UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
When precision pays: Costs, benefits, and constraints in good-enough production

Permalink
https://escholarship.org/uc/item/2cj8v96v

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Mankewitz, Jess
Zettersten, Martin
Jacobs, Cassandra
et al.

Publication Date
2026

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

When precision pays: Costs, benefits, and constraints in good-enough production

Jess Mankewitz[1,2], Martin Zettersten[3], Cassandra L. Jacobs[4] & Robert D. Hawkins[1]
[1]Linguistics, Stanford University
[2]Psychology, University of Wisconsin - Madison
[3]Psychology, University of California, San Diego
[4]Linguistics, University at Buffalo

Abstract

Lexical selection involves a trade-off between message alignment (saying the most accurate word) and accessibility (saying the easiest word). We test predictions of a resource-rational model of lexical retrieval against a fixed-cost alternative, using a novel word-learning task. Participants learn words for novel compass directions and use these words to label prototypical and novel angles on the compass. In Experiment 1, we manipulated the relative frequency of the new vocabulary terms as well as the response deadline to test whether good-enough effects scale with accessibility differences and time pressure. In Experiment 2, we manipulate the payoff for precise responses to test whether speakers adjust their strategies when precision is more valuable. Together, these experiments provide a parametric test of the hypothesis that word choice reflects rational trade-offs between retrieval costs and communicative benefits, with implications for understanding language production as a form of resource-rational decision-making.

Keywords: language production; resource rationality; pragmatics; Rational Speech Act model

Introduction

Speakers readily adapt their utterances to context, producing more precise and informative descriptions when their audience requires them (Brennan & Clark, 1996; Degen et al., 2020; Yoon et al., 2012). Yet accessibility exerts its own pull. Speakers favor words that are easier to produce over those that better fit their intended meaning. For example, in an artificial language production study manipulating word frequency, Koranda et al. (2022) found that speakers often produced high-frequency words even when a low-frequency alternative would have been more accurate. Koranda et al. (2022) accounted for these data by highlighting that lexical choice only partially reflects a speaker's meaning and goals, as limitations on cognitive resources lead people to use "good-enough production" strategies (Ferreira & Patson, 2007; Goldberg & Ferreira, 2022; Harmon & Kapatsinski, 2017). Alternatively, speakers might weigh different alternatives depending on different pressures, modulating their choices depending on the anticipated effort needed to be precise or the likelihood of a reward.

But it remains unclear *how* exactly accessibility shapes word choice. Under a fixed-cost account, speakers allocate a roughly constant budget of effort to retrieval, and when the most informative word is difficult to access, this budget may be insufficient. An alternative interpretation, suggested by resource-rational approaches to cognition (Griffiths et al., 2015; Lieder & Griffiths, 2020), makes this trade-off explicit: speakers should invest more in retrieval when the expected communicative payoff justifies it, and default to accessible words when it does not. On this view, effort allocation itself becomes a decision variable, not a fixed constraint. A speaker who produces a more accessible but less informative word is not failing to be pragmatic. They are making a trade-off that reflects the relative costs and benefits of choosing a more specific term, as formalized by Futrell (2023). We distinguish these accounts using a process-level model of lexical retrieval and conduct two experiments that manipulate the difficulty of the production task by manipulating the time pressure (strict or lenient deadline) and the reward (bonuses for greater precision). Drawing on activation-based accounts of lexical access (Dell, 1986; Levelt et al., 1999), we model retrieval as a race.

Race models generate two distinct predictions, one about time pressure, and the other about the stakes of communicating. First, consider the *overall* accessibility bias toward high-frequency words. Under resource-rational and good-enough accounts, high-frequency words begin with a baseline activation advantage, but semantic information accumulates over time. Under the race model, the better a low-frequency word matches the intended meaning, the more quickly it gains activation. Therefore, additional time helps the more precise word overcome the frequency disadvantage. The magnitude of the overall frequency advantage should scale with the frequency ratio. The critical question is how time pressure affects word choice across different positions. Uniformly, or differentially depending on the stakes?

Time pressure and semantic uncertainty should jointly modulate lexical choice. Consider the *shape* of the decision function across spatial positions (Figure 1, Panel C). Here the accounts diverge. In the paradigm introduced by Koranda et al. (2022), high-frequency and low-frequency words cover neighboring spatial regions, so the stakes of word choice vary continuously with position: near the center of the low-frequency region, the low-frequency word is clearly more informative; near the boundary, the two are roughly equivalent. Under a fixed-cost account, speakers deploy constant retrieval effort regardless of these stakes. Time pressure should affect all regions equally, shifting the entire decision function toward high-frequency words without changing its shape.

Under the resource-rational account, effort allocation is sensitive to stakes. Near the low-frequency prototype, waiting a little longer is valuable because it allows speakers to produce

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

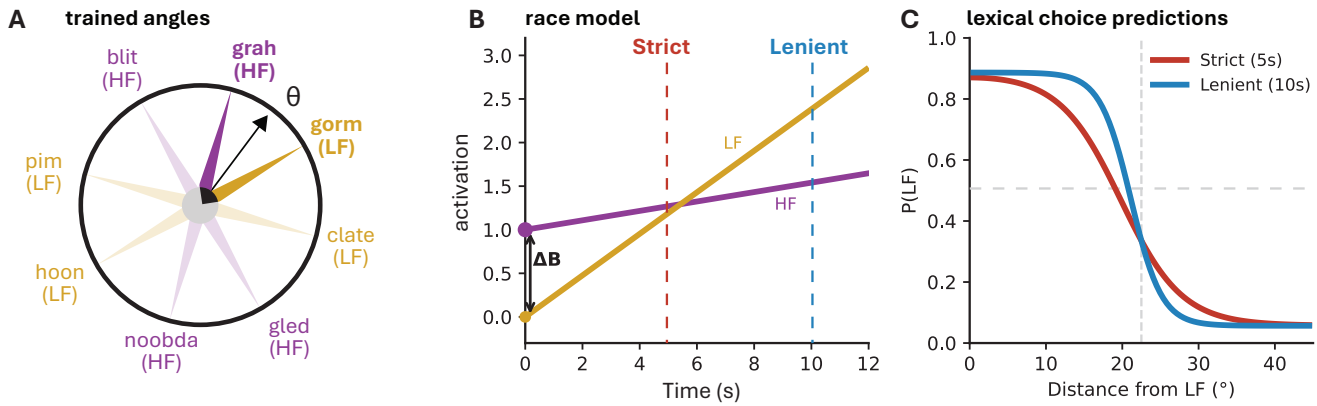

Figure 1: *Resource-rational analysis*. (A) We instantiate directions of different frequencies. Angles between HF and LF directions require the speaker to choose between HF and LF words. (B) Although HF words start with a baseline advantage (ΔB = frequency effect), the LF word's stronger semantic match allows it to accumulate activation faster, eventually "winning" if given sufficient time. (C) This model predicts the probability of producing the LF word as a function of distance from the LF prototype.

the more precise word. Near the boundary, waiting provides little benefit because the words are roughly equally informative, so the fast high-frequency response dominates. Critically, time pressure should magnify this asymmetry: with less time available, speakers at boundary positions never accumulate enough semantic evidence to overcome the baseline advantage, while speakers at prototypical positions still have time to do so. This yields a distinctive predicted signature: time pressure should *flatten* the decision gradient, affecting boundary regions more than prototypical ones, even if it does not shift the overall bias.

We test these predictions in two experiments extending Koranda et al.'s paradigm. In Experiment 1, we manipulate frequency ratio and response deadline to test whether (a) the overall high-frequency bias scales with frequency ratio but not time pressure, and (b) time pressure flattens the decision gradient as predicted by the race model. In Experiment 2, we manipulate the payoff for precise responses. If speakers allocate effort adaptively, they should show steeper gradients when precision is rewarded—a direct test of the resource-rational claim that effort is a decision variable, not a fixed constraint. We first formalize these intuitions in a race model that integrates pragmatic objectives with retrieval dynamics, then report two experiments testing its predictions.

A resource-rational analysis of lexical choice

We formalize speaker word choice by integrating two components: (1) a pragmatic objective function based on the Rational Speech Act framework (Frank & Goodman, 2012; Goodman & Frank, 2016), which captures why speakers care about word choice, and (2) a retrieval process grounded in lexical access dynamics (Jescheniak & Levelt, 1994; Levelt et al., 1999), which captures the costs of producing different words. The key insight is that retrieval cost is not a fixed penalty but depends on *time*. The decision to invest additional time in retrieval can then be analyzed as a value of computation (VOC) problem (Lieder & Griffiths, 2020).

Message accuracy Following the RSA framework, we assume speakers choose words to maximize informativity for the listener. Given a target location θ and a word w with prototypical meaning μ_w, the informativity of w is $\log P_L(\theta \mid w)$, the probability that a literal listener would recover the intended meaning. We model word meanings as Gaussian distributions, so informativity is highest when the target matches the word's prototype and decreases with distance. In our setup, three words race for selection: two high-frequency (HF) words at locations 0 and 1, and one low-frequency (LF) word at location 0.5. At the LF prototype ($\theta = 0.5$), LF is maximally informative. At boundary positions ($\theta = 0.25$), LF and HF are roughly equally informative.

Retrieval cost While informativity determines the *goal* of word choice, retrieval dynamics determine the *cost*. We model lexical access as a race where each word's activation accumulates over time (Levelt et al., 1999):

$$A(w,t) = B(w) + \delta \cdot M(\theta, w) \cdot t \tag{1}$$

where $B(w)$ is baseline activation (higher for HF words due to frequency), δ is the drift rate, and $M(\theta, w)$ is the semantic match between target θ and word meaning μ_w, modeled as a Gaussian:

$$M(\theta, w) = \phi(\theta; \mu_w, \sigma) \tag{2}$$

where $\phi(\theta; \mu_w, \sigma)$ denotes the Gaussian density evaluated at θ with mean μ_w and standard deviation σ. The width parameter σ controls how sharply word meanings are tuned: smaller values produce narrower, more precise meanings. Semantic match operationalizes the informativity intuition—words whose prototypes are closer to the target provide stronger bottom-up activation.

Speakers select words via a softmax over activations at the time of response:

$$P(w \mid \theta, t) = (1 - \lambda)\,\mathrm{softmax}(A(w, t)) + \lambda \cdot \frac{1}{3} \quad (3)$$

where λ is a lapse parameter capturing guessing due to motor errors or incomplete learning. We adopt a reduced parameterization, fixing the LF baseline $B_{\mathrm{LF}} = 0$ so that B_{HF} represents the *relative* frequency advantage (i.e., $\Delta B = B_{\mathrm{HF}} - B_{\mathrm{LF}} = B_{\mathrm{HF}}$; see Figure 1B). This leaves four free parameters: σ (semantic width), B_{HF} (frequency advantage), λ (lapse rate), and T (effective decision time). This formulation captures an important asymmetry: frequency effects are immediate, but semantic effects require time. At $t = 0$, only baseline activation matters. As time increases, semantic match increasingly dominates. In practice, we fix $\delta = 1$ for identifiability and estimate an *effective decision time* T per condition — the value of t at which the softmax in Eq. 3 is evaluated — capturing how much semantic evidence accumulates before a response is committed.

Value of Computation The decision to invest retrieval time can be framed as a value-of-computation (VOC) problem. Waiting longer allows semantic information to accumulate, improving the chance of selecting the most informative word. But waiting also has negative externalities, including missing response deadlines, incurring opportunity costs, and expending cognitive effort. The optimal strategy depends on the expected benefit or utility (EU) of additional processing:

$$\mathrm{VOC}(\mathrm{wait}) = \Delta\mathrm{EU}(\text{better word choice}) - \mathrm{cost}(\mathrm{time}) \quad (4)$$

Critically, this benefit varies with position: high near the LF prototype (where precision pays) and negligible at the boundary (where words are equally informative). This predicts an interaction between retrieval constraints and informativity stakes. Factors that increase the cost of waiting should have larger effects at boundaries than at prototypes.

Experiment 1: Cost pressure

We test this model in a conceptual replication and extension of Koranda et al. (2022). While Koranda et al. used a fixed 4:1 frequency ratio and 5-second response deadline, we manipulated both factors in a 2 (frequency ratio: 1:2 vs. 1:4) × 2 (time pressure: 5s vs. 10s) between-subjects design. This allowed us to test whether good-enough production effects scale with the magnitude of accessibility differences and whether time pressure modulates the trade-off between accuracy and accessibility.

Participants

We recruited 172 participants via Prolific. All participants were fluent English-speakers from the USA, the UK, or Canada. All participants successfully completed the task and were included in the final analysis.

Materials

Participants learned eight novel "Elvish" direction words corresponding to unfamiliar compass directions spaced at 45° intervals (i.e. 15°, 60°, 105°, etc). The pseudowords were *blit*, *grah*, *pim*, *gorm*, *clate*, *noobda*, *gled*, and *noom*, drawn from Amato and MacDonald (2010). Four words were designated high-frequency (HF) and four low-frequency (LF), with HF words appearing more frequently during training according to the assigned ratio condition (either 2:1 or 4:1). The mapping between specific words and compass angles was fully randomized for each participant. Frequency designation (HF vs. LF) was tied to angular position rather than to specific lexical items, ensuring that observed frequency effects could not be attributed to particular words.

Procedure

Training Phase We began by exposing participants to each of the eight direction-word pairings: a compass arrow pointing in one of the eight trained directions appeared alongside the corresponding word. Participants typed the word to confirm that they encoded it; we repeated the trial until they were correct. Participants then completed a series of two-alternative forced-choice (2AFC) training trials. On each trial, a compass arrow appeared with two word options (the target and a foil from a different compass region), and participants typed the word they believed corresponded to the displayed direction. Incorrect responses triggered auditory feedback. We again repeated the trial until correct. Training blocks continued until participants achieved ≥80% first-attempt accuracy within a block. The frequency manipulation was implemented in this phase. In the 1:2 condition, HF directions appeared twice as often as LF directions; in the 1:4 condition, HF directions appeared four times as often. Following each familiarization block, participants completed an untimed recall test, typing the direction name for each of the eight compass angles (presented in random order) without feedback. Participants were required to recall all eight direction names to proceed to the test phase. If the criterion was not met, participants returned to familiarization for additional training. This cycle repeated until criterion was reached (with a maximum of 10 cycles).

Test Phase The critical test phase assessed production under time pressure. The task was presented as a game in which participants were providing directions to elves in search of hidden treasure. On each trial, a compass arrow appeared and participants were asked to type one of the eight trained compass direction names. Compass arrows could appear at any angle, such that the target angle always fell at an untrained position between two of the eight learned compass directions. How close a compass arrow fell to its nearest neighboring compass direction varied randomly across trials, such that on some trials, the target angle was near one of the two compass directions, while on other trials, the target angle was closer to the midline between two compass directions. Participants were allocated a fixed amount of time to respond with one of

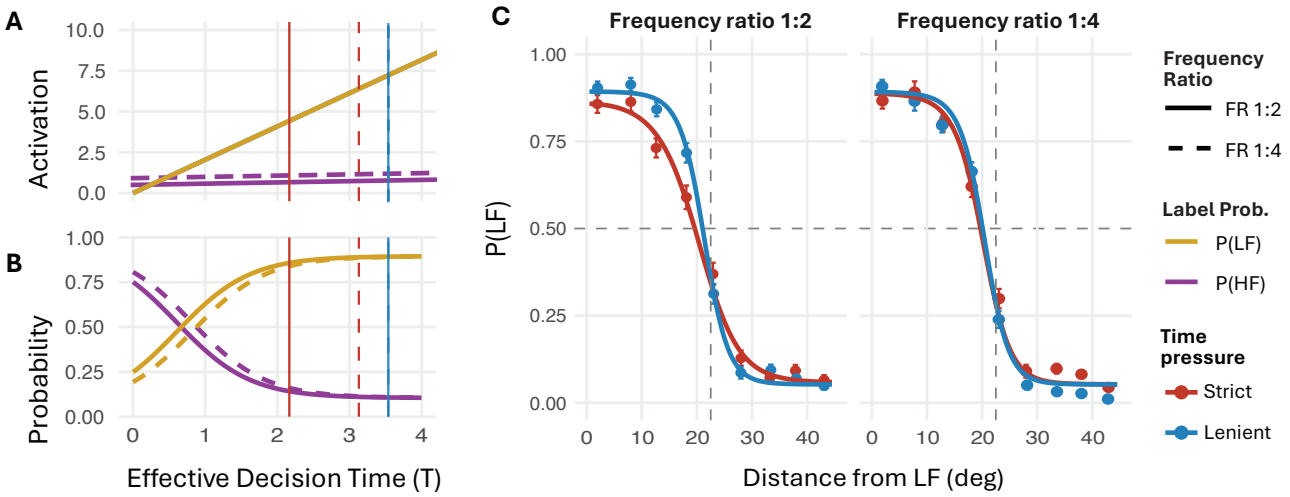

Figure 2: *Experiment 1: Race model dynamics and fit.* **(A)** Activation accumulation over time at the LF prototype ($\theta = 0.5$). HF (purple) and LF (gold) activation curves show baseline advantage and semantic evidence accumulation. Vertical lines mark fitted effective decision times T for strict (5s, red) and lenient (10s, blue) deadlines. **(B)** Production probability as a function of effective decision time T at the LF prototype. As T increases, P(LF) (gold) rises while P(HF) (purple) falls, showing how additional processing time favors the semantically matched word. Vertical lines indicate fitted T values. **(C)** Model fit to behavioral data. Probability of producing the LF word as a function of distance from the LF trained angle, by frequency ratio (facets) and time pressure (color). Points show binned means (± SE); curves show fitted race model predictions. Dashed line marks the boundary between trained angles (22.5°).

the eight compass directions (5 or 10 seconds, depending on condition). A visible countdown displayed the remaining time to respond and trials that exceeded the deadline were marked as timeouts. Responses not matching any trained word (e.g., typos) were excluded. Following Koranda et al. (2022), participants received performance-contingent bonuses calculated as a linear function of angular distance between the response and the target angle. Responses matching the target angle exactly earned maximum points, with points decreasing linearly to zero at 45° angular distance. Responses more than 45° from target, invalid words, or timeouts earned no bonus. Visual feedback displayed after each trial showed an elf discovering treasure (for responses within 45° of target) or an empty hole (for misses), along with the bonus earned. The relative frequencies of the compass direction names nearest to the target angle were maintained throughout the test trials to sustain differences in exposure or familiarity between angles. Participants completed 124 or 136 trials to maintain the frequency distributions. To reduce fatigue, breaks were offered every 40 trials. Following the main test block, participants completed an additional block of eight trials testing only the exact trained compass directions (in random order), also under time pressure. Participants completed a final untimed recall test of all eight directions to assess retention independent of time pressure. They then completed an exit survey including task difficulty and confidence ratings, descriptions of learning strategies, and optional demographic questions. Participants also self-reported whether they had used AI tools during the experiment.

Results

Participants required an average of 4 training blocks to reach the 80% accuracy criterion, with no significant differences across conditions, indicating that the frequency manipulation did not affect overall learnability. Timeout rates were substantially higher under the 5-second deadline (11.1%) than the 10-second deadline (1.5%), confirming that the time pressure manipulation successfully constrained response time. Timeout rates were not systematically biased with respect to LF or HF trials (5s: 10.8% vs. 11.7%; 10s: 1.0% vs. 1.5%; χ^2 tests $p > .12$). We analyzed all test trials in which the target angle fell between an LF and HF trained direction, excluding timeouts.

Model fitting. We fit the race model (Eqs. 1–3) to trial-level choice data. The drift rate δ was fixed to 1 for identifiability; we estimated semantic width (σ), lapse rate (λ), baseline advantage (B_{HF}), and effective decision time (T). We tested whether parameters varied across conditions using likelihood ratio tests between various reduced models. We found a significant effect of frequency ratio on baseline activation: B_{HF} was higher in the 1:4 condition than the 1:2 condition ($\chi^2(1) = 14.94$, $p < .001$). We also found a significant effect of time pressure on effective processing time ($\chi^2(2) = 41.73$, $p < .001$), as well as a significant frequency ratio × time pressure interaction on T ($\chi^2(1) = 20.19$, $p < .001$). Since models are fit to pooled data without participant random effects, p-values are approximate.

Parameter estimates. Estimates for shared parameters were $\sigma = 0.20$ and $\lambda = 0.16$. The baseline frequency advantage was larger under the 1:4 ratio ($B_{\text{HF}} = 0.92$) than under 1:2 ($B_{\text{HF}} = 0.50$), confirming that greater exposure asymmetry produced stronger baseline differences. The T fits revealed the predicted interaction (Figure 2A–B). Under lenient deadlines (10s), both frequency conditions converged to similar effective processing times ($T = 3.5$). Under stricter deadlines (5s), however, the conditions diverged: participants showed greater speedup in the 1:2 condition ($T = 2.2$, $\Delta T = 1.4$) than in the 1:4 condition ($T = 3.1$, $\Delta T = 0.4$). That is, time pressure reduced effective processing time more when the frequency gap was smaller, consistent with resource-rational predictions. Finally, the model makes an implicit response time prediction about response times (RTs): conditions with higher fitted T also showed slower empirical RTs (2.87s vs. 3.67s), despite T being estimated from choices alone.

Gradient flattening. Figure 2C shows the probability of producing the LF word as a function of distance from the LF prototype. Decision gradients are steeper under lenient deadlines and flatter under strict deadlines, with the 1:2/5s condition showing the flattest gradient (lowest $T = 2.2$). The race model captures this pattern well. This gradient flattening is the signature predicted by the resource-rational analysis. Time pressure reduces the opportunity for semantic evidence to accumulate, but this disproportionately affects boundary positions where LF and HF are roughly equally informative. Near the LF prototype, speakers produce LF at high rates even under time pressure; at boundaries, reduced processing time tips the balance toward the more accessible HF word. A fixed-cost account would predict uniform effects of time pressure across positions; instead, the interaction between position and time pressure supports the view that speakers allocate effort based on both task constraints and communicative stakes.

Experiment 2: Benefit pressure

Under the resource-rational framework, speakers should trade off the cost of producing a low-frequency word with the benefit of producing a more accurate word. Both Koranda et al. and Experiment 1 used a linear payoff structure that provided partial credit for less-accurate responses, such that producing a nearby HF word when an LF word would be closer still earned some reward. To test whether speakers adjust effort allocation based on the expected benefits of precision, we introduced a binary, all-or-nothing incentive structure in which participants only earn bonus payments if they produce the word for the angle *nearest* to the target, regardless of overall distance. In Experiment 2, we thus manipulated payoff structure (linear vs. binary), time pressure (5s vs. 10s), and frequency ratio (1:2 vs. 1:4) in a $2 \times 2 \times 2$, between-subjects design. The resource-rational account predicts that binary payoffs should increase speakers' willingness to invest retrieval effort: with no partial credit for "good enough" responses, the expected benefit of waiting for semantic evidence to accumulate is higher. This should manifest as longer effective processing times and steeper decision gradients under binary payoffs, particularly under lenient deadlines where speakers can afford to wait.

Participants

We recruited 400 participants via Prolific. Like Experiment 1, all participants were fluent English-speakers from the USA, UK, or Canada. All but four participants successfully completed the task and were included in the final analysis.

Procedure

Training followed the same protocol as Experiment 1: participants were exposed to each of the eight direction-word pairings, completed a series of 2AFC training trials until reaching 80% accuracy, and demonstrated retention through an untimed recall test.

Test Phase Participants were randomly assigned to one of two incentive conditions. In the *linear* condition, identical to Experiment 1, bonuses decreased linearly with angular distance between the target and produced word's trained compass direction; responses within 45°earned partial credit proportional to their accuracy. In the *binary* condition, participants earned a fixed bonus only if they produced the word for the angle *nearest* to the target (regardless of how close or far that angle was). Responses that were valid trained words but not the nearest option earned no reward. This manipulation held constant the difficulty of retrieval while varying the payoff for precision: in the linear condition, producing a nearby HF word when a LF word would be more accurate still earns partial credit. In the binary condition, only the nearest, most accurate word is rewarded. All other design features, including the time pressure manipulation, training criteria, and test trials were identical to Experiment 1.

Results

Training was comparable to Experiment 1, with participants requiring an average of 4.2 training blocks to reach the 80% accuracy criterion. Timeout rates were 7.7% under strict (5s) and 1.5% under lenient (10s) deadlines.

Model fitting We fit the same race model to trial-level data ($N = 26{,}022$ trials from 396 participants). To enable direct comparison with Experiment 1, we fixed $\sigma = 0.20$ and $\delta = 1$, estimating lapse rate λ, baseline B_{HF} per frequency ratio, and effective decision time T per condition. The frequency advantage B_{HF} was again larger under the 1:4 ratio ($B_{\text{HF}} = 0.42$) than under 1:2 ($B_{\text{HF}} = 0.18$), replicating the qualitative pattern from Experiment 1. The lapse rate was near zero ($\lambda \approx 0$).

Effect of Payoff Structure. A key test of the resource-rational account is whether payoff structure affects processing time. Following the same nested model comparison procedure used in Experiment 1, we compared nested models differing in

FR	Payoff	T (5s)	T (10s)	Mean RT (5s)	Mean RT (10s)
1:2	Linear	1.43	1.69	2.79	3.36
1:2	Binary	1.63	1.76	2.68	3.23
1:4	Linear	1.71	1.63	2.65	3.47
1:4	Binary	1.67	1.87	2.83	3.59

Table 1: Effective decision time parameter T fits by condition ($\sigma = 0.20$ and $\delta = 1$ fixed) and mean human RT (s) by condition.

how T depends on conditions. Likelihood ratio tests between nested models revealed a significant main effect of payoff, $\chi^2(1) = 5.91$, $p = .015$ (Fig. 3). Specifically, binary payoffs increased effective decision time relative to linear payoffs (Table 1). Neither the Payoff × Time nor Payoff × Frequency Ratio interaction improved fit over this additive model. However, the full model with condition-specific T values showed a significant three-way interaction, $\chi^2(2) = 16.12$, $p < .001$, indicating that the payoff effect varied across the combination of time pressure and frequency ratio in a pattern not reducible to simpler two-way interactions. This supports the resource-rational prediction: when precision is incentivized (binary payoffs), speakers invest additional retrieval effort. In the linear condition, producing a nearby HF word still earns partial credit, reducing the incentive to wait for the LF word to win the race. In the binary condition, only the nearest word is rewarded, increasing the value of additional processing time.

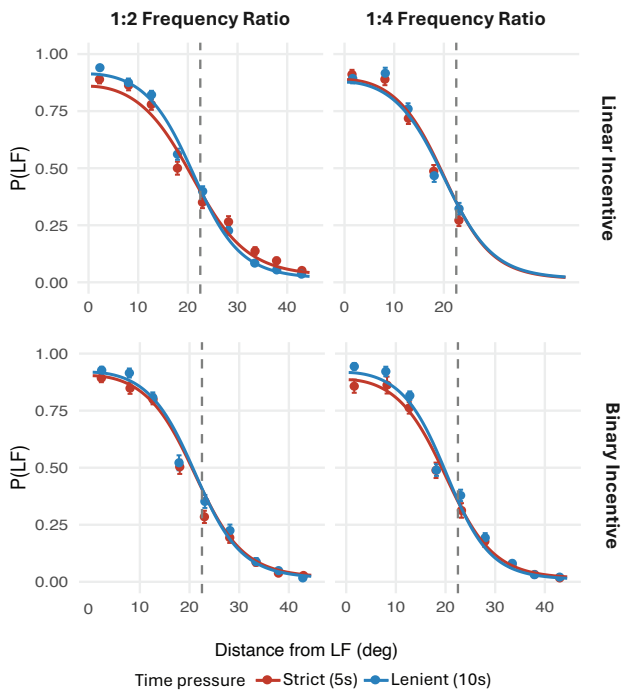

Figure 3: *Experiment 2: Model fit.* Probability of producing the LF word as a function of distance from the LF prototype, by payoff structure and frequency ratio (facets) and time pressure (color). Points show binned means (± SE); curves show fitted race model predictions. Dashed line marks the boundary (22.5°).

General Discussion

We propose a process-level account of good-enough production that links information-theoretic objectives (RSA) to classic activation dynamics in lexical retrieval. Across two experiments ($N = 568$), a single race model captures how speakers adapt word choice to retrieval costs (time pressure, frequency) and communicative benefits (payoffs). In Experiment 1, time pressure flattened decision gradients, reducing LF production more at boundaries than prototypes; in Experiment 2, rewarding precision increased effective processing time T.

This gradient flattening is the signature prediction of resource-rational effort allocation. While a fixed-cost account predicts uniform effects, we found time pressure increased HF production at boundaries where waiting provides little communicative advantage. At prototypes – where precision pays – LF production remained high. The interaction between time pressure and frequency ratio on T reinforces this interpretation: speakers wait longer when the accessibility gap is small enough that additional processing could change the benefits.

By integrating pragmatic objectives with retrieval dynamics, the race model explains *when* utility considerations override accessibility. RSA specifies what makes one word more informative than another; activation dynamics specify how long it takes to access each option. Together, they predict that good-enough production is not a failure of pragmatics but an adaptive response to time-sensitive retrieval constraints. This activation-based framing builds on information-theoretic accounts (Futrell, 2023) and inhibition-based models (Kapatsinski, 2022) to more directly explains why frequency effects diminish when form accessibility is equalized across alternatives, as in forced-choice tasks (Harmon & Kapatsinski, 2017).

One limitation is that T is a latent parameter fit to choices, not directly validated against response times. Future work should jointly fit choice and full RT distributions (or time-locked neural measures) to more directly validate the processing-time interaction. Additionally, our artificial lexicon isolates frequency from real-world correlates of accessibility; extending these signatures to naturalistic retrieval remains an open question.

Finally, we note that gradient flattening might simply reflect increased noise in the decision process. However, a pure-noise account would predict comparable flattening at prototypes and boundaries alike. Instead, we observed LF productions remained high at prototypes even under strict deadlines. More broadly, our work contributes to a growing body of research framing language production as resource-rational decision-making (Futrell, 2023; Gibson et al., 2019), where apparent "errors" or biases reflect adaptive responses to cognitive and communicative constraints rather than failures of the language system. On this view, good-enough utterances are not simply lapses in pragmatics but predictable consequences of rational control over time-sensitive retrieval.

References

Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*(1), 143–148.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition*, *22*(6), 1482.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological review*, *127*(4), 591.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Ferreira, F., & Patson, N. D. (2007). The "good enough" "approach to language comprehension. *Language and linguistics compass*, *1*(1-2), 71–83.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998.

Futrell, R. (2023). Information-theoretic principles in incremental language production. *Proceedings of the National Academy of Sciences*, *120*(39), e2220593120.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407.

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, *26*(4), 300–311.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, *7*(2), 217–229.

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

Jescheniak, J. D., & Levelt, W. J. (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.

Kapatsinski, V. (2022). Morphology in a parallel, distributed, interactive architecture of language production. *Frontiers in Artificial Intelligence*, *5*, 803259.

Koranda, M. J., Zettersten, M., & MacDonald, M. C. (2022). Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science*, *33*(9), 1440–1451.

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.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*, e1.

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*(4), 699–707.