figure 3 shows, the effect of phonological overlap was such that reading a word that overlapped in onset with the dominant name of a picture resulted in fewer dominant names produced relative to the condition in which participants read a word that overlapped with the secondary name of a picture. For example, participants were less likely to say *cup* immediately after producing the word *cast*, than after reading aloud the word *glare*.

Parameter	Estimate	SE	z	*p*
Intercept	0.91	0.26	3.44	< 0.01
Dominant vs Secondary Overlap	-0.86	0.26	-3.45	< 0.01
Trial	0.52	0.52	1.00	0.32
Overlap * Trial	0.67	0.83	0.80	0.42

Table 1. Model estimates for naming distribution of Dominant and Secondary picture names under Dominant and Secondary overlap. Results of mixed-effects logistic model (Jaeger, 2008) predicting lexical selection of dominant name (reference group) or secondary, controlling for trial (centered).

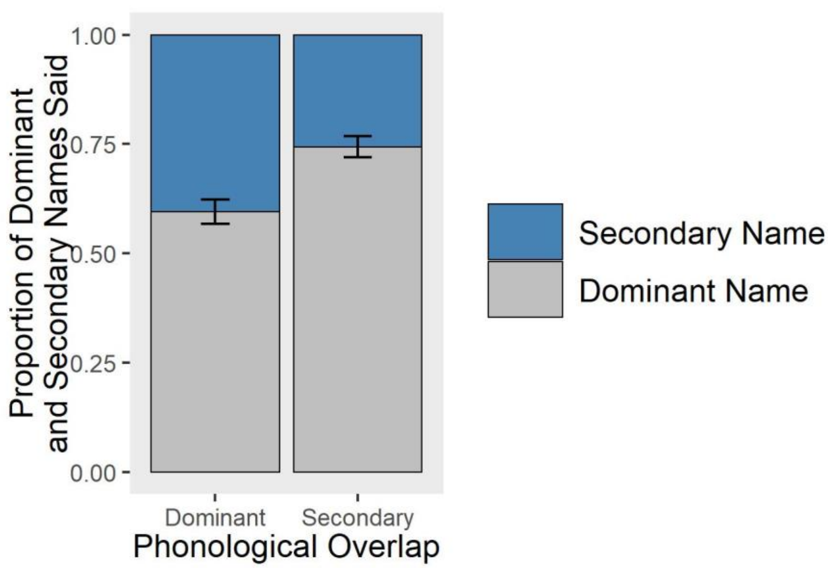

Figure 3. Proportion of Dominant (blue) and Secondary (gray) names produced for target pictures as a function of phonological overlap (Dominant vs Secondary).

3.3.3 Discussion

This experiment showed that the phonological form of a recently uttered word affected participants' picture naming performance. We observed an overlap effect, such that naming probability decreased for picture names that shared phonological overlap with the prime word, compared to the probability of saying a name without overlap to the prime. This result is consistent with other findings of phonological information on lexical selection, specifically interference from phonological overlap (Jaeger et al., 2012b)

We sought to illuminate the nature of these overlap effects in Experiment 3. In Experiment 2, the word preceding a target picture contained some kind of phonological overlap in all critical trials, so that the key comparison was overlap with the dominant vs. secondary name. As a result, it is impossible to know whether phonological overlap with dominant names, secondary names, or both, affect production rates. Identifying the reach of the phonological overlap effect could be important for identifying the range of phonological influences on lexical selection. In Experiment 3, we include a control condition in which the word preceding the target picture has no phonological onset overlap with either picture name. This addition will allow us to see the degree to which the Dominant and Secondary phonological overlap conditions each change naming distributions from the control condition.

3.4 Experiment 3: No Interference Condition

The main addition to this experiment was a No Overlap control condition, in which the onset of the word preceding the picture had no phonological overlap with the onset of either the dominant or secondary picture name. Some small improvements to the materials and procedure were also added. As this experiment retains Experiment 2's two overlap conditions in, it also affords an opportunity to replicate the effects in Experiment 2.

We also collected naming initiation latencies in this study. While we have emphasized the importance of selection data in these studies (e.g., the production of *glass* vs. *cup* under different conditions), the initiation latencies may prove to be informative in conjunction with the naming distributions. For example, it is reasonable to expect that initiation latencies for trials in which the participant produced the secondary name would be longer than latencies in which the dominant name is produced. That is, situations in which the speaker settled on the secondary name would generally be assumed to be ones in which, for whatever reason, the dominant name was less accessible; such situations could be expected to yield longer initiation latencies. A second possible outcome is that the secondary name interference condition, which increased rate of dominant name production in Experiment 2, might yield shorter naming times of the dominant name compared to other conditions. The source for this prediction comes from studies of the basic effect of name agreement on picture naming, where pictures with high name agreement (a single very dominant name) tend to have shorter name initiation latencies than those for pictures with low name agreement (Barry et al., 1997). The secondary interference condition, by reducing the viability of the secondary name, may promote not only more choice of the dominant name but also shorter times to initiate the naming response.

3.4.1 Method

3.4.1.1 Participants

Native English speakers (n = 101) from the University of Wisconsin-Madison completed the experiment for course credit or pay. No participants were excluded.

3.4.1.2 Materials and Procedure

The stimuli and procedure were identical to those of Experiment 2, with the following exceptions. A) A control No Overlap condition was created for each target picture. Control words were selected so that the first two phonemes of the word did not match that of either dominant or secondary picture names. Like Dominant and Secondary Overlap words, control words were semantically unrelated to the target picture and were either one and two syllables in length, with one longer exception; see Appendix for all stimuli. B) To maintain even distribution of conditions within participants (6 trials for each type of overlap), two target pictures were removed (*cup/mug, bird/pigeon*). C) A small number of prime words in the dominant and secondary overlap conditions were replaced to better standardize phonological overlap between the prime word and picture names. For example, the secondary overlap prime word for the glass/cup target, *cast*, was changed to *cut*, to remove the potentially facilitative effect of a near-rhyme between *glass* and *cast.* D) Fixation crosses were added between each trial.

3.4.1.3 Coding

All target utterances were transcribed and coded using the same method outlined in Experiment 2. Participants' productions of the prime words were also verified, and none were excluded. Trials were excluded from analyses where no target utterance was made (n = 4) or contained partial or disfluent utterances (n = 128; total excluded = 132). Audio from responses to target pictures were processed using automated software to detect speech onset (Roux et al., 2017). Target picture utterances with onset latencies greater than 3 seconds were excluded (n = 162).

3.4.2 Results

We first addressed the question of whether the results from Experiment 2 were replicated in this study. The same model from Experiment 2 was fit to the two overlap conditions in the Experiment 3 data. The Dominant vs. Secondary overlap factor again significantly predicted selection decision, meaning that participants were more likely to say the Dominant name following a word with secondary name overlap than in response to Dominant name overlap, and vice versa, ($z = -2.95$, $p < .05$). These results replicate the main effect of Experiment 2 (see Table 2 and Figure 4).

To test effects of the overlap conditions relative to the No Overlap condition, a separate model was fit to a categorical predictor with No overlap as the reference condition, comprising a parameter for Dominant vs. No overlap and Secondary vs. No Overlap. Parameters were also fit to trial order, as well as interaction terms with both predictor variables. Proportion of Dominant and Secondary names produced did not differ between Dominant Overlap and No Overlap conditions ($z = -1.71$, $p = .09$; see Table 2). However, naming proportions did differ between Secondary and No Overlap conditions ($z = 2.19$, $p < .05$), such that proportion of dominant names produced was higher under the Secondary Overlap condition compared to the No Overlap condition.

Parameter	Estimate	SE	z	*p*
(Intercept)	1.10	0.19	6.05	< 0.01
Dominant vs Secondary Overlap	-0.85	0.31	-2.95	< 0.01
Dominant vs No Overlap	-0.33	0.19	-1.71	0.09
Secondary vs No Overlap	0.53	0.24	2.19	0.03
Trial	0.38	0.26	1.14	0.26
Overlap * Trial	0.19	0.71	0.27	0.80

Table 2. Model estimates for naming distribution of Dominant and Secondary names under Dominant, Secondary and No Overlap. Results of mixed-effects logistic model (Jaeger, 2008) predicting lexical selection of dominant name (reference group) or secondary name by trial (centered). Effect estimates for Intercept, Trial and Overlap * Trial are from the replication model, Dominant vs Secondary overlap, but effect estimates and significance are similar across both models.

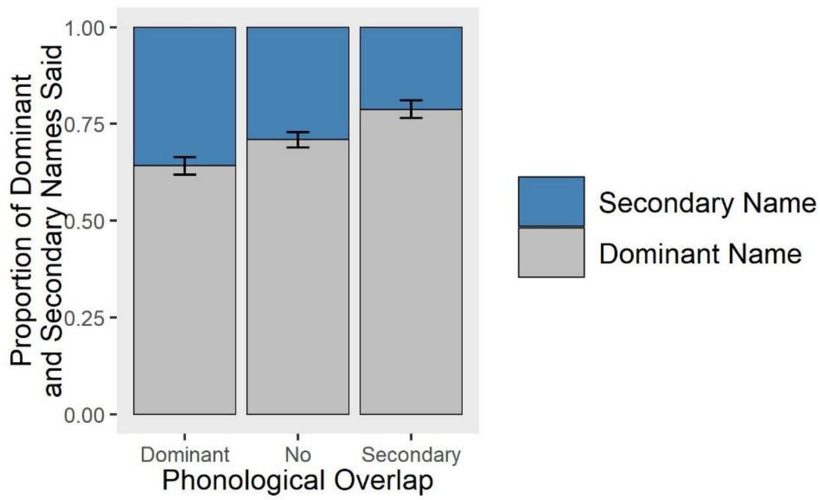

Figure 4. Effect of Overlap on Dominant vs Secondary Names. Proportion of Dominant and Secondary names produced as a function of phonological overlap (Dominant name, No, or Secondary name overlap).

We next investigated the effects of overlap condition and name produced on initiation latencies. Latencies were log transformed before analysis. A mixed effects model predicting naming latency was fit to name said (Dominant vs Secondary), Overlap condition (Dominant vs No, and Secondary vs No), the interaction between Name Said and Overlap condition, and Trial Order (centered). No effects were significant ($p > .05$) except trial order, where participants became faster over the course of the experiment ($t = 4.11$, $p < .01$). Mean log-transformed latencies are presented in Table 3 and Figure 5.

Model	Estimate	SE	t	*p*
(Intercept)	6.99	0.05	128.05	
Name Said (Dom vs Sec)	-0.11	0.05	-2.41	
Overlap (Dom vs No)	0.01	0.02	0.35	
Overlap (Sec vs No)	**0**	**0.03**	**0.06**	
Trial	0.13	0.04	3.2	
Name Said (Dom vs Sec) * Overlap (Dom vs No)	0.12	0.06	2.09	
Name Said (Dom vs Sec) * Overlap (Sec vs No)	0.03	0.06	0.55	

Model	Estimate	SE	t	*p*
Overlap (Dom vs No) * Trial	0.04	0.06	0.73	
Overlap (Sec vs No) * Trial	-0.06	0.06	-0.98	

Table 3. Mixed effects model estimates for picture naming latency (log) of Dominant vs No Overlap and Secondary vs No Overlap, controlling for Dominant and Secondary Name Said, Trial, interaction of Name Said and Phonological Overlap, and interaction of Phonological Overlap and Trial.

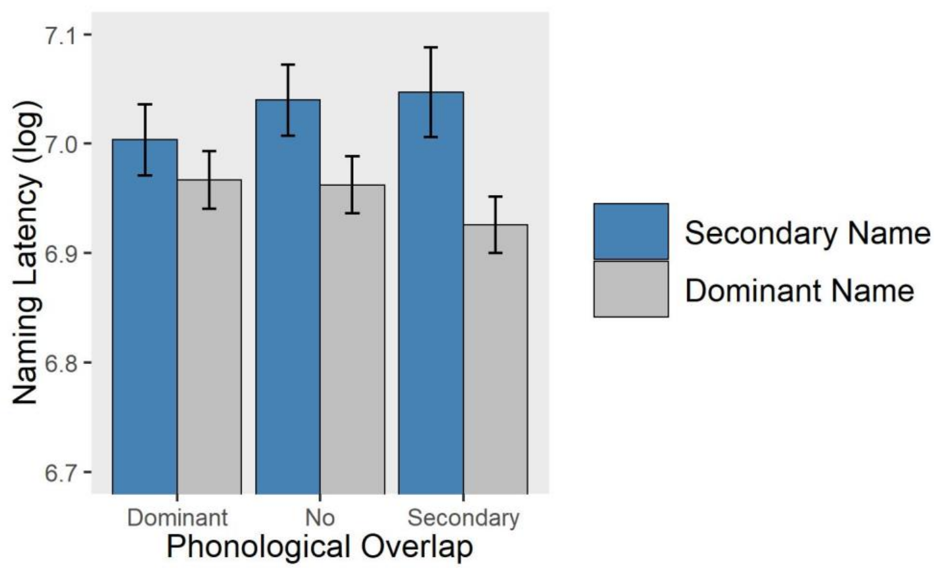

Figure 5. Experiment 3 naming latencies (log transformed) by overlap condition and names produced (bar color). Y axis reflects log transformed naming onset latency with standard error bars.

Discussion

The results of Experiment 3 replicated the naming patterns of Experiment 2. Comparison of the two overlap conditions with the new No Overlap condition showed a reliable that the rate of lexical selection of the dominant name is affected by the activation strength of the phonological form of alternative names. That is, given an identical picture, participants were less likely to say *glass* than *cup* after having said *glare* (Dominant overlap), than after having said *cut* (Secondary overlap).

Dominant naming was overall reliably faster than Secondary naming, however overlap conditions had no reliable effect on naming latency. A lack of effect is somewhat surprising considering latency effects of phonological interference (Fox et al., 2015; Peterson & Savoy, 1998). One possible explanation is that speed of naming may be a stronger constraint in our paradigm, and the viability (and selection) of alternative names is a protective factor.

3.5 General Discussion

In three experiments, we investigated the effect of phonological form on lexical selection, specifically whether the distribution of dominant vs. subordinate names for pictures was affected by the phonological form of a word read aloud just before picture naming. In two experiments, the probability of saying the dominant vs. secondary name for target pictures was influenced by the phonological form of the immediately preceding production trial. Specifically, participants were less likely to produce a name that had phonological overlap with the onset of the preceding word. Similarity in phonological onset interferes with lexical selection, sometimes leading people to produce words they otherwise would not have produced. In the next sections, we discuss several implications and future directions for these results.

As described in the introduction, our method integrates aspects of single word picture naming methodologies, which are common in many studies of lexical selection, with priming methods and an emphasis of choice of utterance form rather than latencies, both of which are common in studies of sentence production. We also aimed to expand the number of items over which we could test our hypotheses compared to some previous investigations of phonological effects on lexical selection (Ferreira & Griffin, 2003; Jaeger et al., 2012b; Rapp & Samuel, 2002). Normative data for our experimental items and other pictures are available in the supplemental materials, and comparisons with previous name agreement assessments (Szekely et al., 2005) are informative, as they provide an estimate of how picture naming patterns may vary with regional dialect or over time. Because it is not trivial to find pictures with low name agreement, the set of items identified from Experiment 1 (see Appendix) may be of use for future studies of alternative naming.

3.5.1 Scope of Phonological Interference

The current studies shed new light on the scope of phonological interference effects in language production. The classic effect of phonological overlap, disfluencies when producing tongue twisters, typically is elicited by partial phonological overlap of adjacent words within an utterance, as in *She sells sea shells.* Word exchange errors also show a phonological influence, as words with phonological similarity are more likely than chance to exchange with each other, as in the error *I wrote a mother to my letter* (Dell & Reich, 1981). Similarly, the phonological interference effects on lexical selection reported by Jaeger et al. (2012b) were also within a single utterance plan of a simple sentence, where phonological overlap between the subject of the sentence and the adjacent verb affected the particular verb produced. By contrast, the phonological interference effects in the current experiments come from two distinct utterance plans: participants read a word aloud, and only after this utterance was completed did they see a target picture to name (see also Bock, 1987). These results suggest that phonological interference between words is not necessarily dependent on a shared plan.

Indeed, in some other studies of phonological form on lexical selection, the priming word was never produced aloud by the participant. For example, in Rapp and Samuel's (2002) studies, the rhyming word prime was presented in a sentence that was read silently or heard, not produced by the participant. And in V Ferreira & Griffin's (2003) study, the homophone prime word was not even presented but merely expected based on a sentence context that participants read silently. Together, these results suggest that the phonological effects are fairly abstract, at the level of phonological encoding, and not from lower-level articulatory codes, which are not expected to be activated for words that are read, heard, or anticipated during reading (Reilly & Blumstein, 2014b).

3.5.2 Interactivity in Lexical Selection

These results are readily accounted for by interactive theories of lexical selection (e.g., Dell, 1986). During spoken production, as activation spreads from conceptual (grammatical) stages of production, multiple alternative names become activated for potential selection. Whereas serial order models argue that selection should be constrained strictly by activation from the conceptual stages (Levelt, 1999), our results suggest that phonological activation from the previously said word also constrains selection. Having produced a word form which overlaps with the form of a candidate word to be selected, decreases that candidate word's activation such that the probability of that word being selected is lower. This pattern is consistent with the effect of homophones on probability of naming errors (Ferreira & Griffin, 2003), and presents a challenge for serial order accounts.

These studies also implicate an ongoing debate as to whether production stages are interactive. Within the scope of single-word production, evidence for interaction in lexical selection has been limited to naming speed and error patterns (Ferreira & Griffin, 2003; Peterson & Savoy, 1998) . Both latency and error effects can be difficult to interpret because theories disagree on whether to interpret changes in latency as arising from pre-lexical vs post-lexical stages of production (Levelt, 1999; Mahon et al., 2007; Oppenheim & Balatsou, 2019).

Conversely, changed naming distributions due to phonological onset overlap provides clear evidence that phonological factors influence lexical selection.

3.5.3 Good-enough production

The various design advantages of these studies lend to the strongest evidence to-date of a "good-enough" constraint in lexical selection. First, because of the number of different target stimuli, the phonological effects observed account for by-item random effects, supporting generalization that the effect is not limited to idiosyncratic properties of our items (e.g., semantic). Rather, item selection was motivated by name agreement observed in Experiment 1. Second, by implementing priming across, rather than within, production trials, we were able to localize an effect of lexical selection. Whereas previously, lexical selection has been uniquely, and importantly constrained to be a meaning-driven process, these results suggest that at least in the limits, alternative names are reliably co-opted to circumvent phonological interference and ease production. An analogous effect in comprehension literature, *Good-enough comprehension*, has been studied for some time. Comprehenders are prone to make errors during online reading or listening, especially when a surprising word appears, revealing the active role of expectations.

This has implications for research that assumes differences in word use (exclusively) reflect underlying differences in meaning for those words. For example, Mahowald, Fedorenko, Piantadosi and Gibson (2013) investigated the production of abbreviated versus full forms (*fridge* vs *refrigerator*) as a function of transition probability from the previous word. They found that abbreviated forms occurred when the target concept was more probable relative to non-abbreviated forms. They concluded that the reliable difference in use was indicative of difference in meaning between forms. Our data and others' (Jaeger et al., 2012b) suggest that phonological overlap is also a reliable cause of differences in naming patterns. For illustration, consider potential noun modifiers *free* and *repaired* for *fridge* vs *refrigerator*. While word co-occurrence may be partially driven by semantics, it is plausible that the phonological onset overlap within the noun-phrases also affect lexical selection distributions. However, it is worth noting that accessibility and meaning constraints need not be mutually exclusive (Bock CITE?).

Future work should investigate the relationship between effects of phonological overlap and message-relevant constraints on lexical selection among alternatives. Evidence for interactivity is largely held to be limited to near-semantic neighbors (Dell & O’Seaghdha, 1991; Goldrick, 2006). While these experiments were underpowered to evaluate the semantic trade-off between alternatives, there is some indication that effects are not limited by degree of semantic similarity. Moderate differences were found among naming alternatives such as *cube* vs *box* and *rose* vs *flower*. Systematic investigation of the types of alternatives and boundary conditions would help inform the extent of interactivity in lexical selection.

3.5.4 Conclusion

In summary, the lexical selection of words is clearly subject to non-message factors. Alternative words are more likely to be selected when the preceding utterance poses a potential phonological overlap with a target word. The robustness of this effect across naming multiple, simple line drawings presents a novel avenue for future research. Finally, the potential for substantial lexical variability, owed to non-message factors, implies that what producers say is not a full reflection of what they mean, illuminating a separation of message and utterance, and predicting potential ambiguity for the comprehender.

4 The Probability of Lexical Selection

4.1.1 Message alignment

A foundational assumption of language use is that the words people say are a close reflection of what they mean. Current descriptions of language production assume that when something is reliably called *coat*, it resembles a message more closely aligned with the meaning “coat” than “jacket”. This assumption underlies the intuition that language successfully communicates meaning, but it is unclear to what extent reliable utterances resemble stable underlying meanings. In this study we investigate the possibility that in addition to meaning-driven activation, selection of words is also a product of differences in the strength of activation of word form representations, themselves. Synthesizing various aspects of language production research, which typically investigate the time course and accuracy of selecting a single, prespecified word, we establish an approach for addressing the extent to which activations for word forms might fluctuate as a function of recent utterances, causing them to occasionally be selected. We then test this question building on the paradigm developed and used in Chapter Three.

Researchers agree that the full scope of language production comprises three ordered stages of processing that each lead to a distinct selection outcome: a message intention, a lexical semantic word, and a lexical phonological form (Goldrick, 2006). Serial models assume that processing of an earlier stage is complete before later stage processing occurs (Levelt, 2001; Levelt et al., 1999). These model assumptions reflect a core assumption of the communicative function of language—the words we say reliably convey our intended meaning.

The meaning-based assumption is a description of the hallmark of language, communicating. There is much empirical work showing supporting the role of meaning in expression. For example, to describe a friend the producer might have the option of *Betty* or *she*

(Arnold, 2008). Producers are more inclined to say a pronoun *she* when the friend has already been introduced in the discourse. Though the producer has *Betty* in mind, a pronoun as the producer's choice of referring expression reflects the needs in that moment.

Meaning-based production also includes what is termed as *producer design* (MacDonald, 2013). In another paradigm (Ferreira & Dell, 2000a), producers are inclined to disambiguate *bat* from *bat* when it is conceptually, but not linguistically ambiguous. Both audience and producer-design emphasize *design;* hypotheses about which decisions are or are not part of the word-selection process. (However, as we will see, producer design is potentially compatible with the view proposed here.) Production proceeds from meaning to word.

4.1.2 Form Accessibility

While meaning drives production choices, form-based effects appear to be relevant, too. Accessibility in language production are cases of decisions in language production that stem from cognitive ease or habit. For example, the order of words for an unplanned sentence is influenced by the phonology of the most recent words recently said. In the classic study of this phenomenon, participants described pictures, and read words aloud, one at a time. On trials when the phonological onset of a preceding word overlapped with the subject of a sentence, participants were more likely to produce the sentence's object first, resulting in a passive sentence, than when the preceding word's phonology overlapped with the object of the sentence. Bock interpreted the overlap to have caused interference during word retrieval. Because word order can be flexible, delayed retrieval of the subject promoted more passive sentences.

However, more recent serial accounts have adapted to accommodate these effects**.** A cascade account is a serial account with the caveat that before selection of an earlier stage, activation from competitors spreads and initiates activation in subsequent stages (Peterson & Savoy, 1998). Cascade models account for mixed error effects, such as *cat*→*rat* > *cat*→*cab*

(Goldrick, 2006). Cascade can account for mixed, feed-forward activation, but not feedback activation.

Evidence for feedback activation looks like phonological properties affecting lexical semantic selection. Earlier work showed a selection effect in sentence contexts (Jaeger et al., 2012b; Rapp & Samuel, 2002). In both study designs participants were asked to produced sentences, where earlier words were carefully constrained to exhibit phonological overlap with a possible later word. Researchers measured the probability of participants producing the phonologically overlapped word vs an alternative word, compared to matched control sentences without phonological overlap. In both studies, phonological overlap predicted word choice. This is consistent with an interactive account but problematic for a cascade account. Jaeger, Furth and Hilliard (2012b) reason that post-selection monitoring is possible. On this view, monitoring would reject the encoded word form with phonological overlap, and then trigger re-selection of an alternative word.

To investigate whether participants could be induced to give alternative names in single-word production, In Chapter Three I elicited production in a simple picture-naming paradigm. They manipulated phonological overlap to an upcoming word such that access for one word was artificially interfered. Participants reliably selected alternative names under manipulation. Further, latencies were, if anything, faster for alternative words relative to overlapped words and respective control condition latencies. This finding shows that word choice is subject to activation of phonological form, even when such activation spills over from a previous, unrelated production.

Both of these cases of phonological activation affecting naming probability are directly predicted from interactive accounts, but they are not incompatible with serial accounts. Lexical representations might comprise learned biases for phonological overlap sequences (Dell et al., 2008), which in turn predictively down-regulate the activation of words whose phonology overlaps with the last word said. More direct evidence for an interactive account of lexical selection would be to demonstrate influence of the activation of words to be selected, on probability of selection, controlling for their relevance to the message to be conveyed, the focus of the present study.

4.1.3 Lexical Priming

Investigations of lexical activation on production are often conducted using a lexical priming paradigm. In the course of naming pictures and or words one at a time, a participant encounters the same word more than once. The central finding is that saying a word they have previously encountered, whether by naming a picture or reading aloud from print (Barry et al., 2001) occurs more quickly than if they have not recently said the word. The interpretation is that after the first utterance of a word, the strength of that word's activation increases, causing it to be more quickly retrieved the next time.

The effect of prior production is long-lasting. For example, Cave (1997) tested the duration of detectable repetition effects by varying the delay between prime and later picture naming. Participants completed an initial picture naming study and were asked to come back for an additional study, between hours and months later. Repeated naming of pictures in the first session was faster even months later compared to non-repetition naming. Suggesting that repetition is a form of implicit learning with long term consequences (Cave, 1997; for review see Cave, 2014).

While the effect of priming on the same word's activation is fairly clear, it is less clear what effect, if any, words have on the production of semantically adjacent words, that is, alternative words. Some insights come from studies on naming latency. Two paradigms, semantic blocking and picture-word interference, employ manipulation of semantic neighbors to test effects on naming latency of target words (See Nozari, 2020). In semantic blocking, participants name pictures one at a time that are semantically related, whether in immediate succession or interspersed among unrelated trials, the effect is increased naming latency (Biegler et al., 2008). For example, naming a picture *cat* on one trial, will slow the retrieval for the word *dog* on a later trial, relative to naming mat. The interpretation is that repeated naming induces increased activation among words that are near-semantic neighbors, which increases processing difficulty and time when selecting the target name.

In picture-word interference paradigms, a target picture to be named is presented, and after a short period (e.g., 400 ms), a printed word appears, overlayed on the picture (Jescheniak et al., 2020; Schriefers et al., 1990). When the content of the printed word is semantically related, naming is slower, relative to when the content is semantically unrelated. As with interpretations of semantic blocking latencies, one view is that semantic activation from the non-target word creates direct interference to producing the target word. Accounts diverge as to whether increased delays reflect competition among semantic neighbors in lexical selection (Aristei et al., 2012), or increased demands in monitoring of message-alignment (Mahon et al., 2007).

However, an open question is whether such changes seen in naming latencies similarly amount to changes in naming probabilities, as in Koranda and MacDonald (2018). Measuring factors that might affect lexical selection requires items that are amenable. In Chapter Three the probability of names for low name agreement pictures were reliably measured. Low-name agreement items are pictures for which a modal name exists, termed the dominant name, as well as additionally common and less common responses, including secondary names. Here we ask whether manipulations to the secondary name can affect the probability of saying the dominant name. For example, *jacket* and *coat* are the dominant and secondary names for a picture. Focusing our manipulation on secondary names increases the likelihood of detecting an effect, while simultaneously reducing the likelihood of detecting the manipulation, because participants are otherwise more likely to give the dominant name.

In order to manipulate lexical form using lexical priming, without subtle effects on form-meaning mappings, it would be beneficial to focus the prime only lexical form. Francis et al. (2008) manipulated aspects of encoding to identify the locus of repetition effects. Participants were primed with either a picture, a printed word or both. Both picture and word priming additively facilitated later naming. They interpreted this result as evidence for facilitation at multiple stages of production, including specifically, articulation of word form. Barry et al. (2006) found a clear, albeit diminished effect of priming from printed word to picture naming. These results suggest that production-independent word form activation plays a role in processing efficiency.

It may be that effects of latency among high name agreement pictures translates to effects of probability among low name agreement pictures. If activation of alternative names directly affects the selection process, then such activation change may shift the distribution of naming. Prior production of words should affect probability of naming pictures for which those words are viable. Alternatively, it may be that latency interference effects do not translate to the decision process of lexical selection.

4.2 Experiment

To test whether lexical selection is biased by prior experience with lexical forms, we implemented a cross-modal lexical priming picture-naming paradigm. Similar to Chapter two, participants were be asked to name pictures and read words aloud, one at a time. In order to test the effect of repetition on naming probability, rather than using high name agreement pictures, for which a target name (and prior repetition) are all but certain, we used low name agreement pictures. Given two viable names for a target picture, if prior naming (priming) of one name increases probability of the same name on later picture-naming trials, this would be evidence for lexical variability in non-adjacent contexts.

4.3 Method

4.3.1 Participants

Native English speakers (n = 99) from the University of Wisconsin-Madison completed the experiment for course credit or pay, whose data were used for subsequent analyses. An additional 5 participants were excluded due to technical difficulties. Data from University of Wisconsin-Madison participants that were previously collected (n = 101) were also used.

4.3.2 Materials

4.3.2.1 Pictures

Black and white line drawings with a wide distribution of name agreement (Snodgrass & Vanderwart, 1980; Szekely et al., 2005) or freely available on the internet (3) were used as target trials (n = 18) and fill trials (n = 82).

Target pictures were selected for having two strong alternative names (dominant and secondary name), as determined in previous norming. Target trials were selected agnostic of the relationship between alternative names, which included abbreviations (*gas/gasoline*), near synonyms (*jacket/coat*) and category level differences (*parrot/bird*) (Szekely et al., 2005; Vitkovitch & Tyrrell, 1995). Plural variants were counted as the same (*brick/bricks*). Across targets, care was taken to minimize repetition of words (e.g. only one target comprised *bird* as a dominant or secondary name). Fill pictures were a mixture of high and low-name agreement pictures.

4.3.2.2 Printed text

Phonological primes.

Each target trial was associated with three prime trials, which comprised a printed word that overlapped in phonology with the target trial's dominant name, secondary name, or neither. Overlap was always with the initial phoneme (onset overlap), and often extended one or two phonemes. For example, the target picture seen in Figure 1, which elicits *jacket* (dominant name) and *coat* (secondary name) was associated with the primes *jackal (Dominant overlap prime), code (Secondary overlap prime),* and *walrus (No overlap prime).*

Repetition primes.

Each target trial was associated with one repetition trial corresponding to a pre-identified alternative name. *A priori* these names were selected categorically as secondary names, in order to maximize likelihood of detecting an effect. Also, by definition secondary names are less likely to be produced given the target, which has the added benefit of decreasing the task demand of repetition priming.

Fill trials.

An equal proportion of words to picture fill trials were selected to be read aloud for fill trials. Words were initially randomly selected from a word generator designed for psycholinguistic research (wordpool), that had relatively high frequency and concreteness, low salience and orthographic complexity, and non-overlap with picture and prime trials. Authors then replaced a few words to remove spurious semantic correlations.

4.3.3 Design

To test whether repetition priming contributes to Good Enough production in novel lexical selection, participants were asked to name pictures and words one at a time aloud, which covertly comprised both repetition and phonological primes. Phonological priming manipulation was developed in Chapter Three. In that word and picture naming experiment, target images were preceded by printed words with phonological overlap to a candidate target word and followed by four fill pictures and words.

For the Secondary name prime Condition, a block of pictures and words to be named was created, including the secondary name for to-be-named target pictures as printed words, interspersed with fill pictures and words. There are two advantages to focusing our priming manipulation exclusively on the secondary names. First, we avoid any potential ceiling effects associated with the higher naming likelihood of dominant names, and second, since dominant names are overall expected more, priming dominant names would have the unfavorable quality of drawing more attention to word repetitions.

4.3.3.1 Word priming block

Secondary Name trials (n = 18) were randomly distributed among 32 picture trials. To increase distance between Secondary text primes and targets, a subsequent block of 14 additional picture trials and 6 text fill trials were presented, for a total of 70 trials in the Repetition block. Relative order of prime words and target trials varied to minimize order effects, however efforts were made to maximize distance between prime and targets. Between 58 and 136 trials separated corresponding text prime and target trials. For example in one counterbalance order the printed word *jacket* (Secondary name) appeared 2^{nd} , and *pail* 18^{th} out of 18 primes presented, and the corresponding picture of *jacket/coat* appeared 3^{rd} and *pail/bucket* 4^{th} out of 18 targets, respectively. Participants in the No Repetition Condition immediately started on the Phonological priming block.

4.3.3.2 Phonological priming block

The phonological priming block previously designed (Koranda and MacDonald, in preparation), comprised trials appearing one at a time, in the order of prime → target → fill trial

(4x). Text and picture fills were randomly ordered throughout. Twelve fixed lists were generated with various target trial orders, and counterbalanced for phonological prime condition, such that each participant, on average received 6 dominant onset overlap trials, 6 secondary overlap trials, and 6 no overlap trials for a total of 108 trials in the phonological priming block.

4.3.4 Procedure

Instructions and pictures were presented with a custom script using python libraries. Computer set up included a microphone. The script recorded audio. Participants were seated in front of the computer and instructed to name pictures as quickly and accurately as possible. Each picture or printed word was presented one at a time in one of sixteen random counterbalanced orders. Trial presentation for the Secondary name prime Condition proceeded from the Repetition Prime Block to the Phonological Prime Block without difference in presentation within or label between blocks.

On each trial, experimenters manually advanced the trial once participants had finished producing a name out loud, at which point a centered fixation cross appeared for 500ms followed automatically by the next trial. Participants were asked to read aloud words or produce names for pictures as quickly and as accurately as possible and could request a short break if needed. After completion of all trials, the experimenter debriefed participants on the nature of the study, and thanked them for their participation.

4.3.5 Coding

Initial transcription was done with an automated software (python package/google?) with a supplied list of likely target words. The full set of transcriptions was inspected and corrected by a coder for accuracy. The coder annotated any idiosyncrasies (e.g. errors) such as hesitations and articles (*um/uh/a/an*), dysfluencies (articulation errors and false starts), delays (more than 3s; excluded later), and corrections (if full word uttered first), or no utterance.

To maximize precision and datapoints for latency, target audio trials were processed using three automated software to detect onset. The number of trials each method extracted naming latency varied, but across all three methods, latency was identified for 1708 trials (97.3%): Chronset (Roux et al., 2017)(trials = 1451), Google Keyword Search (Michaely et al., 2017)(trials = 1412), and FAVE Alignment (Rosenfelder et al., 2015) (trials = 1192). While not as precise as manual coding, averaging latencies across these methods proved reliable enough to detect effects.

4.3.5.1 Exclusion Criteria.

All target utterances occurred above threshold (400 ms). Naming latencies longer than 3s were excluded from analysis (n = 44). Participants produced all repetition and phonological primes.

Valid utterances for analyses included any error-free production (Szekely et al., 2005). Criteria for an invalid trial was: lack of response, disfluency, correction, hesitation. The number of error- based exclusions was 68.

All analyses focused on the proportion of dominant vs secondary utterances. Any other valid utterances were excluded from analysis (n = 188). The pattern of results does not differ when the denominator is instead all valid utterances, for evaluating proportion of dominant vs secondary responses.

Data for the No Repetition condition were previously collected and coded with similar methods described below, except that they did not undergo an initial automated speech-to-text transcription. Rather, additional quality control and correction by the first author. The full description of these methods is reported in Chapter Three and will not be redescribed.

4.4 Results

4.4.1 Prior Utterances on Word Use

In order to test lexical and phonological priming, we fit a generalized linear model to dominant (1) and secondary (0) naming as a function of dominant (.5) and secondary (-.5) phonological onset overlap, and secondary name priming (.5) vs none (-.5), controlling for trial order. Overall, dominant names were said more often than secondary names ($z = 4.27$, $p < .05$), and this was not affected by trial order ($z = 1.1$, $p > .05$).

Our first question was whether saying a potentially related word previously had consequences for the name given to an unrelated production trial. Participants read aloud words that were the second most likely response to a future to-be-named picture. The effect of producing the secondary word earlier in the experiment decreased the likelihood of saying the dominant name in response to target pictures ($z = -4$, $p < .05$). Having said a word before increased the likelihood of that word being selected for an unrelated future communicative goal.

4.4.2 Phonological Onset Overlap

Our second question was whether any such effect would be additive or confounded with the previously identified effect of phonological onset overlap of an immediately preceding word, decreasing the probability of naming a picture similarly. Replicating prior work, phonological overlap predicted the distribution of names said for target pictures, such that there was a higher likelihood of producing dominant names following a prime word sharing onset with the secondary name ($z = 4.31$, $p < .05$). The interaction between lexical and phonological prime was not significant ($z = -0.19$, $p < .05$).

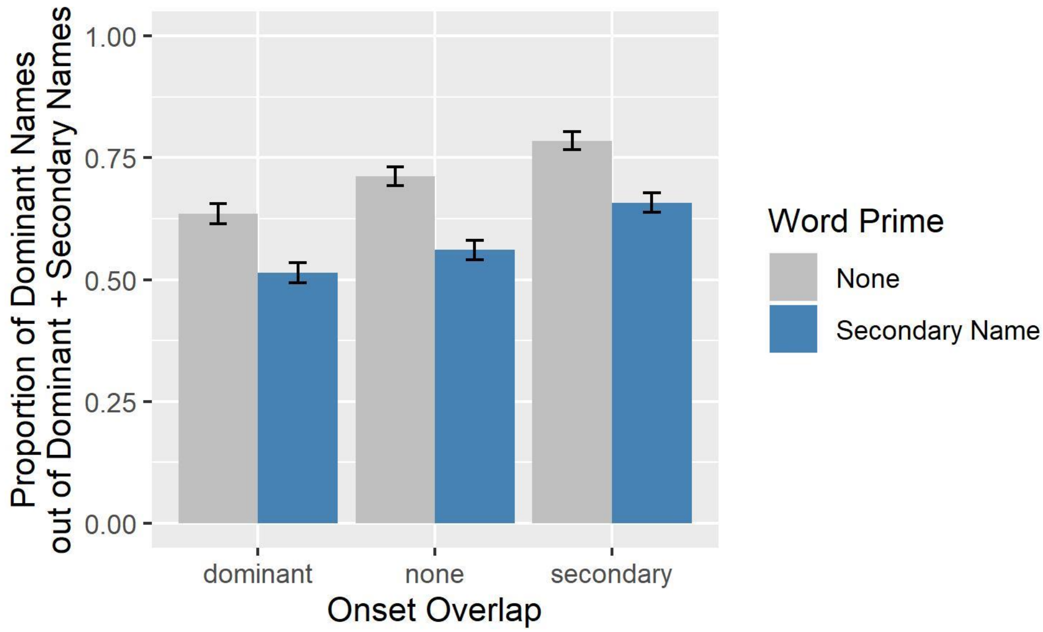

Figure 2. Effect of Repetition and Phonological Prime on Likelihood of Alternative Picture Names. Proportion of dominant names (out of dominant and secondary names) said as a function of phonological overlap (dominant name, none, or secondary name overlap) with the immediately preceding trial condition and repetition condition (none, or secondary name). Groups of bars reflect phonological onset overlap condition, and colors reflect repetition prime condition.

Together these effects suggest that even when communicating about an identical picture, multiple factors from unrelated production plans can (additively) contribute to lexical variability. Rather than communicative goals being a limiting factor for the influence of access, Phonological and Repetition priming are additive.

4.4.2.1 Naming Latency

To capture the effect of lexical priming on naming latency for naming pictures with either dominant or secondary names, data were separated by dominant and secondary names, and fit to a model, respectively. A main effect of latency was found for dominant names, $t = 2.43$, $p < .05$.

Strikingly, having produced the secondary name earlier in the experiment also slowed the onset for saying dominant names by 67ms. Additionally, the phonological overlap effect was replicated ($t = 2.78$, $p < .05$), such that overlap slowed dominant names said for target pictures by 80ms. For dominant names the interaction between phonological and repetition prime did not affect latencies ($t = 0.21$, $p > .05$). Picture naming slowed by about 56ms over the course of the entire experiment ($t = 1.52$, $p > .05$).

For secondary names, the full latency model did not converge, likely because there were fewer observations of producing the secondary name (n = 1028). As such, phonological and repetition primes were tested in separate mixed effects models with trial order, and random by-subject and by-item intercepts. Overall, naming latency for secondary names was 1160ms, ($t = 22.7$, $p < .05$), and slowed by about 122ms over the course of the entire experiment ($t = 3$, $p > .05$). Having produced the secondary name earlier in the experiment sped the onset for saying secondary names by 91ms ($t = 2.59$, $p < .05$), and slowed dominant naming by 67ms ($t = 2.43$, $p < .05$). Secondary naming latencies were numerically slower under secondary (vs dominant) phonological overlap by 45ms, but this was not significant ($t = 1.2$, $p > .05$).

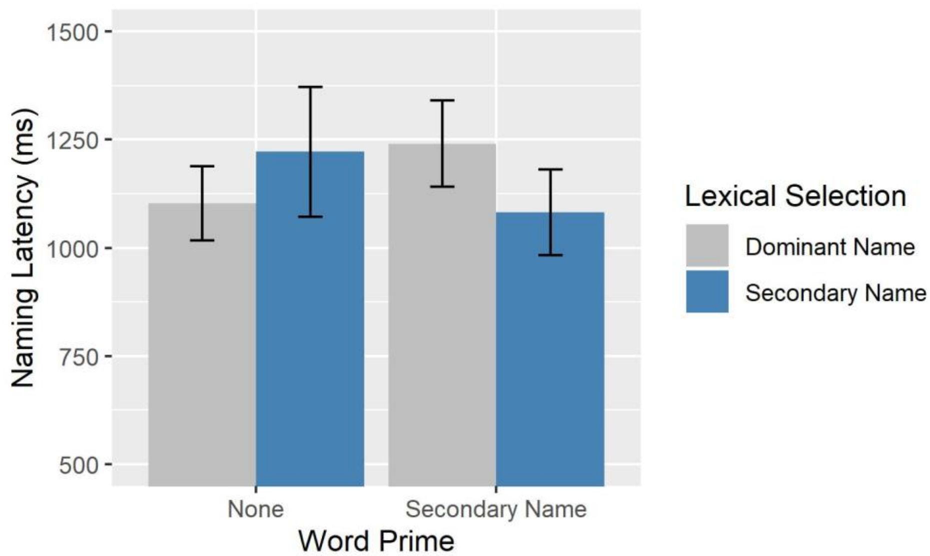

Figure 3. Effect of Lexical Prime on Picture Naming Latency for Dominant and Secondary Names. Naming latency (ms) as a function of word prime (None or Secondary Name) with no phonological overlap on the preceding trial. Bar colors reflect names said (Dominant names in gray and Secondary names in blue).

This study provides several important results. First, we replicate the finding that phonological interference influences lexical selection. Second, a similar effect is found in lexical priming. These further add to the case that lexical selection is variable. Unexpectedly, we also found that participants were slower to produce dominant names for pictures if they had previously produced the secondary name, providing support for theories of competition in lexical selection.

4.5 Discussion

Recent work has demonstrated variability in lexical selection, however, the extent to which factors outside the immediate context might cause lexical variability was unclear. We manipulated lexical priming in a picture naming paradigm. Participants in the lexical priming condition, who had previously read aloud a word, were more likely to select the same word to name a picture with low name agreement, compared to participants who had not previously read aloud that word. This result shows that prior use of words changes the latent activation sufficiently to influence the probability of that word being selected on a later, unrelated production event. Additionally, we replicated the finding that immediate phonological context, namely the phonological onset of a word just said, decreases the likelihood of producing a word of the same phonological onset on the next, unrelated production trial. Finally, having said a word aloud on a prior trial affected the naming latency of later picture naming, whether saying the same, secondary name, or an alternative name, suggesting direct inhibition to alternative name activation, supporting the case for competition in lexical selection.

These results pose a unique challenge to serial models of language production, including those that allow cascading activation from semantic to phonological stages (Peterson & Savoy, 1998). Recent evidence in support of interactive models come from phonological overlap on selection probability (Jaeger et al., 2012b; Rapp & Samuel, 2002). While effects of lexical variability in general are more compatible with interactive accounts of language production, they do not rule out a serial account of production in which phonological representations also entail connection weights that capture phonological transition probabilities in word-adjacent contexts (Dell et al., 2008). Baseline activations of lexical representations might fluctuate over the course of production to facilitate words with more likely phonological properties and inhibit less likely ones. On this view, activation for target picture names which share phonological properties of the immediately preceding phonological prime, are down-regulated (along with other phonologically overlapping words in the lexicon). The lexical priming effect in this study demonstrates non-adjacent effects of phonological influence, which are not straightforwardly predicted by a serial account of language production. By using printed words as primes, early production of those words is unlikely to have systematically altered semantic properties of the dominant and secondary names, and especially not in a way so as to affect the message and message alignment of an upcoming target picture. An interactive account of spreading activation provides a much more straightforward explanation. Increased activation to the primed lexical forms leads to a boost in activation of corresponding lexical-semantic representations, via feedback between representations during spreading activation, and critically, before selection. The relative boost manifests in a probabilistic increase in those lexical names being selected, overall.

Little work in lexical selection has directly evaluated the degree to which lexical selection is subject to naming alternative words. Investigations have focused on the impact of semantic or phonological priming on errors and latencies of dominant names. In almost all studies there are no viable alternative names (e.g., Alario et al., 2004; Barry, Morrison, & Ellis, 1997; Costa, Alario, & Caramazza, 2005; Dell, Oppenheim, & Kittredge, 2008; Meyer & Schriefers, 1991; Nozari & R. Hepner, 2018; Peterson & Savoy, 1998; Vitkovitch, Cooper-Pye, & Ali, 2010). In the few studies that investigate production of low name agreement pictures (e.g., Vitkovitch & Tyrrell, 1995) or the influence of alternative names (Peterson & Savoy, 1998; Rose et al., 2019), effects are evaluated in terms of error patterns and latency strictly with respect to dominant names.

Another interesting result is that our effect of lexical priming and phonological overlap were fully additive, in that the presence or absence of one did not diminish the impact of the other. In our data, there was a difference of about 20 percent in dominant vs secondary names for target pictures. Moreover, the effect was additive with phonological overlap (Chapter Three), such that a difference of about 40 percent in dominant vs secondary names for target pictures was observed. This is consistent with the literature of similar effects on dominant picture naming latencies, where lexical and phonological priming separately and additively influence naming latencies (Perea & Rosa, 2000). Future work should investigate the extent to which both effects on lexical selection apply to various production circumstances (i.e., share boundary conditions).

These results also implicate a related longstanding debate about the timing and mechanism of co-activation during lexical selection; a debate encumbered by methodological limitations (Spalek et al., 2013). Lexical selection in a non-competitive model proceeds strictly based on the more relevant, strongly activated word (Mahon et al., 2007). Considering only our naming probability data, a non-competitive account would be that activation for secondary names was sufficiently primed to be faster than dominant names, for some participants. The crux of the argument, however, lies in how alternative name activations affect the selected name's latency. In a non-competitive account, regardless of the activation of an alternative word (to the one selected), the selected words' latency should be unaffected. Our data are problematic for this account. Dominant names given for target pictures were relatively slower to be selected among participants who had previously said the secondary name aloud vs participants who had not. This suggests the contrary view, that selection is competitive, and that activation of an alternative name, even when that name is not selected, affect the latency of the selected name. More specifically, the activation of near-semantic neighbors can directly interfere with each other.

A key advantage of this study in being able to address the competition debate, pertains its methodological innovation (Spalek et al., 2013). Previous studies of competition have restricted their investigations to the latencies of naming high name agreement pictures, using tasks with unclear interpretations, such as picture-word interference (Mahon et al., 2007; Schriefers et al., 1990) and semantic blocking paradigms (Belke, 2017; Bloem et al., 2004). While not new to psycholinguistic investigations, the present study implemented a novel application of lexical priming to the less-commonly studied outcome of naming probabilities. Such a combination is similar to work in Chapter Three, and cross-trial manipulations such as in Bock (1986).

In Chapter Two I showed that when participants learn an artificial vocabulary with mixed experience across the lexicon, they reliably produce high frequency words in lieu of known, more accurate but less frequent words. These results suggest that lexical selection in novel contexts can be biased by accumulated experience of prior use, in the context of limited vocabulary and controlled ambiguity. It is another question whether such effects hold in natural language production.

In conclusion, this results of this work demonstrate that lexical variability is not limited to the influences of the immediate context, but that activation changes to forms of somewhat recently said words will affect naming latencies and probabilities, providing strong support for an interactive, competitive account of lexical selection. When communicating, the words people say are not just a consequence of communicative intentions and word meanings, but also those words' forms.

5 Lexical variability

5.1 Summary of Findings

Existing theories of lexical selection predict relatively little if any effects of lexical variability, the possibility that producers systematically say alternative words owed to non-message factors. Instead, research in lexical selection has long studied the "semantically and syntactically appropriate" selection of a word (Levelt, 1999, p. 223). These studies sought to investigate lexical variability under fixed message constraints, and variable experiences.

In Chapter Two, under controlled conditions of an artificial language game, high frequency words less aligned with the message were produced over more-aligned, low frequency words. These results held whether participants received direct feedback on their word choices or not (Experiments 3 and 4). Experiment 4 controlled for the frequency of non-linguistic experience.

In Chapter Three, I showed that phonological overlap affects the probability of naming pictures with the dominant vs secondary names. Experiment 3 showed the effect was bi-directional, with both decreased and increased probability of target word production for both dominant and secondary names.

In Chapter Four, I showed that reading a printed word aloud increased the likelihood of naming a later target picture with that word, despite that the word was a less-common response (secondary name). Latencies for dominant names given were slower, and faster for secondary names. Additionally, the effect of lexical activation was additive with the effect of phonological interference on lexical selection.

Together these results provide strong support for lexical variability in language production. Serial models and even proponents of interactive models of language production have argued that bottom-up effects such as phonology are real, but limited (Dell & O'Seaghdha, 1992; Jaeger et al., 2012). Across three experiments in Chapters Two and Three, we found three medium-sized effects of phonological accessibility on lexical selection. In Chapter Four, phonological and lexical accessibility combined led to a difference of 40 percent in dominant and secondary naming of target pictures—what appears to be a large effect. This potentially broadens the relevance of interactivity in accounting for production utterances.

The picture naming paradigm allows a methodological gain in a few important ways. One is to shift the unit of analysis from a *word* and its naming latency to *words* and their naming distribution, given a picture as a message. The immediate implication is a novel, descriptive discovery—lexical variability given a fixed message.

These results take predictions of a two-stage, interactive production model predominantly describing error and latency data in target word production (Dell, 1986), and apply them to word choice, a measurement previously limited to explanations of communication goals (Arnold, 2008). Previous work in lexical selection investigated biases that implicate speed of production and rate of errors (Fenk-Oczlon, 1989; Jescheniak & Levelt, 1994). Studies in Chapter Two demonstrate how probabilistic, interactive effects emerge from variably accessible word forms with an artificial language model. The modeled effects generalize to natural language production, where phonological activation (Chapter Three), and activation of the word form (Chapter Four) affect selection probability. These results point to multiple new avenues of interactivity in lexical selection.

5.2 Limitations

Throughout this dissertation for purposes of contrasting hypotheses, lexical variability is characterized by so-called *non-message* factors. This is a stylistic decision to contrast the prevailing, often unstated assumption to assume a close relationship between word use and message factors. However, on close inspection, classifying an effect as message or non-message falls apart, and more importantly, misses the point. A potentially more durable description for the non-message effects presented here—word frequency, phonological overlap and lexical activation—is that they are factors present *independent* of specific messages, whose implications arise only as a consequence of particular properties of a specific message being communicated, in other words, they are message-*agnostic*, in that the forms, and not the messages underlying the source of influence are the basis of priming.

In that the mainstay of language production research posits a close relationship between lexical variability and message variability, alternative lines of reasoning are available to attribute contextual variation without appealing to variable mapping between messages and words. Consider the effect of phonological interference, given the possibility of two phonologically similar, adjacently produced words. Dell et al. (2008) argue that network activation of phonology, and related word forms is responsive to phonological co-occurrence probabilities, and predictively adjust pre-activation for upcoming words. Such an account puts variability on a factor outside the message-to-word mapping, phonology. However, it minimally requires a strongly interactive production model, in order for phonological effects to penetrate to lexical selection.

A similar line of reasoning can be applied to some effects of accessibility on lexical variability. Uniform Information Density (Mahowald et al., 2013) is one such theory that argues a relationship between cognitive constraints particular to the unfolding message and lexical selection between two near-synonyms. When the upcoming word is one the listener is less likely to predict, a longer word form among viable alternatives will tend to be selected. This line of reasoning attempts to map distributional word usage more finely to equally fine-grained, context-specific messages. Beyond optimizing distributional data, is difficult to hold this view to the choices at hand for the producer. Given *code,* calling the subsequent picture a *jacket* would be deemed more communicative than calling it a *coat.* Distributionally, participants say both variants, and appear to be idiosyncratically communicative across messages, because pictures varied in their baseline name distributions and degree of lexical variability.

The present research did not investigate the communicative consequences of lexical variability. This is to say, while differences between alternatives were observable, it is unclear that pragmatic consequences, namely, comprehension of a message, were affected. It could be that lexical variability emerges only in those cases where miscomprehension given an alternative word is unlikely. However, there are reasons why this is unlikely. First, this approach requires strong support for rational communication, essentially, that all production choices, including disfluencies and selections between synonyms are purposeful. For illustration, consider the view described in the discussion of Chapter Two, in which, high frequency words over repeated selection over more message-aligned alternatives, become used (and understood to) encompass those meanings. Considering the compass data, this is analogous to the Elves understanding that, given *nooba* is a high-frequency word, it is more vague representing a wider span of possible directions for the gold, than a low-frequency word. Critically, individual instances of production are winner-take-all, and not a probabilistic blend between lexical alternatives. A rational account of comprehending lexical variability, at best, can only account for word choices at the level of distributions. Its predictions for word choices fall directly from the semantic context, and are otherwise static, invariable within and across individuals. A rational approach to language could formulate an account for word frequency and phonological overlap effects because these effects draw from broad, distributional properties of words, that are going to be relatively stable across individuals. However, the effect of lexical variability from priming secondary names is especially difficult for a rational account to explain, because it stems from the particulars of an individual's experience, and crucially, particulars that could easily fall outside the immediate discourse context. In the same way that priming was evident despite naming a picture for the first time, it is reasonable to speculate that self-priming could occur across conversations where the listener changes, such as a professor teaching two classes in a row, or a person making two phone calls in a row.

Moreover, ambiguities compound for the comprehender when considering that a word is not only an alternative for one other word, but an alternative for many potential words, and thus meanings. In the case where a high frequency word such as *noobda* is adjacent to two low-frequency words, the comprehender must account for the possibility that *noobda* is aligned with a message closer to either of its semantic boundaries. There is evidence that the process of lexical selection is not equally sensitive to comprehension ambiguities (Ferreira et al., 2005). Producers are less likely to provide disambiguation between two homophonic meanings of *bat* (i.e., baseball bat vs type of mammal), than between two sizes of the same type of bat. This shows that production is not perfectly motivated to combat comprehension ambiguities, and predicts that, in principal, cases of lexical variability can also straddle and mislead comprehension ambiguities.

Finally, it is worth mentioning that these studies (by design) do not consider interactive constraints in multi-word production. Effects of accessibility in sentence production are highly overlapping with precisely this objective. For instance, as discussed in Chapter Four, phonological overlap influences ordering in short phrases (Janssen & Caramazza, 2009) and word choice in sentences (Jaeger et al., 2012b; Rapp & Samuel, 2002). However, unlike the present studies, studies of accessibility design their stimuli such that viable alternative utterances are semantically matched (including those mentioned in this paragraph; see also footnote on p. 7). A speculative lesson from the present studies is that semantically matching alternatives may be unnecessarily restrictive. The stimuli and results across the studies here either directly demonstrated the viability of alternatives that differed in message alignment as seen in Chapter Two, or were selected agnostic of semantic properties (Chapter Three and Four).

5.3 Future Directions

Relaxing the restriction of semantically matched alternatives is, itself, a valuable line of future research. Chapter Three and Four empirically demonstrate the reciprocity of lexical alternatives given a message. Additional studies are needed to characterize what licenses such reciprocity between two words (viability). Viability need not be restricted to semantic neighbors. Consider that in Chapters Two and Three, the criteria for stimuli selection was simply messages with viable alternative names, agnostic of the basis of relation between those names. Some name pairs were near-synonyms, some were taxonomically related and others were associatively related. It is unclear whether these factors differently constrain lexical variability. Frequency itself (Chapter Two) may be a source of viability of alternative names, as might other effects of accessibility.

A reasonable assumption is that lexical variability is constrained by certain message factors. Once baseline effects of variability are better understood, deviations in effect sizes may potentially provide insight about constraints to message factors. In this sense, lexical variability can be used as a tool to investigate the relative degree of meaning-based constraints on production. For example, the stimuli from Figure 1 in the introduction, WOLF and DOG, were also stimuli in Experiment 1, where baseline naming distributions were collected. The word *dog* was an uncommon response to WOLF and the overwhelming response to DOG. It is an open question, however, the extent to which naming distributions for each picture would change, under phonological or lexical priming. One prediction is that the size of changed name distributions is proportionate to the degree of distribution in baseline naming. Since baseline name agreement for WOLF is greater than that of DOG, the degree of lexical variability—percentage increase and decrease of *dog* utterances across priming conditions—would be greater for WOLF than DOG. In the context of comparing distributions for picture naming, this observation may be trivial. However, most of language production is not accompanied by a picture description of the message. Baselines for low and high lexical variability can then be used to estimate the relative constraints from message factors in instances where a distribution of utterances can be observed, but not their underlying messages, such as corpus data. Such an approach, while highly speculative, offers a means for characterizing messages in their pre-verbal state, something that we know little about (Konopka & Brown-Schmidt, 2014), but assume so much.

5.4 Conclusion

These studies are an investigation of the effect of various cognitive dimensions on language use. They characterize some key reasons why when we produce words, despite knowing more accurate words, we might say less accurate alternatives. It will happen more often when, compared to the accurate word, the alternative has been used more frequently, more recently, and when its phonological form differs from the last word said. These effects demonstrate lexical variability. It lends support to the intuition that people do not always say the best words for what they mean. Though oriented to communicate, language production is a process of systematic inaccuracies in word use, inaccuracies that are sure to be found throughout this dissertation and other artifacts of human language.

6 References

Abrams, L., White, K. K., & Eitel, S. L. (2003). Isolating phonological components that increase tip-of-the-tongue resolution. *Memory & Cognition*, *31*(8), 1153–1162. https://doi.org/10.3758/BF03195798

Acheson, D. J., & MacDonald, M. C. (2009). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language*, *60*(3), 329–350. https://doi.org/10.1016/j.jml.2008.12.002.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155.

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791–810. https://doi.org/10.1037/0278-7393.33.4.791

Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*(1), 143–148. https://doi.org/10.1016/j.cognition.2010.04.001

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527. https://doi.org/10.1080/01690960801920099

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554. https://doi.org/10.1080/01690960802299378

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

Barry, C., & Seymour, P. H. K. (1988). Lexical Priming and Sound-to-Spelling Contingency Effects in Nonword Spelling. *The Quarterly Journal of Experimental Psychology Section A*, *40*(1), 5–40. https://doi.org/10.1080/14640748808402280

Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of Acquisition, Word Frequency, and the Locus of Repetition Priming of Picture Naming. *Journal of Memory and Language*, *44*(3), 350–375. https://doi.org/10.1006/jmla.2000.2743

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 560–585. https://doi.org/10.1080/027249897392026

Baus, C., Strijkers, K., & Costa, A. (2013). When does word frequency influence written production? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00963

Belke, E. (2017). The Role of Task-Specific Response Strategies in Blocked-Cyclic Naming. *Frontiers in Psychology*, *07*. https://doi.org/10.3389/fpsyg.2016.01955

Biegler, K. A., Crowther, J. E., & Martin, R. C. (2008). Consequences of an inhibition deficit for word production and comprehension: Evidence from the semantic blocking paradigm. *Cognitive Neuropsychology*, *25*(4), 493–527.

Bloem, I., van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, *51*(2), 307–323. https://doi.org/10.1016/j.jml.2004.05.001

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*(1), 1–47.

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, *26*(2), 119–137. https://doi.org/10.1016/0749-596X(87)90120-3

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 467–484. https://doi.org/10.1016/S0022-5371(80)90321-7

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67. https://doi.org/10.1016/0010-0277(85)90023-X

Brennan, S. E. (1990). Conversation as direct manipulation: An iconoclastic view. In B. K. Laurel (Ed.), *The art of human-computer interface design* (pp. 393–404). Addison-Wesley.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430–1450. https://doi.org/10.1037//0278-7393.27.6.1430

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*(2), 105–135. https://doi.org/10.1016/j.cognition.2004.10.006

Collins, A., & Loftus, E. (1975). A Spreading-Activation Theory of Semantic Processing. *A History of Psychology in Autobiography, Vol. IX.*, *82*(6). https://doi.org/10.1037/11571-006

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological Review*, *98*(4), 604–614. http://dx.doi.org.ezproxy.library.wisc.edu/10.1037/0033-295X.98.4.604

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314. https://doi.org/10.1016/0010-0277(92)90046-K

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*(4), 583–608.

Engelhardt, P., Bailey, K., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573. https://doi.org/10.1016/j.jml.2005.12.009

E-Prime 2.0. (2012). Psychology Software Tools, Inc. https://www.pstnet.com.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340.

Ferreira, V. S., & Dell, G. S. (2000). Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology*, *40*(4), 296–340. https://doi.org/10.1006/cogp.1999.0730

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340.

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis) selection. *Psychological Science*, *14*(1), 86–90.

Ferreira, V. S., Morgan, A., & Slevc, L. R. (2018). Grammatical Encoding. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics*. https://doi.org/10.1093/oxfordhb/9780198786825.013.18

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*(Supplement C), 97–117. https://doi.org/10.1016/j.jml.2015.04.002

Friendly, M. (2018). *WordPools: Classical Word Pools Used in Studies of Learning and Memory*. https://CRAN.R-project.org/package=WordPools

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*(2), 141–176. https://doi.org/10.1016/j.cogpsych.2012.03.002

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*(7–8), 817–855. https://doi.org/10.1080/01690960600824112

Harley, T. A., & MacAndrew, S. B. (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research*, *30*(4), 395–418.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22–44. https://doi.org/10.1016/j.cogpsych.2017.08.002

Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: Evidence from subject-verb agreement. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(5), 891–904. https://doi.org/10.1037/0278-7393.31.5.891

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

Jaeger, T. F., & Levy, R. P. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems* (pp. 849–856).

Jaeger, T. F., Furth, K., & Hilliard, C. (2012a). Incremental Phonological Encoding during Unscripted Sentence Production. *Frontiers in Psychology*, *3*, 481. https://doi.org/10.3389/fpsyg.2012.00481

Jaeger, T. F., Furth, K., & Hilliard, C. (2012b). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1439–1449.

Janssen, N., & Caramazza, A. (2009). Grammatical and Phonological Influences on Word Order. *Psychological Science*, *20*(10), 1262–1268. https://doi.org/10.1111/j.1467-9280.2009.02429.x

Jescheniak, J. D., Wöhner, S., Bethcke, H. S., & Beaupain, M. C. (2020). Semantic interference in the picture-word interference task: Is there a pre-lexical, conceptual contribution to the effect? *Psychonomic Bulletin & Review*, *27*(2), 373–378. https://doi.org/10.3758/s13423-019-01667-w

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*(1), 113–139. http://dx.doi.org.ezproxy.library.wisc.edu/10.1037/0033-2909.120.1.113

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54.

Kapatsinski, V. (2010). What is it I am writing? Lexical frequency effects in spelling Russian prefixes: Uncertainty and competition in an apparently regular system. *Corpus Linguistics and Linguistic Theory*, *6*(2). https://doi.org/10.1515/cllt.2010.007

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. *The Oxford Handbook of Language Production*, 3–20.

Koranda, M. J., & MacDonald, M. C. (2018). The breadth of phonological interference in word choice. Presented at CUNY Sentence Processing, Davis, CA.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Labov, W. (2011). *Principles of linguistic change, volume 3: Cognitive and cultural factors* (Vol. 36). John Wiley & Sons.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1), 1–22.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38.

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122.

Levelt, W., & Maasen, B. (1981). Lexical Search and Order of Mention in Sentence Production. In W. Klein & W. Levelt (Eds.), *Crossing the Boundaries in Linguistics: Studies Presented to Manfred Bierwisch* (pp. 221–252). Springer Netherlands. https://doi.org/10.1007/978-94-009-8453-0_12

Lorimor, H., Adams, N. C., & Middleton, E. L. (2018). Agreement With Conjoined NPs Reflects Language Experience. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00489

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3636467/

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3636467/

MacKay, D. G. (1982). The Problems of Flexibility, Fluency, and Speed-Accuracy Trade-Off in Skilled Behavior. *Psychological Review*, *89*(5), 483-507.

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*(3), 503.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

Michaely, A. H., Zhang, X., Simko, G., Parada, C., & Aleksic, P. (2017). Keyword spotting for Google assistant using contextual speech recognition. *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 272–278. https://doi.org/10.1109/ASRU.2017.8268946

Mirković, J., & MacDonald, M. C. (2013). When singular and plural are both grammatical: Semantic and morphophonological effects in agreement. *Journal of Memory and Language*, *69*(3), 277–298.

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morgan, A. M., von der Malsburg, T., Ferreira, V. S., & Wittenberg, E. (2020). Shared syntax between comprehension and production: Multi-paradigm evidence that resumptive pronouns hinder comprehension. *Cognition*, *205*, 104417. https://doi.org/10.1016/j.cognition.2020.104417

Nozari, N., & R. Hepner, C. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, 1–15. https://doi.org/10.1080/02643294.2018.1476335

Oppenheim, G. M., & Balatsou, E. (2019). Lexical competition on demand. *Cognitive Neuropsychology*, *36*(5–6), 216–219. https://doi.org/10.1080/02643294.2019.1580189

Perea, M., & Rosa, E. (2000). Repetition and form priming interact with neighborhood density at a brief stimulus onset asynchrony. *Psychonomic Bulletin & Review*, *7*(4), 668–677. https://doi.org/10.3758/BF03213005

Perret, C., & Bonin, P. (2018). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1100-1

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 539.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse Constraints on Syntactic Processing in Language Production: A Cross-Linguistic Study in English and Spanish. *Journal of Memory and Language*, *42*(2), 168–182. https://doi.org/10.1006/jmla.1999.2668

Rapp, D. N., & Samuel, A. G. (2002). A reason to rhyme: Phonological and semantic influences on lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 564–571. http://dx.doi.org.ezproxy.library.wisc.edu/10.1037/0278-7393.28.3.564

Reilly, M., & Blumstein, S. E. (2014). Effect of sound similarity and word position on lexical selection. *Language, Cognition and Neuroscience*, *29*(10), 1325–1341. https://doi.org/10.1080/23273798.2014.917193

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 753–763. https://doi.org/10.1037/xlm0000592

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2015). *FAVE: Speaker fix* (v1.2.2) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.22281

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods*, *49*(5), 1864–1881.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86–102.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91–127. https://doi.org/10.1016/0010-0277(94)90067-1

Slobin, D. (1987). Thinking for Speaking. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 435–445.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174.

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes*, *28*(5), 597–614. https://doi.org/10.1080/01690965.2012.718088

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes*, *28*(5), 597–614. https://doi.org/10.1080/01690965.2012.718088

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Anal'a, L. A., Vargha, A., & Bates, E. (2005). Timed action and object naming. *Cortex*, *41*(1), 7–25.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology*, *48*(4), 822–848.

Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, *22*(6), 996–1003. https://doi.org/10.1016/j.conb.2012.05.003

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*(3), 452–457. https://doi.org/10.3758/BF03196594

7 Appendix

7.1 Stimuli

Table 5. Picture Name Agreements and Overlap Words.

Target pictures with normed name agreement (Experiment 1), and associated manipulations (Overlap) in Experiments 2 and 3. Picture numbers correspond to numbered pictures in Figure 6 below. Overlap names for Experiment 2 and 3 correspond to Dominant and Secondary names identified in Experiment 1.

Picture	Name	Exp 1 Agreement	Exp 2 Overlap	Exp 3 Overlap
1	hat	dominant	half	half
1	cap	secondary	courage	courage
1		none		desk
2	coat	dominant	code	code
2	jacket	secondary	jackal	jackal
2		none		walrus
3	pillar	dominant	pickle	pickle
3	column	secondary	caution	caution
3		none		shoe
4	cube	dominant	cute	cute
4	box	secondary	bucks	bluff
4		none		toffee
5	desert	dominant	dresses	destiny
5	cactus	secondary	kernel	castle
5		none		grape
6	frog	dominant	frost	frock
6	toad	secondary	tone	toe
6		none		vapor
7	garbage	dominant	gas	gas
7	trash	secondary	NA	trial
7		none		airplane
8	glass	dominant	glare	glare
8	cup	secondary	cast	cut
8		none		apron
9	mixer	dominant	mitt	minty
9	whisk	secondary	NA	NA
9		none		corn

Picture	Name	Exp 1 Agreement	Exp 2 Overlap	Exp 3 Overlap
10	bucket	dominant	back	buckle
10	pail	secondary	pain	paint
10		none		cucumber
11	parrot	dominant	pain	parents
11	bird	secondary	burn	burn
11		none		hope
12	rose	dominant	resist	resist
12	flower	secondary	flounder	flounder
12		none		scone
13	sailboat	dominant	savvy	savers
13	boat	secondary	bear	bones
13		none		folder
14	scale	dominant	scare	scare
14	balance	secondary	ballots	ballots
14		none		muffler
15	teeth	dominant	tree	teams
15	dentures	secondary	dances	dense
15		none		granite
16	wolf	dominant	window	window
16	dog	secondary	NA	dollop
16		none		genius
17	rocket	dominant	rod	rod
17	missile	secondary	measure	misty
17		none		juice
18	road	dominant	ribs	ribs
18	highway	secondary	hero	hero
18		none		bottle
19	bird	dominant	burglar	
19	pigeon	secondary	piston	
20	cup	dominant	cut	
20	mug	secondary	madness	

Figure 6. Target pictures for Experiments 2 (all items) and 3 (1-18 only).

7.2 Supplementary Tables

Model	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.840	0.226	3.716	0.000
Overlap (Dom vs No)	-0.335	0.187	-1.788	0.074
Trial	0.449	0.240	1.873	0.061

Table 6. Model estimates for naming distribution of Dominant and Secondary names under Dominant and No overlap.

Model	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.317	0.176	7.487	0.000
Overlap (Sec vs No)	0.603	0.265	2.276	0.023
Trial	0.448	0.257	1.742	0.082

Table 7. Model estimates for naming distribution of Dominant and Secondary names under Secondary and No overlap.

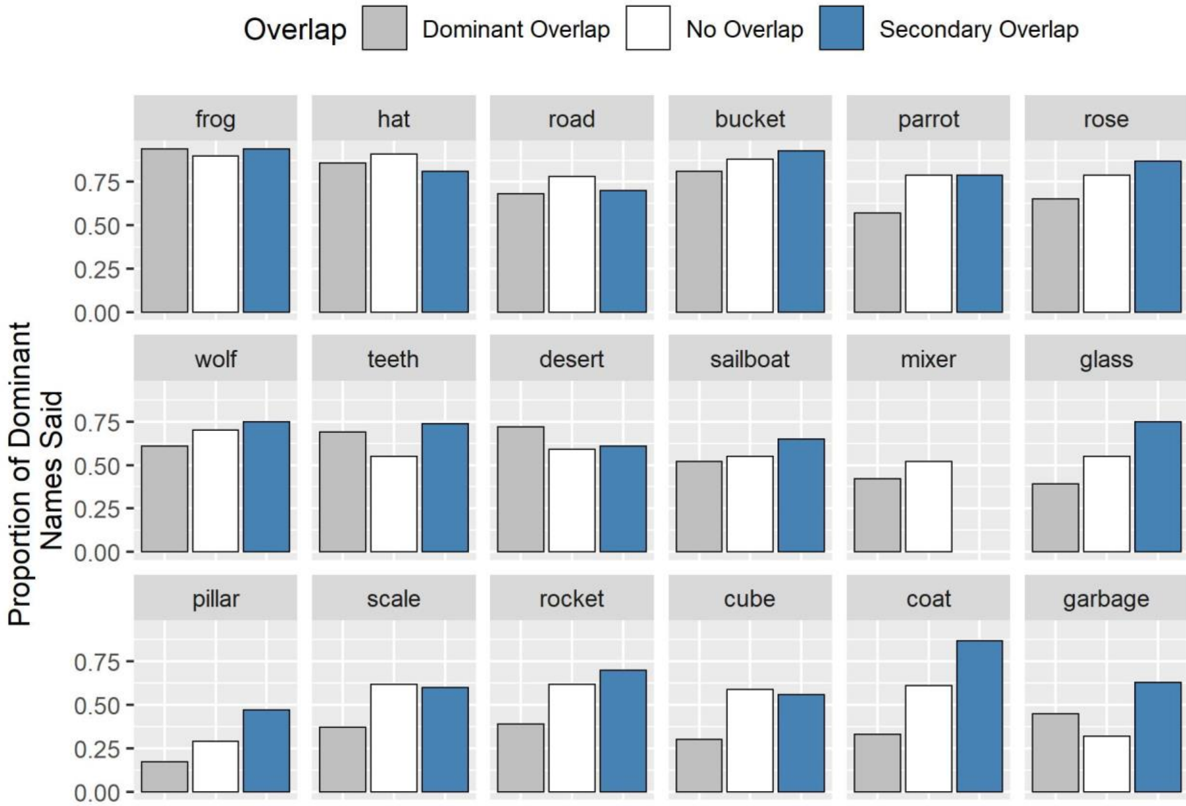

Figure 7. Probability of naming, by target pictures (Experiment 3). Panel names correspond to dominant names in Table 5, ordered by most to least baseline name agreement (Experiment 1). Bars correspond to Overlap condition (gray for Dominant, white for No, and blue for Secondary Overlap).