

Good-Enough Production: Selecting Easier Words Instead of More Accurate Ones



Mark J. Koranda¹, Martin Zettersten^{1,2},
and Maryellen C. MacDonald¹

¹Department of Psychology, University of Wisconsin–Madison, and ²Department of Psychology,
Princeton University

Psychological Science

1–12

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976221089603

www.psychologicalscience.org/PS



Abstract

Dominant theories of language production suggest that word choice—lexical selection—is driven by alignment with the intended message: To talk about a young feline, we choose the most aligned word, *kitten*. Another factor that could shape lexical selection is word accessibility, or how easy it is to produce a given word (e.g., *cat* is more accessible than *kitten*). To test whether producers are also influenced by word accessibility, we designed an artificial lexicon containing high- and low-frequency words whose meanings correspond to compass directions. Participants in a communication game (total $N = 181$ adults) earned points by producing compass directions, which often required an implicit decision between a high- and low-frequency word. A trade-off was observed across four experiments; specifically, high-frequency words were produced even when less aligned with messages. These results suggest that implicit decisions between words are impacted by accessibility. Of all the times that people have produced *cat*, sometimes they likely meant *kitten*.

Keywords

language production, lexical selection, artificial-language game, word accessibility, word frequency, open data, open materials

Received 12/17/20; Revision accepted 2/4/22

Options for a behavior, such as which word to say or which hand to reach with, require a decision process. Recent theories of motor control have hypothesized probabilistic decision-making processes that maximize utility of actions in the face of uncertainty (e.g., Wolpert & Landy, 2012). Language production—whether signed, spoken, or written—is a form of action that offers abundant alternative behaviors to convey a message, including alternative words and phrases. Measuring the costs and benefits between alternatives has been difficult because researchers must know the utility of alternative choices in conveying a producer's intended message, which is not readily available.

Theories of word choice (lexical selection) have not favored the probabilistic approach seen in motor control research. In lexical selection accounts, early grammatical encoding processes settle on words to fit the message, and later processes develop the phonological code for overt production (Levelt et al., 1999). On this view, a speaker's word choices, such as *cat* versus

kitten, are guided solely by which words best fit the intended message (i.e., “message alignment”), not by word “accessibility” factors such as a word’s frequency or length, which could affect ease of production. Probabilistic and interactive production approaches have been invoked to account for related phenomena, including findings that accessibility affects word order (Bock, 1982), speed of production (Sevald & Dell, 1994), and distributions of speech errors (Dell & Reich, 1981). However, within the domain of lexical selection itself, the dominant view is that the producer’s message drives word choice (Jescheniak & Levelt, 1994; cf. Dell, 1986). In this research, we investigated whether lexical selection is determined by message alignment or is more probabilistic, as seen in the motor literature. The

Corresponding Author:

Mark J. Koranda, University of Wisconsin–Madison, Department of Psychology
Email: mjkoranda@wisc.edu

results 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 have investigated whether lexical selection is truly controlled by a single deterministic factor. Ferreira and Griffin (2003) examined errors in picture naming and found that speakers misnamed pictures more often when the phonological form of an incorrect competitor word had been primed than when it had not been primed. Ferreira and Griffin termed this result “good-enough production,” meaning that beyond message alignment, the accessibility of phonological forms also affects lexical selection. Similarly, producers increasingly avoid difficult words more when describing pictures under phonological interference (Jaeger et al., 2012; Koranda & MacDonald, 2018; Rapp & Samuel, 2002), and learners will extend familiar morphemes to novel grammatical categories (Harmon & Kapatsinski, 2017). Together, these studies are consistent with an interactive process account of lexical selection (Dell, 1986), in which accessibility constrains producers’ word choices.

A limitation of these studies is that they have no independent assessment of message alignment, making it difficult to distinguish alignment and accessibility effects on lexical selection. There are reasons to believe that message alignment, for example, between a stimulus picture and a word describing it, could vary across producers and contexts. Dialects and other variations in experience affect word use (Melinger, 2021), but they may also affect producers’ messages. Similarly, studies of word usage in different discourse contexts, for example, saying *tummy* or *stomach* to different audiences (Stoll et al., 2009), could have either an accessibility or a message-alignment interpretation. More generally, highly accessible words might be chosen because they both fit the message and have been used frequently (Bock, 1982). Because of difficulties distinguishing alternative accounts, studies promoting good-enough production have had comparatively little influence on the deterministic approach to lexical selection.

Testing the hypothesis that speakers weigh both message alignment and production difficulty requires that we have evidence of both. This information would help quantify probabilistic decision processes. In other words, is lexical selection effectively a deterministic process except in unusual cases such as homophone production, or is probabilistic integration of several factors part of lexical selection, as is hypothesized in other motor behaviors?

To address this question, we designed a small artificial lexicon that allowed us to precisely manipulate the

Statement of Relevance

An important feature of language is flexibility—we have several words to describe almost any concept. Most language-production theories assume that we choose words that best reflect our intended message. An alternative, drawn from theories of motor control, is that word choices also reflect efficiency, such as saying the common word *cat* instead of the more precise word *kitten*. To date, it has been difficult to distinguish these theories. We developed a novel communication game in which players gave compass directions to elves digging for treasure. We manipulated features of the game to test whether players would prefer easy, well-practiced compass directions over rarer, more precise ones. We found that players frequently sacrificed precision for efficiency, even though precise directions earned more points in the game. These results shed new light on our word choices, including potential sources of miscommunication, and new insight on how language use may share features with motor control.

strength of both message alignment and accessibility in four experiments in order to quantify the interplay between these factors. We assigned novel words equidistantly along a single, continuous semantic space—directions on a compass. Participants communicated compass directions via typed responses to help elves hunt for treasure. The task of using a small number of compass terms to describe many different directions 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. We varied the frequency of the compass-direction names in an initial training phase, thereby affecting participants’ practice with different words and thus their accessibility—the ease with which these words could be retrieved and produced. We then assessed lexical selection behavior in a “treasure-hunt” communication game.

If production choices are driven only by message alignment, then participants should produce directions that best match the message prompt (compass arrow). However, if lexical selection also entails probabilistic integration of accessibility, then producers should sometimes respond with high-frequency words even when the low-frequency alternative is more aligned with the message. Because our artificial lexicon exactly specified the messages in the compass points in both learning

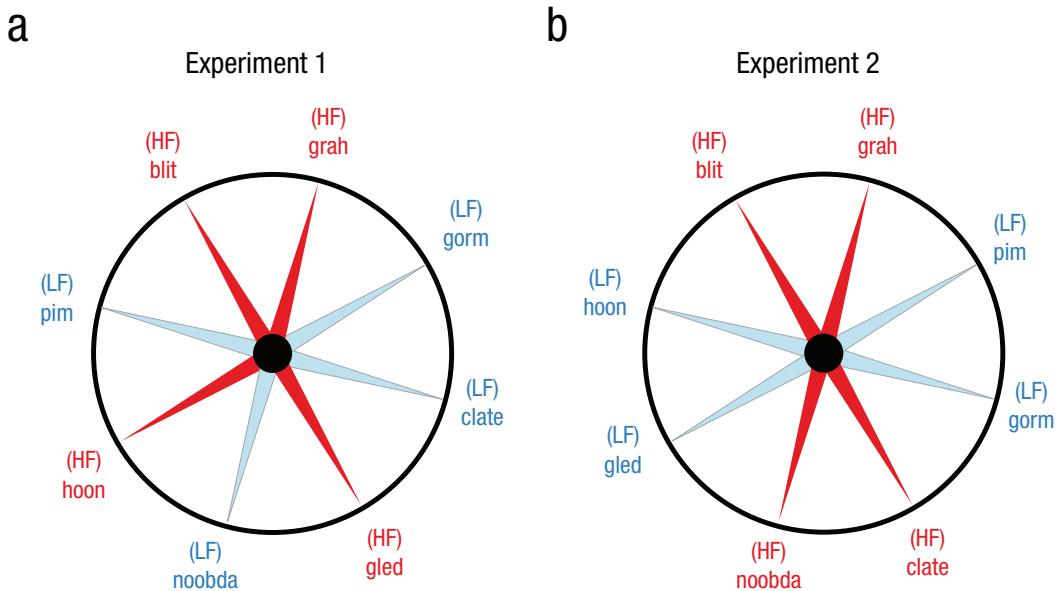


Fig. 1. The eight compass directions learned and the word frequency assigned to each compass direction during training in (a) Experiment 1 and (b) Experiment 2. High frequency (HF) and low frequency (LF) words are shown in red and blue, respectively. Novel words shown are examples; each participant received a different random assignment of words for the eight compass directions.

and test, and because we also controlled the relative frequency—and consequently, the accessibility—of different words in the language, we could quantify how message alignment and frequency affect lexical selection in a way that has not been possible to date.

Experiments 1 and 2

Our small artificial lexicon contained four novel high-frequency words and four novel low-frequency words, each of which referred to a precise direction on a compass. Experiment 1 tested the degree to which message alignment and word frequency affect participants' use of words in the language, and Experiment 2 replicated and extended our results to a different layout of compass points. All experiments were approved by the university's institutional review board.

Method

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

Materials. For each participant, eight novel words were drawn randomly from a set of 18 pseudowords (*pim*, *dak*, *vorg*, *yeen*, *grab*, *skod*, *gled*, *veek*, *blit*, *peka*, *sarp*,

minada, *boon*, *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° face of a compass image: 15°, 60°, 105°, 150°, 195°, 240°, 285°, and 330° (see Fig. 1). These compass positions were chosen to avoid translation to standard directions such as "north."

Each direction was assigned to a high-frequency category or a low-frequency category in one of two counterbalanced compass arrangements. In Experiment 1 (see Fig. 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 (see Fig. 1b) created an even balance between trials in which a high-frequency word was adjacent to another high-frequency word versus a low-frequency word (and vice versa).

Procedure. Participants played a game—programmed in *PsychoPy* (Version 1.85.0; Peirce, 2007)—in which their job was to help elves hunt for gold by indicating a search direction for buried treasure. The experiment consisted of a training phase, in which participants were taught novel words for the eight compass directions (see Fig. 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 a common method in language production, and typing and speaking have been shown to produce comparable results in studies of production choices (Gennari et al., 2012). Typing is known to be sensitive to word frequency, including in tasks with nonwords (Barry & Seymour, 1988; Baus et al., 2013; Kapatsinski, 2010).

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 a 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 4 times more frequently (as both a target and 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 each high-frequency word and one presentation of each low-frequency word. If participants scored below 80% on a 20-trial block, they were presented with another block of 20 word-learning trials. When 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 eight 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 4 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 recheck 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 eight compass directions than to others (see Figs. 2a and 2c). Each test stimulus direction was 0° to 11° away from a previously trained compass direction. For these trials, the 4:1 ratio of high-frequency words 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 on the basis of the compass to direct a group of elves toward a hidden treasure. Trials timed out after 5 s if participants did not begin typing.

In far-distance trials, participants were tested with randomly generated angles that were close to the midline of two compass directions, between 11° and 22° from each, creating conflict between two words that could guide the elves (see Figs. 2b and 2d; $n = 64$). On critical far-distance trials, the angle fell between a low-frequency word and a high-frequency word (Experiment 1: $n = 48$; Experiment 2: $n = 32$), although the compass direction always lay at least 2° closer to one compass direction than another. The trial design and feedback were otherwise identical to those of 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, we gave feedback to participants in Experiments 1 and 2 in the form of a score after each trial (see Fig. 3a), with points proportional to participants' 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 on the basis of the distance of the tested angle from the word entered; closer labels yielded 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 on the basis of the speed of participants' responses. For example, a faster reaction time of 300 ms corresponded roughly to a 4% change in base score (0–3 points). Thus, although 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 retested on their knowledge of the

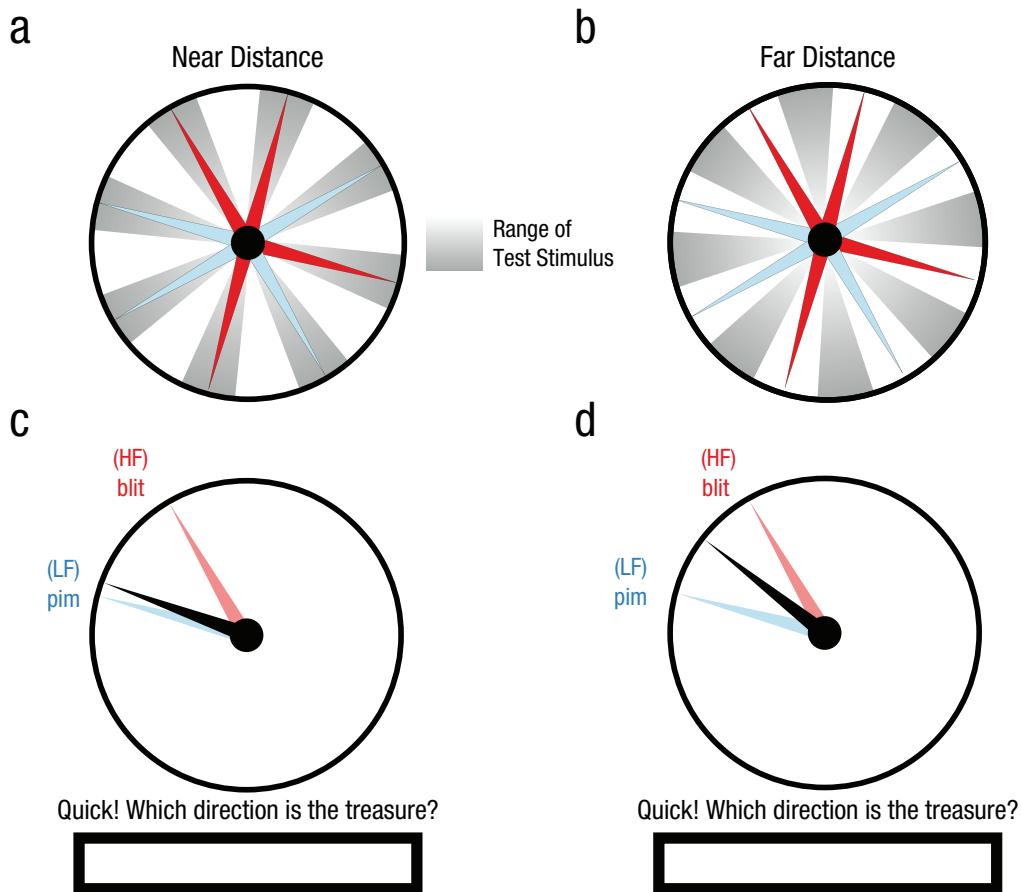


Fig. 2. Directions tested on (a) near-distance trials and (b) far-distance trials in Experiment 1, with examples of a (c) near-distance trial and (d) far-distance trial during the treasure hunt. For directions tested, the same distance manipulation was introduced in all experiments. The frequency of the trained compass directions is shown with the configuration used in Experiment 1. Red compass points were trained 4 times as often as blue points in the training phase. For experiment trials, participants saw only the direction in black. The two nearest compass directions and words in blue (low frequency [LF]) and red (high frequency [HF]) were added for illustration purposes and were not visible to participants.

eight trained compass directions. The first eight retention trials (the eight original 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-hunt 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.

Results

Word-training performance. Participants' accuracy across all word-learning blocks was high (Experiment 1: $M = 95.2\%$, $SD = 3.1\%$; Experiment 2: $M = 95.8\%$, $SD = 3.2\%$). On average, participants completed approximately

five 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 treasure-hunt portion of the game.

Word retention. At the end of the task, participants in both experiments showed high retention of both high-frequency words (Experiment 1: $M = 97.4\%$, 95% confidence interval [CI] = [94.8%, 100%]; Experiment 2: $M = 98.9\%$, 95% CI = [96.3%, 100%]) and low-frequency words (Experiment 1: $M = 97.4\%$, 95% CI = [94.8%, 100%]; Experiment 2: $M = 94.9\%$, 95% CI = [92.3%, 97.5%]) on the untimed retention test. These results suggest that participants maintained high levels of accuracy on both high- and low-frequency words at the end of the treasure hunt. High retention of the eight compass directions across participants ensures that any shifts in decision boundary during the treasure hunt are unlikely to be driven by

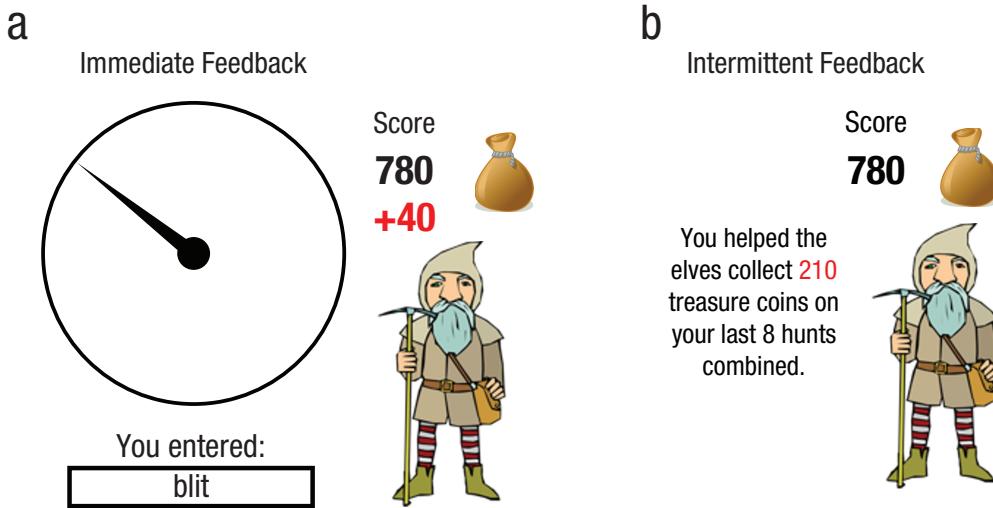


Fig. 3. Examples of immediate feedback (a) and intermittent feedback (b). In Experiments 1 and 2, participants received feedback after each game trial showing their response and gold coins earned. In Experiments 3 and 4, participants received no feedback after each response, and after every eight trials, they were shown the cumulative amount of gold coins earned during those trials.

forgetting particular labels (particularly the low-frequency labels). Participants also showed no significant difference in how quickly they responded correctly to high-frequency words and low-frequency words in the untimed retention task in either Experiment 1 or 2 (measured from the onset of a test prompt to the participant pressing the Enter key after typing the word). On the timed retention block, accuracy and reaction times for high-frequency words were comparable, although participants were slightly more accurate and faster to respond for high-frequency words than low-frequency words, collapsing across Experiments 1 and 2. There was no main effect of experiment version (Experiment 1 vs. 2) on accuracy and reaction times and no interaction between experiment version and frequency for either block, suggesting that the general learning patterns were similar across experiments. For further details, including an overview of participants' recall and response times for high- and low-frequency words in the final two retention tests at the end of each experiment, see Section S4 in the Supplemental Material available online.

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 message-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 used the *lme4* package (Version 1.1–27.1; Bates et al., 2015) in the R programming environment (Version 4.1.1; R Core Team, 2021) to 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], $z = -16.89$, $p < .001$. Crucially, when we controlled for distance from the nearest principal direction, participants were more likely to use the nearest word when it was a high-frequency word compared with a low-frequency word, $b = 0.71$, Wald 95% CI = [0.30, 1.12], $z = 3.37$, $p < .001$ (see Fig. 4a). This effect corresponded to an estimated 3.1° shift (95% CI = [1.3°, 4.9°]) in participants' decision boundary toward high-frequency words compared with low-frequency words. There was no interaction between frequency and trial type (near distance vs. far distance).

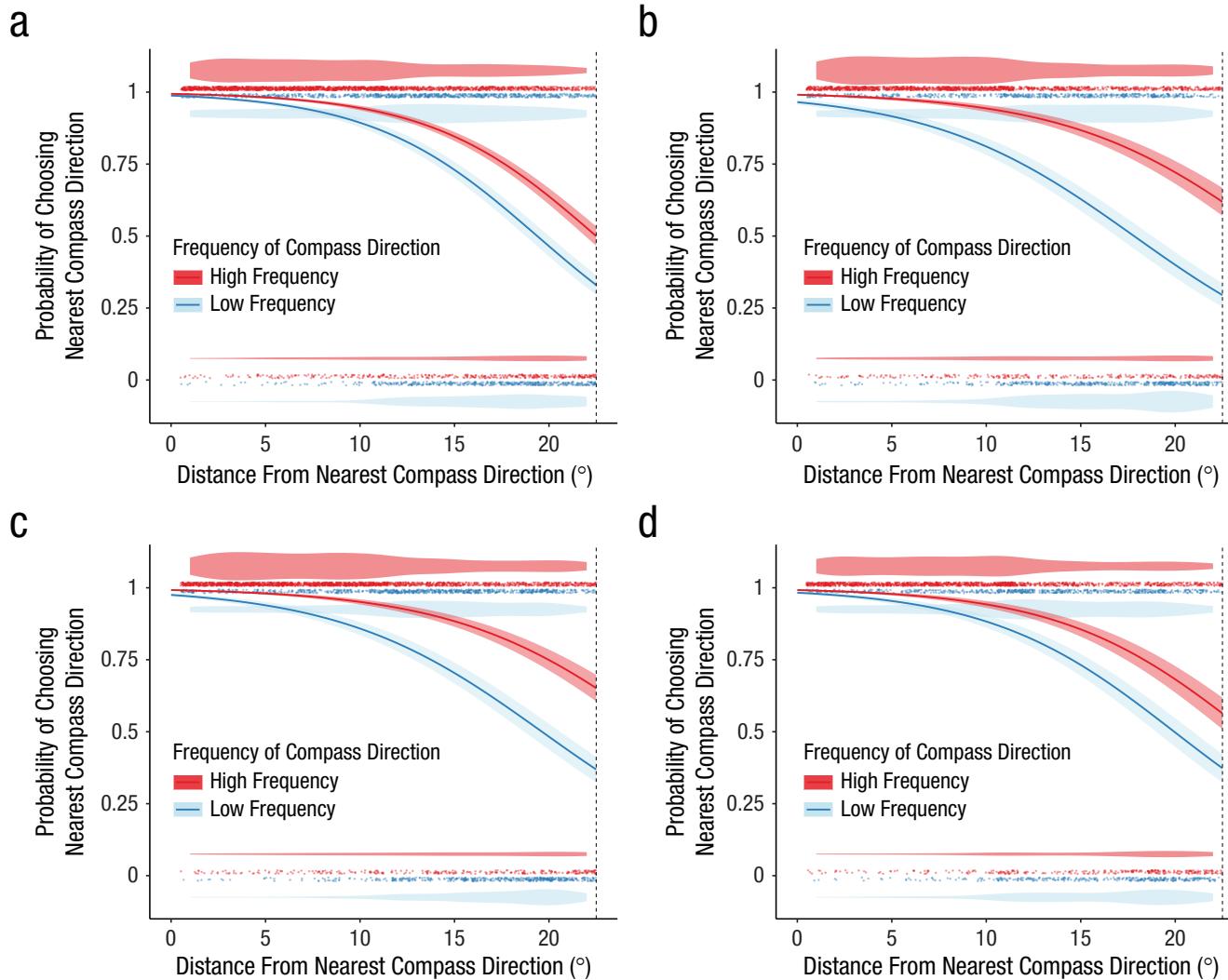


Fig. 4. Probability of choosing the nearest compass direction between a low-frequency word and a high-frequency word as a function of distance from the nearest compass direction, separately for (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, and (d) Experiment 4. Error bands represent ± 1 SE. Dots represent individual participant responses. Violin plots show the density of the response distribution; distributions at the top of the plot correspond to choices for the word corresponding to the nearest compass direction, and distributions at the bottom of the plot correspond to selection of the compass direction that is farther away (i.e., less accurate).

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 fitted 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, $b = 0.72$, Wald 95% CI = [0.31, 1.13], $z = 3.41$, $p < .001$, even after we controlled for participants' average final accuracy on the two neighboring compass directions.

Next, in an even more conservative test of the robustness of the frequency effect, we refitted the original logistic mixed-effects model including only participants who successfully named all compass directions correctly in the untimed retention test 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], $z = 2.76$, $p = .006$.

Experiment 2. To test the impact of frequency on participants' overextension tendencies, we fitted the same model as in Experiment 1. When we controlled 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 compared with a low-frequency

word, $b = 1.36$, Wald 95% CI = [0.75, 1.97], $z = 4.37$, $p < .001$ (see Fig. 4b). This effect corresponded to an estimated 7.2° shift (95% CI = [4.0°, 10.5°]) in participants' decision boundary for high-frequency words compared with low-frequency words. There was no interaction between near-distance trials and far-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], $z = 4.42$, $p < .001$.

Experiment 3

Experiments 1 and 2 showed 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 because 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. Experiment 3 removed frequent feedback and provided only multitrial, aggregated updates on scores.

Method

Participants. A new group of University of Wisconsin–Madison psychology undergraduate students ($N = 55$; 38 women; age: $M = 18.9$ 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 overcollection of data.

Design and procedure. The experimental design and procedure were identical to those of Experiment 2, with the following differences. First, to prevent excessive perseveration on the learning block, we limited each participant to 10 learning-test blocks before they automatically advanced to the treasure hunt. Second, unlike in previous experiments, participants were informed that they would receive intermittent feedback on their performance during the treasure hunt. A single, cumulative score was displayed every eighth trial (see Fig. 3b).

Results

Word-training performance. Participants' accuracy across all pair-learning blocks was high ($M = 94.4\%$, $SD = 4.2\%$). On average, participants completed approximately five pair-learning blocks ($M = 4.64$, $SD = 1.99$) before progressing to the treasure hunt.

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, with the exception that participants were slightly faster to respond to high-frequency words during the timed retention test (for details, see Table S3 in the Supplemental Material).

Test performance. To test the impact of frequency on participants' overextension tendencies, we fitted the same model as in Experiments 1 and 2. When we controlled for angle distance from the nearest compass direction, participants were more likely to use the nearest word when it was a high-frequency word compared with a low-frequency word, $b = 1.17$, Wald 95% CI = [0.58, 1.77], $z = 3.89$, $p < .001$ (see Fig. 4c). This effect corresponded to an estimated 6.0° shift (95% CI = [3.0°, 9.0°]) in participants' decision boundary for high-frequency words compared with low-frequency words. 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], $z = 4.05$, $p < .001$, 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], $z = 2.98$, $p = .003$.

Experiment 4

A concern about Experiments 1 to 3 is that familiarity with the trained compass directions was confounded with frequency of producing a word because every presentation of a compass direction was accompanied by the participant typing that direction's name. Thus, it is possible that the frequency effects in Experiments 1 to 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.

Method

Participants. A new group of University of Wisconsin–Madison psychology undergraduate students ($N = 43$; 24 women; age: $M = 18.7$ years, $SD = 0.88$; all native speakers of English) participated for course credit.

Design and procedure. The experiment design and procedure were identical to those of 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 received exposure only to the eight compass directions, we moved the angle in 45° increments. When 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, we designed the experiment so that the compass directions for which a low-frequency name would later be assigned appeared 4 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 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, we fixed the number of learning-trial blocks for participants. In this experiment, participants advanced to the treasure hunt after five learning blocks, which was the modal number of learning blocks that participants completed in Experiments 1 to 3.

Results

Compass and word-training performance. Participants were highly accurate in their memory for compass directions ($M = 97.7\%$, $SD = 3.8\%$) and across all pair-learning blocks ($M = 93.0\%$, $SD = 7.7\%$).

Word retention. As in the previous experiments, retention for high- and low-frequency words did not differ significantly in either the timed retention or untimed retention tasks, although accuracies were numerically higher for high-frequency words compared with low-frequency words (see Table S3).

Test performance. The model with the full random-effects structure from Experiments 1 to 3 did not successfully converge, leading us to prune the random-effects structure by removing the random slope for angle distance

(the predictor of least theoretical interest) to achieve model convergence. The results including the full random-effects structure from Experiments 1 to 3 yielded qualitatively similar results. Controlling for angle distance from the nearest compass direction, we found that participants were more likely to use the nearest word when it was a high-frequency word compared with a low-frequency word, $b = 0.77$, Wald 95% CI = [0.17, 1.38], $z = 2.51$, $p = .012$ (see Fig. 4d). This effect corresponded to an estimated 3.8° shift (95% CI = [0.8°, 6.8°]) in participants' decision boundary for high-frequency words compared with 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 = [-0.19, 1.04], $z = 1.35$, $p = .18$, indicating that the effect was more sensitive to the inclusion of participants with imperfect final retention of all compass directions in Experiment 4.

Unlike in Experiments 1 to 3, participants could advance to the treasure hunt prior to achieving 100% accuracy on the eight compass directions during the word-learning phase because 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], $z = 2.00$, $p = .045$.

General Discussion

Our novel language and communication game allowed us to quantify, for the first time, the degree to which language producers engage in probabilistic lexical selection and weigh both word accessibility and message alignment. 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 posttest 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 phonologically based accessibility factors may influence

the choice of words in language production (Ferreira & Griffin, 2003; Jaeger et al., 2012; Koranda & MacDonald, 2018; Rapp & Samuel, 2002). Our work extends these findings to effects of word frequency (see also Harmon & Kapatsinski, 2017) and furthermore quantifies the degree to which accessibility results in deviations from message alignment, owing to unique design features of our paradigm that allowed independent manipulation of message and accessibility. More work is needed to determine how consistently these effects appear across populations and various communication situations. For example, although our effects were consistent across several variants of the game, one important step is to extend these results to other modalities of language communication such as speech and sign.

These findings are related to several other language production phenomena, although more work is needed to determine whether similarities reflect related 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, such as speakers' use of high-frequency words such as *cat*, instead of the less accessible but more accurate word, *kitten*. Relatedly, children are sometimes found to produce overextensions for frequent words (e.g., calling a sheep a *dog*) despite comprehending the relevant word meanings (Gershkoff-Stowe et al., 2006; Naigles & Gelman, 1995).

Our results have several implications for theories of language production. First, feed-forward 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) do not predict our finding that implicit production choices balance message alignment and accessibility. Other accounts suggest that word form has only a limited effect on lexical selection because selection is generally completed before word-form computations have begun (e.g., Dell & O'Seaghda, 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 suggest 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). Although an accessible but less accurate word would not benefit the

comprehender, our results suggest that such trade-offs might be communicatively optimal. Piantadosi et al. (2012) presented computational simulations showing that selection of vague, common words over rare precise ones can result in a more efficient communicative system (see also MacDonald, 2013, who argued that increased efficiency for the producer aids the comprehender). In our communication game, points awarded for message alignment reinforce accurate communication to elf "listeners," yet participants still selected frequent words. Other studies have shown that producers regularly produce ungrammatical phrases, impairing comprehension but easing the task for the producer (Morgan et al., 2020). Together with our results, such findings suggest that people may routinely generate utterances that are easier to produce and good enough to communicate a message, despite increasing comprehension difficulty.

This work also has implications for how language production processes modulate language change over time (MacDonald, 2013). Probabilistic decisions favoring more accessible, high-frequency words may account for how the meaning of accessible words changes (Bybee, 2008) or how grammatical patterns are regularized over generations of use (Hudson Kam & Newport, 2005). Recently, Harmon and Kapatsinski (2017) showed that when two semantically equal options could describe a novel meaning, the high-frequency one was more reliably extended. Consistent with Bock's (1982) observation, links between concepts and words may become stronger for frequent words compared with infrequent words, or high activation can itself become a probabilistic cue to message alignment. Our results suggest that such extensions may occur even when high-frequency words are actually less precise, predicting a robust influence of frequency on diachronic change.

In summary, our data reveal that lexical selection is not deterministically driven by alignment with a message but is instead good enough—a probabilistic compromise between utility and efficiency. In other words, what we say is not always what we mean. Because similar utility-efficiency trade-offs also arise in nonlinguistic motor behaviors in humans and other species, this work narrows the range of production behavior that is unique to human language.

Transparency

Action Editor: Vladimir Sloutsky

Editor: Patricia J. Bauer

Author Contributions

M. J. Koranda and M. Zettersten developed the initial study concept. All the authors contributed to the study design. M. J. Koranda and M. Zettersten developed software for

conducting the study, and M. J. Koranda collected the data. M. Zettersten and M. J. Koranda analyzed and interpreted the data under the supervision of M. C. MacDonald. All the authors contributed to drafting and revising the manuscript and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was supported by the National Institutes of Health (NIH T32DC005359), the National Science Foundation (NSF-GRFP-DGE-1747503 and NSF-1849236), and the Wisconsin Alumni Research Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, National Science Foundation, or Wisconsin Alumni Research Foundation.

Open Practices

All data, scripts, and materials have been made publicly available via OSF and can be accessed at <https://osf.io/cvekx/>. The design and analysis plans for the study were not preregistered. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Mark J. Koranda <https://orcid.org/0000-0003-0204-1570>
 Martin Zettersten <https://orcid.org/0000-0002-0444-7059>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/09567976221089603>

References

- 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>
- 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>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Baus, C., Strijkers, K., & Costa, A. (2013). When does word frequency influence written production? *Frontiers in Psychology*, 4, Article 963. <https://doi.org/10.3389/fpsyg.2013.00963>
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89(1), 1–47.
- Bybee, J. (2008). Formal universals as emergent phenomena: The origins of structure preservation. In J. Good (Ed.), *Linguistic universals and language change* (pp. 108–121). Oxford University Press.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321.
- 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. <https://doi.org/10.1037/0033-295X.98.4.604>
- 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](https://doi.org/10.1016/S0022-5371(81)90202-4)
- Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, 14(1), 86–90. <https://doi.org/10.1111/1467-9280.01424>
- 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>
- Gershkoff-Stowe, L., Connell, B., & Smith, L. (2006). Priming overgeneralizations in two- and four-year-old children. *Journal of Child Language*, 33(3), 461–486. <https://doi.org/10.1017/S0305000906007562>
- 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. <https://doi.org/10.1023/A:1010421724343>
- 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>
- Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195. <https://doi.org/10.1080/15475441.2005.9684215>
- Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Incremental phonological encoding during unscripted sentence production. *Frontiers in Psychology*, 3, Article 481. <https://doi.org/10.3389/fpsyg.2012.00481>
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824–843. <https://doi.org/10.1037/0278-7393.20.4.824>
- 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), 157–215. <https://doi.org/10.1515/cllt.2010.007>
- Koranda, M. J., & MacDonald, M. C. (2018, March 15–17). *The breadth of phonological interference in word choice* [Poster session]. The 31st Annual CUNY Sentence Processing, Davis, CA, United States.
- Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, Article 226. <https://doi.org/10.3389/fpsyg.2013.00226>
- Melinger, A. (2021). Do elevators compete with lifts? Selecting dialect alternatives. *Cognition*, 206, Article 104471. <https://doi.org/10.1016/j.cognition.2020.104471>
- 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, Article 104417. <https://doi.org/10.1016/j.cognition.2020.104417>
- Naigles, L. G., & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: Preferential-looking in a study of dog, cat, and cow. *Journal of Child Language*, 22(1), 19–46. <https://doi.org/10.1017/S030500090009612>
- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291. <https://doi.org/10.1016/j.cognition.2011.10.004>
- 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. <https://doi.org/10.1037/0278-7393.28.3.564>
- R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.1) [Computer software]. <http://www.R-project.org>
- 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](https://doi.org/10.1016/0010-0277(94)90067-1)
- Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science*, 33(1), 75–103. <https://doi.org/10.1111/j.1551-6709.2008.01004.x>
- 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>