Fast and slow errors: What naming latencies of errors reveal about the interplay of attentional control and word planning in speeded picture naming

Christina Papoutsi[1], Elli Tourtouri[1,2], Vitória Piai[1,3], Leonie F. Lampe[4], Antje S. Meyer[1,3]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[2] Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

[3] Donders Institute for Brain, Cognition, and Behavior. Radboud University, Nijmegen, The Netherlands

[4] Department of Linguistics, University of Potsdam, Germany

Author Note

Christina Papoutsi: https://orcid.org/0000-0002-7678-031X

Elli Tourtouri: https://orcid.org/0000-0003-0453-7377

Vitória Piai: https://orcid.org/0000-0002-4860-5952

Leonie F. Lampe: https://orcid.org/0000-0001-6020-8506

Corresponding author: Christina Papoutsi

Address: PO Box 310, 6500 AH Nijmegen, The Netherlands

Email address: christina.papoutsi@mpi.nl

Declarations

Acknowledgements and Funding Information

This research was supported by the Max Planck Society for the Advancement of Science, Munich, Germany. During data collection, Leonie F. Lampe was supported by an iMQRES scholarship from Macquarie University.

Conflicts of Interest

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices Statement

All datasets and analyses generated during this study are available on OSF: https://osf.io/s4eq6/ (Papoutsi et al., 2024)

Author contributions

Conceptualization: C.P, E.T., and A.M.; Formal analysis: C.P; Writing – Original Draft: C.P.; Writing – Review & Editing: C.P., E.T., V.P., L.F.L., and A.M.; Supervision: E.T., V.P., A.M.

Abstract

Speakers sometimes produce lexical errors, such as saying “salt” instead of “pepper”. This study aimed to better understand the origin of lexical errors by assessing whether they arise from a hasty selection and premature decision to speak (*premature selection hypothesis)* or from momentary attentional disengagement from the task (*attentional lapse hypothesis).* We analyzed data from a speeded picture naming task (Lampe et al., 2023) and investigated whether lexical errors are produced as fast as target (i.e., correct) responses, thus arising from premature selection, or whether they are produced more slowly than target responses, thus arising from lapses of attention. Using ex-Gaussian analyses, we found that lexical errors were slower than targets in the tail, but not in the normal part of the RT distribution, with the tail effect primarily resulting from errors that were not coordinates, i.e., members of the target’s semantic category. Moreover, we compared the coordinate errors and target responses in terms of their word-intrinsic properties, and found that they were overall more frequent, shorter and acquired earlier than targets. Given the present findings, we conclude that coordinate errors occur due to a premature selection but in the context of intact attentional control, following the same lexical constraints as targets, while other errors, given the variability in their nature, may vary in their origin, with one potential source being lapses of attention.

Keywords: speeded picture naming; target-error relation; response time distribution; lexical selection; attention

1. Introduction

Humans' ability to use language for expressing their ideas and communicating with others is remarkable, but spoken utterances are not always flawless. People occasionally produce words that do not optimally convey their intended message. For instance, consider the scenario where one requests the sugar, or the pepper at the dinner table, when they actually mean the salt. Crucially, speakers might sometimes produce erroneous names despite knowing the accurate term for a specific concept. While certain aspects of lexical errors1, such as their frequency of occurrence (i.e., error rates) and classification (e.g., semantic, visual, etc.), have been examined in production studies, their naming latencies are typically not analyzed. This is mainly because in studies involving neurotypical speakers, errors are rare and rather heterogeneous. However, the latencies of errors can reveal important insights into the cognitive mechanisms involved in word production, such as lexical selection, monitoring, and attentional control, as well as the interplay among these processes. So, how do these word slips arise?

One potential source could be related to a premature decision to produce a word, e.g., when the speaker is under time pressure. Being under pressure to produce a word, speakers may make a hasty word selection which might at times be erroneous. Indeed, language production studies have shown that erroneous word choices occur more often when speaking takes place under time pressure; higher error rates in combination with faster correct responses in time-constraining production tasks have suggested that such lexical errors result from a speed-accuracy trade-off (e.g., Lampe et al., 2023; Lloyd-Jones & Nettlemill, 2007; Vitkovitch et al., 1993; Vitkovitch &

[1] Even though erroneous language production involves various types of errors, such as sound non-word errors and blends (for a detailed list, we direct the reader to Dell, 1986), the present study focuses on whole real word substitution errors (e.g., saying salt instead of pepper), which we refer to as 'lexical errors' throughout the manuscript. These substitutions can be semantically related (e.g., saying salt instead of pepper) or unrelated (e.g., saying cat instead of salt).

Humphreys, 1991). This *premature selection hypothesis* attributes erroneous production to failures intrinsic to the core production processes (e.g., lexical selection) but assumes intact attentional control, allowing the participant to focus on the goal of naming the picture. According to this hypothesis, errors are produced at least as fast as correct responses stemming from a rapid and premature decision-making mechanism. More specifically, as the correct word and the other co-activated words race towards reaching the selection threshold, the word that is ultimately selected depends on their activation levels. Sometimes, the target word reaches the activation threshold quickly, leading to fast correct responses. Other times, the target's activation builds up more slowly, but is still higher than that of other words at the moment of selection, resulting in slower correct responses. Similarly, errors can occur when their activation reaches the threshold before that of the correct word. This can happen either because the alternative word is activated rapidly, leading to fast errors, or because the target's activation is slower, allowing the more strongly activated erroneous word to be selected prematurely even if its activation is not particularly fast, resulting in slower errors. In both cases, errors, just like correct responses, arise from a rapid (but premature) decision, as the system quickly settles on the word with the highest activation at the moment of selection. As variability in word activation levels during speech production can lead either the correct word or an erroneous word to reach the selection threshold first, errors and correct responses are expected to have on average similar latencies. There is no delay when an error occurs compared to when the correct word is selected.

Another potential source of lexical errors could be related to the speakers' momentary distraction from the task at hand (i.e., to name the picture). We will refer to this hypothesis as the *attention lapse hypothesis*. Theories of lexical access suggest that attentional control plays a central role in the success of naming performance (e.g., Roelofs, 2004, 2023a)**.** For example, delays in production onset latencies and higher error rates have been attributed to attentional lapses during word production (San José et al., 2021). The *attention lapse hypothesis* attributes lexical errors to failures in a domain-general executive control system, namely attention. This hypothesis predicts that errors are produced systematically more slowly than correct responses, because lapses disrupt focus, thus delaying word selection and increasing the chance of choosing the wrong word.

In the present study, we aimed to test the two hypotheses by examining how lexical errors and correct responses (henceforth targets) differ in both the means and distributions of their latencies. To do this, we used data from a speeded picture naming task, originally collected and reported by Lampe et al. (2023). In this task, participants were instructed to rapidly name a picture before a specific time deadline (600 ms). This is a task that has been previously used in order to elicit lexical errors from neurotypical speakers (e.g., Lloyd-Jones & Nettlemill, 2007; Moses et al., 2004; Vitkovitch & Humphreys, 1991). The high rate of lexical errors in this dataset, relative to the number of errors typically produced in standard picture naming experiments, allowed us to analyze, for the first time, the latencies of lexical errors and compare them to those of targets. In addition, we conducted a more focused analysis of the most common error type, semantic coordinate errors. In this way, we aimed to control for timing differences between lexical errors and targets resulting from differences in the nature of the errors.

Before describing the present study in more detail, we further discuss how lexical errors arise according to theories of word production, focusing on how the theories explain lexical errors arising due to premature selection and attention lapses.

1.1. Lexical errors may result from premature lexical selection

According to the *premature selection hypothesis*, lexical errors may arise due to a hasty selection within the word planning process. This hypothesis is compatible with models of lexical access, like the WEAVER ++ (Levelt et al., 1999; Roelofs, 1992, 2023b) and the Dark Side model (Oppenheim et al., 2010), yet with some differences in the specifics of the lexical selection process.

Word production theories agree on the following processing stages: concept activation, lexical selection, word form retrieval, and articulatory preparation for the overt production of the word (e.g., Abdel Rahman & Melinger, 2009; Dell, 1986; Levelt et al., 1999; Oppenheim et al., 2010). Some theories also incorporate a monitoring mechanism (further discussed in section 1.3), which checks the covert or overt output for errors (Levelt et al., 1999; Nozari et al., 2011; Roelofs, 2005). The activated concept spreads activation to its lexical representation (e.g., the word *kitten*) as well as semantically related words (e.g., *cat, pet, dog* and *tiger)*, due to their links within the lexical-semantic network. The strength of activation depends on how strong these links are. A main disagreement between models of lexical access lies in the mechanism needed to determine which of the co-activated representations will be selected. Here, we illustrate this potential mechanism of lexical access as implemented in two models with different lexical selection architectures, namely the WEAVER++ model (Levelt et al., 1999; Roelofs, 1992, 2023b) and the Dark Side model (Oppenheim et al., 2010).

In WEAVER++ (Roelofs, 2008, 2023b)[2], activation cascades from concepts to lemmas and word forms. The model posits that co-activated words compete for selection, with the time needed for selection depending on the target's activation level relative to all other co-activated words at the time of selection (relative selection threshold). The closer the activation level of the target and its competitors, the longer the production system needs to resolve the competition. To facilitate the selection, the model incorporates spreading activation coupled with procedural if-then rules based on the task demands. The rules boost the chosen concept's activation and enhance the efficiency of word selection and word form encoding. These rules guide not only word planning—ensuring the selection of a word and its form—but also executive control—ensuring that the goal of naming a picture is achieved. As soon as competition is resolved and there is a "winner", a verification operation takes place that checks whether the word that prevailed corresponds to the representation that was selected at the previous processing level (i.e., conceptual processing). Upon verification, the word is selected.

The WEAVER++ model, originally based on chronometric data of correct word production in healthy speakers, has been extended to account for lexical errors due to impairments, e.g., in people with aphasia (e.g., Roelofs, 2023b)**.** However, lexical errors produced by healthy speakers in speeded naming resemble those produced by aphasic speakers (e.g., Hodgson & Lambon Ralph, 2008; Mirman, 2011)**,** which suggests similarities in the mechanisms that are at play in the production of these errors (Dell et al., 1997). According to WEAVER++, lexical errors may arise if competition is resolved prematurely, either because the activation level of a competing

[2] The initial assumption of the WEAVER++ model was that lemma retrieval and word-form encoding were serial and discrete stages, with word form encoding starting only after lemma retrieval was complete (Levelt et al., 1999). However, accumulating evidence suggesting that activation of word form does not depend on lemma selection led to the omission of the serial-discrete assumption and the adoption of a parallel-continuous ("cascade") assumption.

erroneous word rapidly increases and significantly surpasses that of the target (or other co-activated words) at the moment of selection, or because the word form of the most strongly activated target word is not sufficiently activated for selection within a defined time window (e.g., due to infrequent use or brain damage). In such cases, the system may prematurely resolve any conflicts in the selection process, leading to the selection of another co-activated word that is more accessible at that moment. In this way, the system prioritizes completing the selection quickly rather than ensuring accuracy.

Contrary to the competitive account of lexical selection, Oppenheim et al.'s (2010) Dark Side model posits an absolute selection threshold. The activated word that exceeds this threshold is selected, independently of its activation difference to other co-activated words (Oppenheim et al., 2010; Oppenheim & Nozari, 2024). After selection, the activation levels of the non-selected words decay to zero and their semantic-to-lexical links are weakened. According to Oppenheim et al.'s (2010) model, both correct responses and errors arise from momentary fluctuations in activation levels that influence which word reaches the selection threshold first. For instance, if the target word “dog” and the word “cat” are both activated, fluctuations can cause “cat” to momentarily have a higher activation level. If “cat” surpasses the absolute threshold before “dog”, it will be selected, resulting in an error. This process could lead to both fast and slow errors: Fast errors occur when the erroneous word rapidly surpasses the threshold, for example due to being recently retrieved. Slow errors happen when the activations of both words are low, but the erroneous word still reaches the threshold first over a longer time. Similarly, correct responses can be fast or slow depending on how quickly the target word's activation exceeds the threshold. Since selection depends solely on whether a word's activation exceeds the absolute threshold, the timing of word selection is influenced by these fluctuations.

In summary, both (competitive and non-competitive) models can account for errors arising if the target word does not have a clear advantage over other words in the mental lexicon and by postulating a mechanism, namely attentional control or a flexible selection criterion, that ensures that a word is selected within a specific time window.

1.2 Lexical errors may arise due to lapses of attention

According to the *attention lapse hypothesis*, lexical errors may arise because the speaker is momentarily distracted from the task goal, i.e., to produce a word in a timely manner. The role of attention on cognitive tasks, such as speaking, has been a topic of discussion since the early 1900s, with researchers suggesting that attentional disruptions could influence aspects of naming performance, like the shape of the response time (RT) distribution. More recently, studies of picture naming have shown that momentary lapses of attention can change onset latencies in vocal utterances, by increasing the number of slow (correct) responses, i.e., the tail of distribution (e.g., Jongman, 2017; Jongman et al., 2015; San José et al., 2021), and lead to higher error rates (descriptively shown in San José et al., 2021). But how can lapses of attention affect naming performance in this way? Within the competitive model of lexical access (WEAVER++ model, Roelofs, 1992), San Jose et al. (2021) provided a computational account and empirical evidence to explain how attentional lapses modulate semantic interference in picture naming. More specifically, they proposed that lapses of attention can lead to goal neglect in a subset of trials, with activation spreading unsupervised throughout the lexical-semantic network. As a result of this unsupervised spreading activation, the conflict between co-activated words builds up and it takes longer to resolve, thus leading to a stronger interference effect at the tail of the distribution and (descriptively) to an overall increase of produced errors.

It is important to note that even though goal neglect as a result of lapses of attention has been proposed and implemented in the competitive WEAVER++ model, it is also compatible with non-competitive models of lexical access. For example, attentional disengagement from the task goal could disrupt the flow and rhythm of spreading activation within the lexical-semantic network, thus yielding longer RTs and more errors.

1.3 Role of post-selection monitoring in erroneous production

Some word production theories assume the existence of a self-monitoring mechanism, which detects and repairs errors before (internal) or after (external monitoring) they have been overtly produced (e.g., Levelt et al., 1999; Nozari et al., 2011; Roelofs, 2005). The two most prominent monitoring accounts are the *comprehension-based model* (Levelt, 1983; Levelt et al., 1999; also implemented in WEAVER++, Roelofs, 2005) and the *conflict-based model* (Nozari et al., 2011). Despite differing in specific aspects of internal monitoring, such as where and how they operate within the production system (for a review, see Gauvin & Hartsuiker, 2020), both models similarly predict that time-constraining contexts (as reflected in the task used in the present study) affect naming performance, making error detection more challenging. Here, we briefly describe both models. Note that it is not the aim of the study to adjudicate between the two models.

The *comprehension-based model* describes internal monitoring as a process external to the core production processes (semantic, lexical and phonological encoding). It uses the comprehension system to detect errors, by verifying whether the planned word form matches the initially created message for production. In case of discrepancy, speech is halted and the word planning process restarts (Noteboom, 1980). Monitoring is attention-demanding and error-prone (Roelofs, 2005).

Time pressure can cause lapses in the verification process, leading to undetected errors. Finally, the effectiveness of monitoring varies, as errors that closely resemble the intended word in meaning or sound are harder to detect than unrelated errors.

The *conflict-based model* describes internal monitoring as a process internal to the production system. The monitor detects errors by comparing the level of conflict between co-activated representations (computed as the inverse of the difference between the activation level of the two most highly activated representations) against a set criterion, defined as a level of conflict threshold that determines whether the level of conflict between representations is high or low. If the level of conflict between representations is below the criterion, the conflict is low and the monitor accepts the selection of the most activated representation; otherwise, the conflict is high, and the monitor delays selection in favor of additional processing. This criterion is flexible and is modulated by task demands and goals (Nozari & Hepner, 2019). For example, if the goal is to prioritize speed over accuracy, the criterion shifts, and the monitor more often fails to detect the selection of errors.

In sum, both monitoring models agree that errors are more likely to bypass the monitoring system under time pressure than in less time-constraining contexts, due to reduced attentional resources or adjustments in the error detection criterion.

1.4 Present study

The aim of the present study was to test whether lexical errors arise from a hasty selection within the production system, with attentional control remaining intact (i.e., *premature selection hypothesis)*, or from lapses of attention (i.e., *attention lapse hypothesis*). To do so, we examined how target and error responses differ in both the means and distributions of their latencies by using a relatively large dataset from a speeded picture naming task (Lampe et al., 2023).[3] The *premature selection hypothesis* would predict that, all else being equal, errors are produced at least as fast as targets throughout the RT distribution. On the other hand, the *attention lapse hypothesis* predicts that errors resulting from momentary distractions are uniformly slower than correct responses throughout the RT distribution, all else being equal.

We performed two types of statistical analyses on the data: a Linear Mixed Effects Model (LMEM) analysis and an ex-Gaussian analysis. The LMEM allowed us to examine overall effects of the predictor (response accuracy) on RTs, while accounting for variability between items and participants.

The ex-Gaussian analysis offers complementary information about how RTs are distributed for the target responses and lexical errors. This analysis models the positively skewed RT distributions as the combination of two underlying distributions—a Gaussian (or normal) and an exponential distribution (e.g., Balota et al., 2008; Ratcliff, 1979). The normal distribution represents the bulk of the RTs, while the exponential distribution accounts for a small but significant proportion of extremely long RTs (i.e., tail of the distribution). The ex-Gaussian analysis estimates the mean (μ) and standard deviation (σ) of the Gaussian component, and the mean and standard deviation (τ) of the exponential component, offering insights into distribution shifts (μ effects) and changes in skewness (τ effects). Together, the two analytical approaches provide a more comprehensive understanding of the RT data.

[3] It is important to note that the dataset provided by Lampe et al. (2023), while containing a substantial number of errors produced in the speeded naming task, did not undergo RT analyses of these errors in their original study. Instead, their research focus was on investigating the RTs associated with correct responses, thus not addressing the temporal dynamics of erroneous responses. This presents an untouched area of research that the present study aims to explore.

In the present dataset, participants produced different types of lexical errors, including errors that were semantically related to the targets and semantically unrelated errors. Semantically related errors were by far more frequent. Within the class of semantically related errors, semantic coordinates (e.g., saying “duck” for the item ‘goose’) constituted the most common error type. These errors are of particular theoretical interest because the selection of a target from a set of co-activated members of a semantic category is a major concern of current theories of lexical access and has been widely studied experimentally (using the semantic blocking paradigm, e.g., Abdel Rahman & Melinger, 2009; Belke et al., 2005; Damian et al., 2001; using the picture-word interference paradigm, e.g., Damian & Martin, 1999; Glaser & Düngelhoff, 1984; Schriefers et al., 1990). Therefore, we conducted an additional analysis on semantic coordinate errors, and investigated whether these errors systematically differed from target responses in their naming latencies.

It is well known that psycholinguistic properties of words influence performance in (correct) word production (e.g., Alario et al., 2004; Perret & Bonin, 2019). A related, but relatively sparse, line of research focuses on the relationship between targets and naming errors with respect to psycholinguistic variables, aiming to understand which word-intrinsic properties of the erroneous response cause the selection of this specific lexical representation. Previous research examining the target-error relation in word frequency and age of acquisition has led to inconsistent results, with some studies finding that erroneous words tend to be more frequent and earlier acquired than target words (word frequency: Del Viso et al., 1991; Kittredge et al., 2008; Koranda et al., 2022; age of acquisition: Bormann et al., 2008; Gerhand & Barry, 2000), and others not observing a word frequency and age of acquisition advantage of erroneous responses over target words (word frequency: Best, 1996; Dell, 1990; Vitevitch, 1997; age of acquisition: Kittredge et al., 2008). Finally, there are consistent findings about the target-error relation with respect to their phonological aspects, with errors mirroring the target words in their number of syllables or segments (e.g., Gordon, 2002; Harley & MacAndrew, 2001).

In line with this previous research, we explored the relation between semantic coordinates and target responses in terms of their word-intrinsic properties, namely word frequency, age of acquisition, length, conceptual familiarization and imageability. By comparing targets and coordinate errors in their word-intrinsic properties, we aimed to better understand the factors influencing activation strength of representations in the mental lexicon. This approach can help to explain why participants in the experiment made specific erroneous lexical choices over the targets. Together, the multifaceted data from examining the target-error relations in word-intrinsic properties and the analysis of naming latencies of errors allow us to construct a more comprehensive model of the cognitive processes governing word selection and production.

2. Description of the dataset by Lampe et al. (2023)

2.1. Participants

We used the dataset collected by Lampe and colleagues (in the form of audio recordings) from 82 English native speakers, who were recruited from Macquarie University's Psychology participant pool to participate in their study (2023). As mentioned in the study for which the data were originally collected for (Lampe et al., 2023), participants provided written informed consent to participate and received course credit or monetary compensation for their participation. Before any data pre-processing and analyses took place, the audio data were pitch-shifted using Praat (Boersma & Weenink, 2022) in order to ensure the participants' anonymity.

Participants were 17–35 years old, right-handed, had normal or corrected vision and no history of neurological, language or cognitive disorders.

Participants completed a speeded picture naming task, which was preceded by either one or two standard naming rounds for the same pictures (data reported in Lampe et al., 2021, 2023). For the present study, only the data corresponding to the speeded picture naming round was used. Data from two participants were excluded from the analysis because they performed poorly in the standard naming rounds (less than 60% overall accuracy). Consequently, the data from 80 participants (age: *M* = 20 years, range = 17–33 years; 63 females and 17 males) was included in our analyses.

2.2. Materials

Colored photographs of 297 items taken from the McRae et al. (2005) database were used as stimuli in the original study. All items had high name agreement (>75%), as assessed and reported in Lampe et al. (2021). Stimuli were presented in four experimental blocks (Block 1: *N* = 35 items, Blocks 2 and 3: *N* = 87 items, Block 4: *N* = 88 items) whose presentation order was randomized, thus generating six experimental lists. Lampe and colleagues (2023) presented the stimuli in a pseudorandomized order, such that there were at least two items of different semantic categories intervening between two categorically related items. There were 35 semantic categories that contained on average 8.5 items.

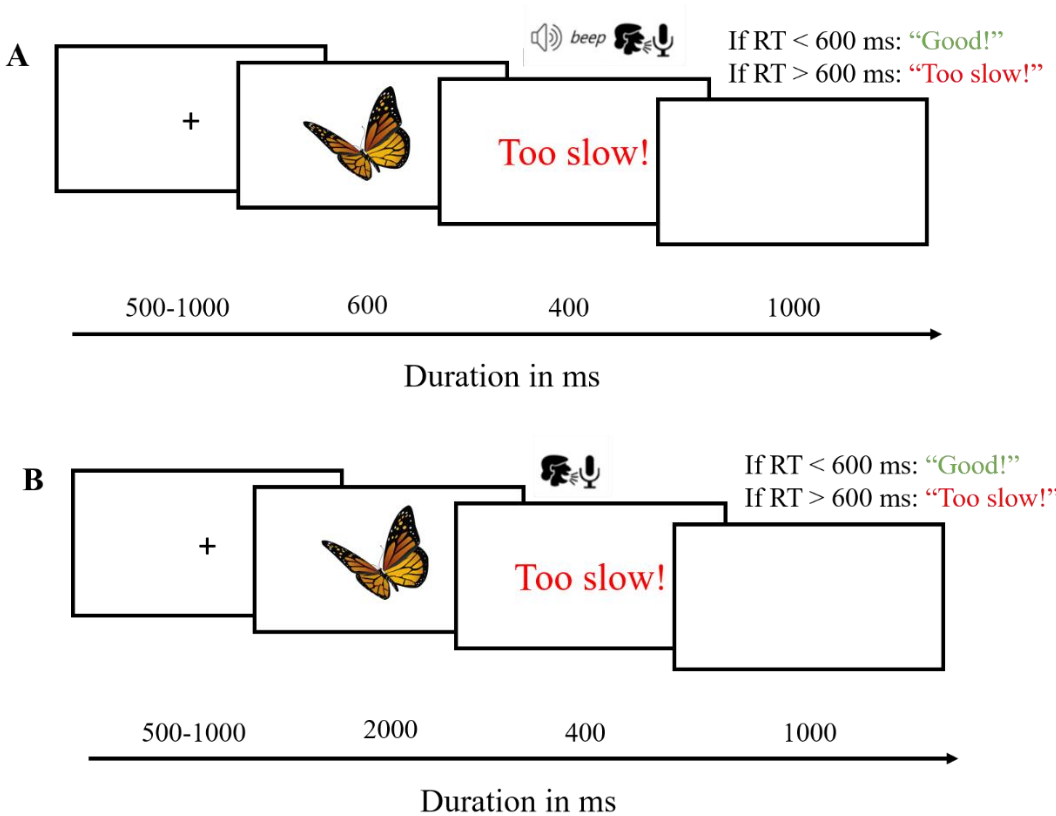
2.3. Procedure of Speeded Naming Task

In the speeded picture naming task, Lampe and colleagues (2023) first familiarized the participants with the expected naming speed using practice trials. This happened five times throughout the experiment, once before each block. Altogether, 25 practice items were presented.

In these practice trials, the picture was shown for 600 ms, after which a beep sound was played and the picture disappeared from the screen. Participants were instructed to name the picture as quickly as possible, ideally while it was still on-screen, and to “beat the beep”, prioritizing naming speed over accuracy. If participants started naming the picture within 600 ms post picture onset, they received positive feedback (“Good!”) at the end of the trial. If they started naming the picture later, they received negative feedback (“Too slow!”). In the experimental trials of the speeded task, the display of the pictures was longer (2000 ms), so that it had the same duration as the picture display of the second standard picture naming task, as comparing performance on the two naming tasks was the research focus of Lampe et al. (2023). However, participants were instructed to name the pictures in the same way as they had done in the practice trials and as if the beep sound were still played. As in the practice trials, positive or negative feedback was given depending on whether response onset was faster or slower than the 600 ms deadline, respectively. Figure 1 shows a trial overview of the practice (Panel A) and experimental phase (Panel B). The study procedure, but not the analyses conducted here, were preregistered by Lampe and colleagues on the Open Science Framework (Lampe et al., 2019; https://osf. io/yw6ma/). The datasets and analyses generated during this study are available on OSF: https://osf.io/s4eq6/.

Figure 1

Overview of a Practice Trial (Panel A) and an Experimental trial (Panel B) in the Speeded Picture Naming Task



Note. The boxes indicate the screens participants saw and the numbers below indicate presentation durations in ms. Abbreviations: RT = Response time. Adapted from "Are they really stronger? Comparing effects of semantic variables in speeded deadline and standard picture naming", by Lampe et al., 2023, Quarterly Journal of Experimental Psychology, 76 (4), p. 769 (https://doi.org/10.1177/17470218221103356). The butterfly image is a representative example of the images used in the experiment. The image is in the public domain under CC0 (Creative Commons Zero) license. Source: https://commons.wikimedia.org/wiki/File:Danaus_plexippus.svg.

3. Coding of responses and naming latencies

Following the anonymization of the audio data, response times (RTs) were manually coded per trial using Praat (Boersma & Weenink, 2022). In order to avoid possible inter-rater biases, we coded the RTs of all responses (targets and errors). The computation of the intraclass correlation coefficient (comparing the RTs for correct responses between the original study and the presentS study) showed that the inter-rater agreement for the RTs of the target responses was high, using the two-way random effect models and "single rater" unit, kappa = 0.99, $p < .001$.

The participants' raw responses and their classification into broad and more detailed categories were already provided by Lampe and colleagues (2023; data are publicly available at the Open Science Framework website: https://osf.io/5r8fp/). This study followed the response classification reported by Lampe and colleagues (2023), which was based on the response classification by Fieder et al. (2019). However, for a small number of responses the coding was modified, as the initial transcription was not in accordance with the raw auditory data[4]. This was the case in 66 trials, corresponding to 0.26% of all data points (N = 24,354). A second independent rater was used to rate these trials and any discrepancies were resolved through consensus discussion.

A detailed description of the response classification with examples is reported in Appendix A. Responses corresponding to the target words were considered *target responses.* These responses included correctly produced target words (with or without a determiner), disfluency on the initial phoneme or syllable or a hesitation marker, elaborations, abbreviations, and synonyms.

[4] In most of the modified trials, there was clear disfluency preceding word onset, but this disfluency was not transcribed in the initially annotated raw response. However, this information is important for the present study, as naming latencies of responses constitute the main datatype of interest.

Responses that did not correspond to the target words were considered as errors. These responses consisted of semantically related errors, other errors, and omissions. *Semantic errors* consisted of the following response types, as coded by Lampe and colleagues (2023): superordinates, subordinates, coordinates, associates, semantic others, part-whole relationships, and incomplete responses that shared at least 50% of their phonemes with a semantically related item or vice versa. Finally, erroneous responses revealing a two-step error including a semantic and a phonological error were also considered semantically related errors. *Other* errors comprised the following response types: semantically and phonologically unrelated words, parts of compound target words, self-corrected responses preceded by disfluencies that were shorter than one English syllable, visually related errors, false starts consisting of less than one syllable that were target-related or target-unrelated and were not followed by a full response, non-words, phonologically related errors (words or non-words), and morphologically-semantically-phonologically related errors, which were classified based on the ambiguity of the response. Finally, trials where participants did not produce a response or expressed failure to respond were classified as *Omissions.*

In the response classification indicated by Lampe et al. (2023), only the target responses were coded for disfluencies (e.g., repetition of initial phoneme or syllable, such as *ho..hose* for the target word *hose*). However, there were also several erroneous disfluent responses (e.g., saying *pah..goat* for the target word *cow*). As this information is crucial for the consideration of these responses in the RT analysis of the present study, we additionally classified all responses as either *fluent* or *disfluent*.

4. Data analyses

Following the exclusion of two participants (see section 2.1), there were a total of 23,760 data points (297 trials for 80 participants), of which 77.2% were *target* responses (*n* = 18,338), 21% were *erroneous* responses (*n* = 4,992), and 1.8% were *omissions* (*n* = 430). All *omissions* were excluded from this and all following analyses. *Disfluent* responses (*n* = 736; 3.09% of total observations) were excluded from the analysis, thus leaving only fluent target and erroneous responses (*n* = 22,347; 94.05% of total observations).

4.1 Primary analysis: Including all lexical errors

For this analysis, we included the data of all participants (*n* = 80), but only a subset of items (*n* = 285) for which both target and error responses were elicited, given our research aim and in order to avoid potential biases that could arise from including items that yielded only correct responses. Regarding the errors included in the RT analysis, only responses corresponding to real words of the English language (without any determiners) were included: from the category *semantic errors,* coordinate, superordinate, subordinate, associate, other semantic, and part-whole responses; from the category *other errors,* phonological and mixed morphologically-semantically- phonologically related errors only consisting of real words, unrelated, visual, responses that were parts of compound target words. Regarding the target responses included in the RT analysis, responses that contained only the target word (i.e., without a determiner, hesitation, or disfluency) were included. Finally, responses that were produced faster than 200 ms (only one observation) were excluded.

This resulted in a final dataset of 21,157 data points, of which 81.1% were target responses and 18.9% were errors, of which 15.9% were *semantic errors* and 3% *other errors*. Table 1 shows the raw count and percentage of the target and erroneous response types that were included in the LMEM analysis.

To test for differences in RTs between target and erroneous responses, we performed a LMEM models analysis using the lme4 package (version 1.1-31, Bates et al., 2015), with *p*-values being derived using lmerTest (version 3.1-3, Kuznetsova et al., 2017) in R Studio (version 2022.12.0.353, R Core Team, 2022). Raw RTs were log-transformed for the analysis in order to reduce the positive skewing in the distribution (Baayen et al., 2008). The factor *Response Accuracy* (target vs. error; treatment-coded, errors as reference) was included as a fixed factor. The random effects structure included by-participant and by-item random intercepts and random slopes for Response Accuracy. In addition, we visually examined the RT distribution using the vincentizing method (Ratcliff, 1979). We rank-ordered the raw RTs from the fastest to the slowest for each Response Accuracy condition and participant and divided them into four 25% quantiles of naming latencies. Subsequently, each quantile was averaged first per participant and then per condition.

Finally, we decomposed the RT distributions into the three ex-Gaussian parameters, μ, σ, and τ. More specifically, we used the quantile maximum likelihood method (Brown & Heathcote, 2003) to estimate the three parameters. Given that this method requires at least 40 observations per condition and participant in order to yield reliable estimations, the ex-Gaussian analyses included only the participants who produced 40 or more targets and 40 or more errors ($N = 53$). The data for these analyses were 14,037 observations (10,830 targets and 3,207 errors). Table 1 shows the raw count and percentage of the target and erroneous response types that were included in the ex-Gaussian analyses. We first estimated the parameters separately for each participant and Response Accuracy condition (target and error responses) using the QMPE software of Brown and Heathcote (2003). The average number of search iterations was 12.8. The extracted ex-Gaussian parameters were then submitted to three linear regression analyses (one per ex-Gaussian parameter). Each regression model included Response Accuracy (target vs. error; treatment-coded, errors as reference) as the main predictor.

Table 1

Raw Count and Percentage of Response Types Included in (a) the LMEM Analysis and (b) the ex-Gaussian analyses

Broad response type	Detailed response type	Raw count		Percentage (%)	
		LMEM	ex-Gaussian	LMEM	ex-Gaussian
Target responses	Correctly produced target word	17,154	10,830	81.1	77.2
Erroneous responses		4,003	3,207	18.9	22.8
Semantic		3,368	2,653	15.9	18.9
	Coordinate	2,384	1,859	11.3	13.2
	Superordinate	520	413	2.5	2.9
	Subordinate	6	4	0.03	0.03
	Associate	85	73	0.4	0.5
	Other semantic	218	189	1.0	1.4

	Part-whole	155	115	0.7	0.8
Other		635	554	3.0	3.9
	Unrelated	293	266	1.4	1.9
	Visual	211	193	1.0	1.4
	Phonological (word)	86	59	0.4	0.4
	Compound	39	30	0.2	0.2
	Morphological-semantic-phonological	6	6	0.03	0.04
Total		21,157	14,037	100	100

Note. LMEM = linear mixed effects model.

4.2 Subset analyses: Including only coordinate errors

The lexical errors included in the previous analysis (section 4.1) consisted of various types of errors (e.g., semantic coordinates, associates, visual, phonological etc.). To control for the type of produced error, we performed a subset analysis in which we explored whether there are systematic RT differences between target responses and coordinate errors, the most common error type in the present data (60% of the lexical errors of the LMEM analysis; 58% of the lexical errors of the ex-Gaussian analysis). Therefore, we conducted LMEM and ex-Gaussian analyses considering only coordinate errors, following the same approach described in section 4.1.

For the LMEM analysis, we included all participants (*n* = 80) and a subset of items (*n* = 204) from which both target responses and coordinate errors were elicited.[5] 13,792 data points (11,706 targets and 2,086 coordinate errors) entered the analysis.[6] The dependent variable was log-transformed RTs. As predictors, we included Response Type (target vs. coordinate error; treatment-coded, coordinate errors as reference). By-participant and by-item random intercepts and random slopes for Response Type were also included in the model. Additionally, we visually inspected the RT distributions for target responses and coordinate errors, as described previously. For the ex-Gaussian analyses, we included only the participants who produced 40 or more targets and 40 or more coordinate errors (*n* = 9). The remaining data for these analyses were 1,455 observations (1,038 target responses and 417 coordinate errors). The ex-Gaussian analyses were conducted using the same steps as those detailed in the preceding section (4.1). The average number of search iterations was 13.1. Given the small sample size, the results of the ex-Gaussian analyses should be interpreted with caution.

4.3 Comparing the Psycholinguistic Variables Between Targets and Coordinate Errors

In addition to the RT analyses, we inspected how coordinate errors and target responses differed in terms of various psycholinguistic variables that are known to influence word production (e.g., Alario et al., 2004; Perret & Bonin, 2019). Specifically, we considered the following psycholinguistic variables for each response: word frequency, word length (number of

---

[5] The type-token ratio for coordinate errors indicated variability in the number of unique errors produced per target item, ranging from 1 to 10 unique errors, with a median of 2 unique errors. This suggests that while some items elicited more alternative coordinate names, for most targets, the number of easily encoded alternatives was limited.

[6] The dataset of the RT analysis we report here is the same as the dataset of the analysis reported in section 4.3, where we include only coordinate errors whose information on their psycholinguistic properties is available. In addition, we ran the same RT analysis on all produced coordinate errors, irrespective of the availability of information on their psycholinguistic properties (n = 15,212 datapoints, consisting of 12,828 targets and 2,384 coordinate errors). The results of this analysis including a slightly larger dataset were comparable with the results reported in the section 5.2 ($\beta = -.02$, S.E. = .01, $t = -1.724$, $p = .09$).

phonemes), age of acquisition, concept familiarity and imageability. A subset of the items of the primary analyses was created that contained all coordinate errors for which information on the psycholinguistic variables was available both for the intended target responses and the produced coordinate errors (*n* = 2,086). More specifically, for targets and coordinate naming errors word length was obtained from McRae et al.'s (2005) database, and age of acquisition, concept familiarity and imageability were obtained from a norming study by Lampe and colleagues (2021)[7]. Finally, word frequency values were obtained from a (spoken) word frequency database based on television subtitles (SUBTLEX-UK; van Heuven et al., 2014). We inspected how target responses (*n* = 11,706) and coordinate errors (*n* = 2,086) differed in terms of the five psycholinguistic variables by performing Welch two-sample *t* tests.

5. Results and Discussion

Here, we first report the results of the primary analysis, which aimed to examine whether the RTs of errors systematically differed from the RTs of target responses (section 5.1), and the subset analysis, which aimed to further examine RT differences between target responses and coordinate errors, being the most common error type in the present data (section 5.2). Finally, we report the results of the analysis examining the relationship between target responses and coordinate errors in terms of their psycholinguistic properties (section 5.3).

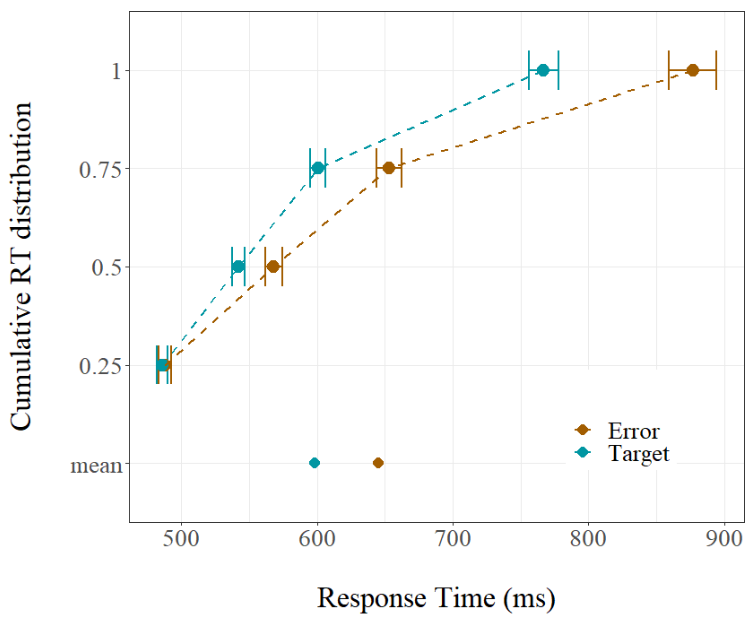
[7] In Lampe et al.'s (2021) norming study, participants rated the age of acquisition, familiarity, and imageability of items based on written words, without being shown corresponding pictures. Since these ratings were given independently of specific images, they are also deemed suitable for assessing the erroneous responses.

5.1 Primary Analysis Comparing RTs between Targets and Errors: LMEM and ex-Gaussian analyses

Figure 2 shows the cumulative RT distribution for the target responses and errors, with the bottom dots showing the mean RTs for each of the two response accuracy conditions. Table 2 summarizes the mean RTs and ex-Gaussian estimates per response accuracy condition. As is evident, mean naming latency was longer for the errors compared to target responses. The LMEM showed that the RT difference between error and target responses was statistically significant (β = -.031, *S.E.* = .008, *t* = -3.458, *p* < *.001*).

Figure 2

Cumulative RT Distribution for Target and Erroneous Responses



Note. Mean RT for each of the two response types is shown at the bottom of the figure. Error bars indicate the standard error for each RT bin. RT = response time; ms = millisecond.

Table 2

Mean Response Times and ex-Gaussian Parameter Estimates per Response Accuracy Condition Across Participants.

Response accuracy	*Mean*	μ	σ	τ
Target responses	598 (16)	489 (5)	34 (2)	98 (4)
Error responses	645 (21)	492 (7)	48 (5)	138 (9)

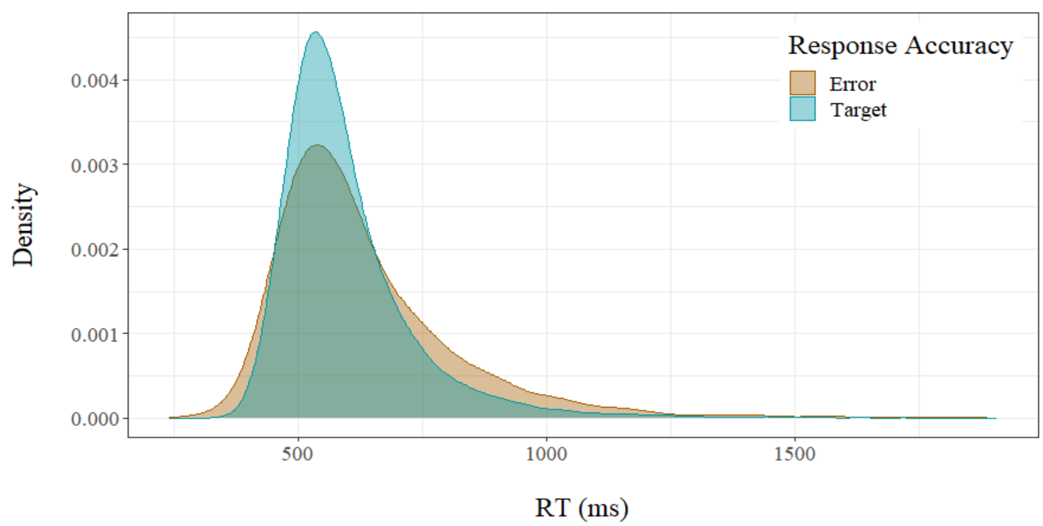
Note: M = mean response time; μ, σ, τ = ex-Gaussian parameter estimates. Mean response times and ex-Gaussian parameter estimates are given in milliseconds. Standard errors of the mean are shown in parentheses.

The results of the ex-Gaussian analysis revealed a statistically significant difference between target and erroneous responses for τ (β = -40.398, *S.E.* = 10.035, t = -4.026, $p < .001$) and σ (β = -14.163, *S.E.* = 5.825, t = -2.431, $p = .017$), but not for μ (β = -2.923, *S.E.* = 8.279, t = -.353, p = .725). Thus, the RT difference between target and errors was present mostly in the tail of the distribution consisting of the slowest responses rather than being a shift of the entire distribution. The effect on σ indicates that the distribution of RTs for target responses was less variable compared to the distribution of RTs for the erroneous responses.

In sum, lexical errors were produced more slowly relative to target responses. This difference was driven largely by only a part of the RT distribution, that is, the tail. The findings suggest a mixed scenario, where lexical errors were produced more slowly relative to the target responses in the bin of the slowest responses (i.e., the tail of the distribution), but their overall RT difference is not significant in the normal (i.e., Gaussian) part of the distribution. A closer look at the RT distribution for target and error responses suggests that the observed tail effect is driven by an over-representation of errors relative to the target responses in the tail of the RT distribution (Figure 3). In other words, there are more extremely slow responses in the error RT distribution than in the target RT distribution, thus yielding a higher mean RT for the slowest erroneous responses. The larger variation in the RT of the slow erroneous responses relative to that of the slow correct responses is also captured in the error bars of the slowest RT bin (Figure 2).

Figure 3

RT Distributions for Target and Error responses



Note. RT = response time; ms = millisecond.

5.2 Subset Analysis Comparing RTs of Target Responses and Coordinate Errors

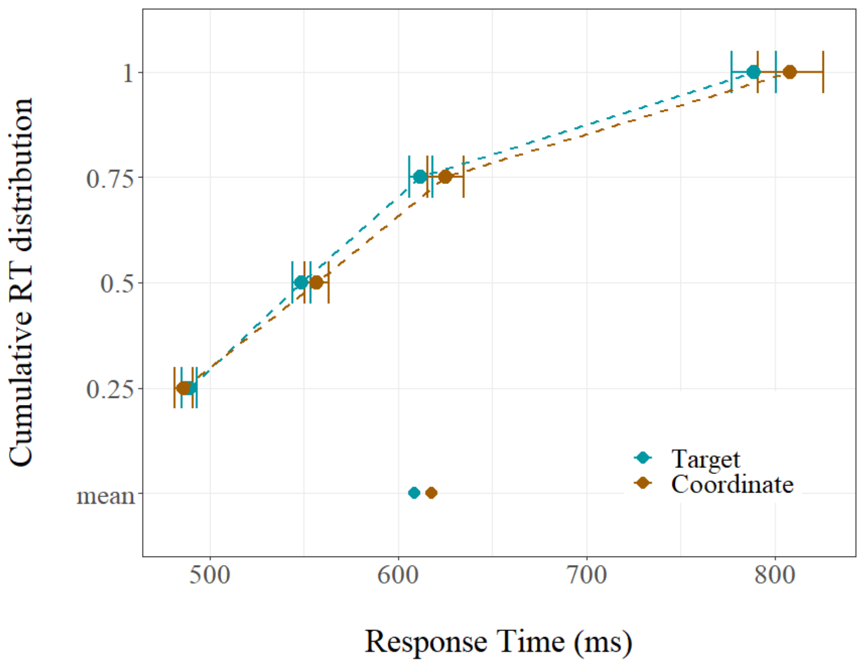
Figure 4 shows the cumulative RT distribution and means for target responses and coordinate errors. A detailed summary of mean RTs and ex-Gaussian estimates for each response type is shown in Table 3.

Visual inspection of the RT distributions for target responses and coordinate errors suggests that RTs for coordinates and targets are similarly distributed in the fastest bin and slightly diverge as responses get slower. Yet, this divergence does not seem to be substantial.

The LMEM showed no significant RT difference between target responses and coordinate errors (β = .002, *S.E.* =.01, t = .181, p = .856). The ex-Gaussian analyses revealed no significant RT difference between target responses and coordinate errors in either of the three parameters (μ: β = - 21.28, *S.E.* =19.65, t = -1.083, p = .295; σ: β = -7.808, *S.E.* =13.508, t = -.578, p = .571; τ: β = 42.13, *S.E.* =21.94, t = 1.920, p = .07).

Figure 4

Cumulative RT Distribution for Target Responses and Coordinate Errors



Note. Mean RT for each of the two response types is shown at the bottom of the figure. Error bars indicate the standard error for each RT bin. RT = response time; ms = millisecond.

Table 3

Mean Response Times and ex-Gaussian Parameter Estimates per Response Type Across Participants.

Response type	*M*	μ	σ	τ
Target Responses	609 (17)	471 (13)	34 (9)	98 (15)
Coordinate Errors	618 (18)	493 (15)	42 (10)	56 (16)

Note: M = mean response time; μ, σ, τ = ex-Gaussian parameter estimates. Mean response times and ex-Gaussian parameter estimates are given in milliseconds. Standard errors of the mean are shown in parentheses.

In this analysis we focused on coordinate errors, i.e., the most common lexical errors in the present dataset, and examined why these lexical errors arise, by comparing their RTs and word-intrinsic properties to those of target responses. The RT analyses showed that the naming latencies of coordinate errors do not significantly differ from the naming latencies of target responses throughout the RT distribution.[8] This suggests that coordinate errors arise due to premature selection rather than attentional disengagement. The present RT results show a different pattern from the results in the primary analysis suggesting that the tail effect we observed is driven from errors other than the coordinates. In order to confirm this assumption, we conducted a post-hoc analysis contrasting the RTs of coordinates, targets and all other errors. This analysis is reported in the next section.

5.2.1 Post-hoc Analysis: Comparing RTs of target responses, semantic coordinates and all other errors

In our primary analysis (section 5.1), we found a tail effect when comparing the RTs of targets and errors, with errors being as fast as targets in the normal part of the distribution, and slower than targets in the tail. In the analyses reported in section 5.2, we focused on the most common

---

[8] This finding remained consistent when controlling for participants' familiarity with the target name of the depicted item and responses driven from difficulties in the visual identification of the items, by comparing target responses and coordinate errors on items that had been named correctly in the previous standard picture naming round (see Appendix B for an exploratory analysis showing that RTs of targets and error responses in the speeded naming task were differentially affected by whether they matched or mismatched the responses in the previous standard naming round). Specifically, in this subset of data (4,081 targets and 617 coordinate errors) we found that RTs of coordinates and targets overlap, as shown in their mean latency (targets: *M* = 614 ms, *SE* = 17 ms; coordinates: *M* = 614 ms, *SE* = 16 ms) and their RT distribution (see Figure C1 in Appendix C).

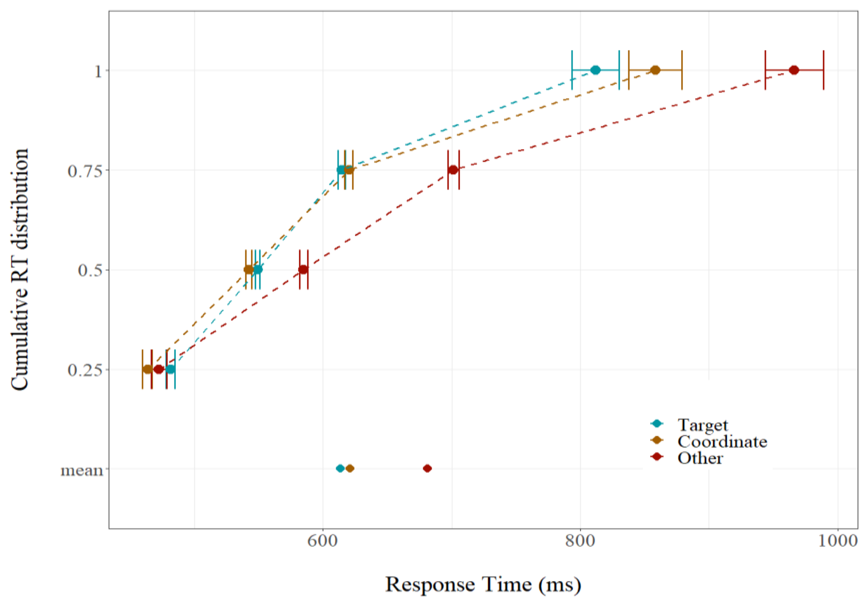
type of produced errors, i.e., the semantic coordinates, and we found that they show no significant RT differences from the target responses throughout the distribution In light of the findings in sections 5.1 and 5.2, this analysis aims to investigate whether the tail effect in 5.1 is indeed driven by errors other than the coordinates, by comparing the RTs of targets, semantic coordinates and all other (semantically related or unrelated) errors.

Similarly to the previous RT analyses, we performed a LMEM to compare the RTs for target responses, semantic coordinates, and other related and unrelated errors. We included only participants (*n* = 80) and items (*n* = 198) from which target responses, coordinates and other errors were elicited. This led to the inclusion of 14,553 data points (11,257 targets, 2,135 semantic coordinates, and 1,161 other related and unrelated errors). Of the 1,161 other errors, 685 were semantically related (but not coordinate) errors and 476 were unrelated errors. The model included log-transformed RTs as the dependent variable and Response Type (target vs. coordinate vs. other error; treatment-coded, coordinate errors as reference) as the predictor. By-participant and by-item random intercepts and random slopes for Response Type were also included in the model. Moreover, we visually examined the RT distributions for target responses, semantic coordinates, and other errors. Note that RTs could not be rank-ordered and averaged for each participant first, as not all participants had sufficient data points for the other errors. It was not possible to perform ex-Gaussian analyses, as there were no participants who produced 40 or more responses corresponding to all three response types.

Figure 5 shows the cumulative RT distribution and means for target responses, semantic coordinates, and other errors. As the figure shows, the RTs for targets and coordinate errors show overlap in the second and third bins, and a slight divergence in the fastest bin, with coordinates being faster than the targets, and in the slowest bin, with coordinates being slower than the targets. The RTs of all other errors were much slower than both targets and coordinates in all bins except for the fastest bin. This observation was also reflected statistically in LMEM, with the overall RTs of coordinate errors (*M* = 621 ms, *SE* = 20 ms) being not significantly different from the RTs of target responses (*M* = 613 ms, *SE* = 17 ms; β = -.02, S.E. = .01, *t* = -1.603, *p* = .11), but being significantly faster than other errors (*M* = 681 ms, *SE* = 24 ms; β = .05, S.E. = .01, *t* = 3.717, *p* < .01).

Figure 5

Cumulative RT Distribution for Target Responses, Coordinates and all Other (Related or Unrelated) Errors



Note. Mean RT for each of the two response types is shown at the bottom of the figure. Error bars indicate the standard error for each RT bin. RT = response time; ms = millisecond.

These findings suggest that the tail effect observed in the primary analysis (section 5.1), where all errors were included, was not driven by coordinate errors (as was also suggested in the analysis on coordinate errors only, section 5.2), but rather by all other types of (semantically related or unrelated) errors included in the analysis.

5.3 Comparing the Psycholinguistic Variables Between Targets and Coordinate Errors

Table 4 summarizes the mean values of each psycholinguistic variable for target responses and coordinate errors. The *t*-tests showed that target responses and coordinate errors significantly differed in their word frequency ($t(2725.5) = -12.708$, $p < .001$), word length ($t(3050.4) = 11.187$, $p < .001$, and age of acquisition ($t(3061.1) = 6.165$, $p < .001$), with coordinate errors being more frequent, shorter and acquired earlier than target words[9]. These findings suggest that coordinate errors are selected over target names, especially under time constraints, due to their greater accessibility.

Table 4

Psycholinguistic Properties for Target Responses and Coordinate Errors.

Response type	*Word frequency*	*Word length*	*Age of acquisition*	*Familiarity*	*Imageability*
Target responses	4.01(0.56)	4.68 (1.72)	3.19 (1.02)	5.94 (1.90)	5.42 (0.74)

[9] These effects were observed despite the fact that there were on average more items in the McRae et al. (2005) database that (1) were equally or less frequent than the target word (on average, 59% of items were equally or less frequent than the target word and 41 % were more frequent than the target), (2) were equally long or longer in word length than the target word (on average, 61% of items were equally long or longer than the target word and 39% were shorter than the target), and (3) were acquired as late as or later than the target word (on average, 59% of items were acquired as late as or later than the target word and 41% were acquired earlier than the target word). Given these percentages, if coordinate errors were selected by chance, we would have expected to observe the opposite effects or no effects.

Coordinate errors	4.19 (0.62)	4.26 (1.57)	3.05 (0.92)	5.92 (1.92)	5.45 (0.81)
Mean difference	0.18*	-0.42*	-0.14*	-0.02	0.03

Note: Mean values for each psycholinguistic property of target responses and coordinate errors. Standard deviations are shown in parentheses. The “Mean difference” row represents the difference between Target responses and Coordinate errors for each variable. Asterisks (*) indicate psycholinguistic variables for which there is a significant difference between target responses and coordinate errors.

6. Discussion

In the present study, we compared the temporal characteristics of lexical errors and correct responses with the aim of assessing whether lexical errors result from a hasty selection due to a premature decision to speak (*premature selection hypothesis)* or due to momentary lapses of attention (*attention lapse hypothesis*). On the one hand, the premature selection hypothesis predicts that errors should be at least as fast as targets. On the other hand, the attention lapse hypothesis predicts that errors should be slower than target responses across the entire distribution. To this end, we analyzed data from a speeded picture naming task, where participants were presented with pictures and were asked to name them as fast as possible and ideally within 600ms after picture onset (Lampe et al., 2023). In our study, we used linear mixed effects models and ex-Gaussian analyses to investigate potential systematic differences in naming speed between erroneous responses and target responses. By doing so, we aimed to gain a deeper understanding of the temporal dynamics underlying error production under time constraints, while also shedding light on the broader mechanisms governing word selection.

Table 5 summarizes the results of the primary analyses including all lexical errors (grouped together), the subset analyses including coordinate errors only, and the post-hoc analysis including the coordinates and all other (semantically related and unrelated) errors separately. The results of the primary analysis showed that production of lexical errors was overall slower than the production of targets. This finding was nuanced by subsequent ex-Gaussian analyses, the results of which gave support to a mixed pattern in the RT distribution of errors in relation to targets. Specifically, lexical errors were produced more slowly compared to the target words in the tail of the RT distribution (effect on τ), but not in the normal (Gaussian) part of the distribution (effect on μ), as shown by the findings of the ex-Gaussian analysis. The tail effect observed in the cumulative RT distribution and the ex-Gaussian analysis is driven by an asymmetric representation of target and erroneous responses in the tail of the distribution. More specifically, the τ parameter, which captures the mean and standard deviation of the slow responses (i.e., tail of the distribution), was larger for errors than for targets. This suggests that there were more slow erroneous responses compared to slow target responses, as shown in Figure 3. Moreover, RTs for the slowest erroneous responses showed a greater variability relative to the RTs for the slowest target responses (as shown in Figure 2). Finally, we found an effect on the σ parameter, accounting for the standard deviation of the normal part of the distribution, with RTs of errors showing more variability relative to RTs of target responses. In light of the two hypotheses, these findings suggest, at first sight, that fast errors are associated with a hasty and incorrect selection, in line with the *premature selection hypothesis*, while slow errors potentially occur due to momentary attentional disengagement from the task goal, in line with the *attention lapse hypothesis*.

In the subset analyses (section 5.2), we focused on coordinate errors, i.e., the most common errors in the data, and examined whether and how they differed from target responses with respect to their naming latencies. We found that RTs of coordinate errors did not significantly differ from target responses throughout the entire distribution (see also Figure 4). This finding suggests that coordinate errors arise due to premature selection and do not follow the mixed RT pattern we observed in the primary analysis. In order to assess whether the tail effect of the primary analysis is driven by the other produced errors rather than the smaller sample of the data, we performed a post-hoc analysis of the complete dataset (similarly to the primary analysis) comparing the RTs of targets, coordinate errors and all other (related and unrelated) errors (section 5.2.1). The cumulative RT distribution showed a divergence of other errors from both target and coordinate errors throughout most of the distribution (see Figure 5) and the LMEM results showed that other errors were produced significantly slower than coordinates, whose RTs did not significantly differ from those of target responses. Taken together, these results suggest that the tail effect observed in the primary analysis, where all errors were included, was primarily driven by these other errors.

Table 5

Summarized findings of all analyses about differences in RTs (and word-intrinsic properties) between targets and lexical errors

Analyses			No. of errors	No. of targets	Results
Primary RT analyses	all errors vs. targets (section 5.1)	LMEM	4,003	17,154	RT errors > RT targets
		ex-Gaussian	3,207	10,830	μ errors = μ targets σ errors > σ targets τ errors > τ targets
Subset RT analyses	coordinate errors vs. targets (section 5.2)	LMEM	2,086	11,706	RT coordinates = RT targets
		ex-Gaussian	417	1,038	μ errors = μ targets σ errors = σ targets τ errors = τ targets

Post-hoc RT analysis	coordinate errors vs. targets vs. all other errors (section 5.2.1)	LMEM	2,135 coordinates 1,161 other	11,257	RT coordinates = RT targets RT coordinates < RT other
Analyses in psycholinguistic properties	coordinate errors vs. targets (section 5.3)	*t*-test	2,086	11,706	frequency: errors > targets length: errors < targets AoA: errors < targets

Note: LMEM = linear mixed-effects model; RT = response time; μ = mean of the Gaussian component (normal part of distribution); σ = standard deviation of the Gaussian component; τ = mean and standard deviation of the exponential component (tail of the distribution); frequency = log word frequency; length = word length in number of phonemes; AoA = age of acquisition; The symbols '>' and '<' indicate significantly greater than and significantly lower than respectively. The symbol '=' indicates no significant difference between the values of the two compared response types.

The present results suggest that different mechanisms lead to different types of errors, which is behaviorally captured in their timing. Coordinate errors were produced as fast as targets throughout the RT distribution, which suggests that they arise from a premature selection but in the context of intact attentional control. By contrast, the other (semantically related or unrelated) errors were overall produced more slowly than coordinate errors (and thus targets). However, a visual inspection of their RT distribution shows a rather mixed pattern, in which other errors substantially diverge from coordinates and targets in most of the distribution except for the bin with the fastest responses (see Figure 5). This suggests that, while a subset of these errors (i.e., the fastest ones) may also arise due to premature selection, this is not the case for the errors in the other three bins. For those errors, the results are more in line with the attention lapse hypothesis.

Of these other errors, the majority of the responses that were semantically related to the target word were superordinates (e.g., bird for seagull) and other responses with semantic relationship to the target (e.g., staircase for basement). The majority of the semantically unrelated errors were completely unrelated (e.g., pelican for lime) or visually related to the target (e.g., stick for cigar). Given the different types of errors included in this broad category, we can only speculate about the causes that might lead to the production of these errors. One possibility is that difficulties in the visual processing of the pictures (e.g., in the cases where they produced visual errors) prolonged the time it took participants to name the items, and were eventually named erroneously. If these errors result from momentary lapses of attention, these lapses could allow activation to spread not only to semantically related but also unrelated representations, e.g., if they were (co-)activated earlier in the experiment. As a result, selecting among co-activated representations becomes more time-consuming, sometimes leading to the production of words that are neither the intended target nor semantically related. Another potential explanation for these findings could relate to the operation of the self-monitoring system following the selection of these representations. This aspect will be further explored in the discussion on coordinate errors below. Nevertheless, given the small number of the other errors, it was not possible to delve into their RT distribution and further dissect them in order to pinpoint exactly which of these scenarios might drive the observed effects.

Focusing on coordinate errors, which are thought to be co-activated along with the target, the lack of RT difference from the target responses throughout the distribution suggests that, when under time pressure, using a non-target label of the same semantic category (e.g., goose) to name an object (e.g., duck) does not inherently require more cognitive processing time compared to using the correct label (i.e., duck). This can be explained by a race-to-threshold process where both the target word and the coordinate error are activated and race for selection. Depending on their activation levels on each trial, the target or the coordinate reaches the selection threshold first, sometimes quickly (resulting in fast responses) and other times more slowly (resulting in slow responses). As previously mentioned, this finding is in line with the premature selection hypothesis, according to which the speaker is attentive to the task goal (to name the picture fast) but makes a hasty (and often erroneous) word selection.

A premature semantic error can be explained both by competitive and non-competitive models of lexical access. Both models attribute coordinate errors to the way in which activation spreads through the representations, with a word other than the target being favored for selection. In the WEAVER++ model (Roelofs, 2023b), this could be explained if activation of the target's word form is not sufficient to allow its selection, and verification is skipped. Alternatively, in Oppenheim et al.'s (2010) non-competitive model, this could be explained by changes in the semantic-to-lexical links of co-activated representations following lexical selection, in which the link of the selected representation is strengthened and that of the unselected representations is weakened. If a previously co-activated but unselected representation becomes the target later on, its retrieval may be hindered by its weakened semantic-to-lexical links, resulting in the eventual selection of a co-activated representation with stronger links (e.g., a previously selected target).

Along with the RT differences between coordinate errors and target responses, differences in their word-intrinsic properties may explain why these errors were selected over the targets. In the present study, we found that coordinate errors had an overall advantage over targets in a number of psycholinguistic variables that are found to influence naming performance (Perret & Bonin, 2019), with word frequency, number of phonemes, and age of acquisition showing a significant effect (see section 5.3). The results are in line with previous work examining the target-error relation in their word-intrinsic properties (e.g., Del Viso et al., 1991; Kittredge et al., 2008). This finding, in combination with the conjectured locus of these variables in lexical access (e.g., Jescheniak & Levelt, 1994; Kittredge et al., 2008) is consistent with the proposal that semantically related errors arise when their word form is more accessible than that of the target word (Roelofs, 2023b). In addition, using the same dataset as this study to investigate the target-error relation in various semantic variables, Lampe et al. (2024) found that coordinate errors were overall more typical category representatives, had more semantic features than the target words, and were largely near semantic neighbors of the targets (though not the closest ones compared to other near semantic neighbors). Lampe and colleagues (2024) suggested that these properties lead to stronger activation of the semantic representations of erroneous words, making them more accessible, and thus more likely to be selected, explaining their frequent selection over the target and their predominance over other errors in the present dataset.

In addition, the present results inform us about the way in which self-monitoring operates. More specifically, the higher number of coordinate errors compared to other errors suggests a monitoring system that is more “lenient” when it comes to selected words that belong to the same semantic category as the target compared to categorically unrelated words. This finding can be explained by the comprehension-based monitoring model (Levelt, 1983; Roelofs, 2005), according to which, the comparison of the planned word to the intended message leads to less successful error detection when the planned word is semantically related to the target word and speaking takes place under time pressure. With respect to the conflict-based monitoring model (Nozari et al., 2011; Nozari & Hepner, 2019), the predominance of categorically related errors over other errors could be explained by considering their strong connection in the semantic-lexical space and a flexible conflict threshold that allows these errors to go undetected given the task demands (e.g., speed over accuracy). However, two questions remain unanswered: (1) Why would the monitor fail to detect coordinate errors more frequently than other errors that may also be semantically related? (2) Why would these other errors be produced more slowly than the coordinates? One possibility is that self-monitoring may operate in such a way that the connection strength (i.e., semantic similarity) between the most highly activated erroneous representation and the target plays a critical role within the monitoring process. The stronger the connection is (e.g., in terms of semantic properties), the more likely the erroneously selected word will go unnoticed by the monitor. In other words, if the selected word is a “good-enough” approximation of the speaker’s message, it is more likely to be “approved” by the monitoring system relative to a representation that is semantically more distant to the concept. In contrast, an erroneously selected word that is semantically more distant (e.g., an unrelated word) is more likely to be detected as an error by the monitor, leading to an attempt to repair this selection. This attempt may take time and may not be successful, resulting in the late production of the erroneous word that was initially selected. This proposal highlights the role of the monitoring system in lexical selection (Dhooge & Hartsuiker, 2012), and encourages further research to better understand whether and how factors like semantic similarity affect speakers’ word choices and naming latencies.

A potential limitation of our study is the influence of prior production on naming responses due to the inclusion of earlier naming rounds. Participants named the same items in one or two standard naming rounds before the speeded naming round, which could lead to perseveration errors and impact the overall latencies of errors. We attempted to control for prior production effects on the latencies of coordinate errors (being the most common errors), by comparing a subset of target responses and coordinates on items that had only been named correctly in the previous standard naming round. This approach excluded coordinate responses that had been repeated across the standard and speeded naming rounds by the same participant for the same item. However, we acknowledge that perseverations may not arise solely from responses to the same item in the task. While it is indeed the case that the majority of coordinate errors correspond to names that were previously produced by the same participant (as correct or incorrect responses), this was also the case for the majority of target responses. In fact, of the 11,706 target responses included in the subset analysis (reported in sections 4.2 and 5.2), 94% (n= 11,033) were previously produced and 6% (n= 673) were not previously produced. Similarly, of the 2,086 coordinate errors, 94% (n= 1958) were previously produced and 6% (n = 128) were not previously produced. The likelihood of a name being previously produced is numerically the same for both targets and coordinate errors, and therefore repetition cannot have caused most of the errors. In addition, the small proportion of non-previously produced responses (6%) likely stems from different underlying causes, as the not previously produced coordinates were "new" errors, whereas the not previously produced targets were corrections of errors in the previous rounds (see also discussion in Appendix B). Therefore, any observed differences in RTs between targets and coordinate errors in this small dataset are difficult to interpret meaningfully. Ideally, future studies should aim to control for repetition effects by placing the speeded naming round first, eliminating prior naming rounds that may introduce perseveration effects. However, given that both target responses and coordinate errors were similarly affected by repetition in our study, we believe that our main findings remain robust and that our research question can be addressed with the current data.

The findings of the present study add to the growing body of research that examines target-error relations (e.g., Lampe et al., 2024) as well as the notion of "good enough" production, according to which speakers implicitly weigh accessibility difficulty and message alignment factors in their lexical choices (e.g., Goldberg & Ferreira, 2022; Koranda et al., 2022; Lee et al., 2022). Recent studies have explored whether speakers' word choices are guided solely by message alignment, that is, selecting a word that best fits the speaker's intended message, or whether word selection follows a rather probabilistic approach, that is, selecting a word that is more accessible without being the optimal choice for the intended message (e.g., Jacobs & MacDonald, 2023; Koranda et al., 2022; Lee et al., 2022). Koranda et al. (2022), for instance, using an artificial language production task, found that, under time pressure, word selection is not deterministically driven by message alignment, but instead it can settle for utterances that are "good enough" for current communicative purposes, representing a trade-off between clarity and ease. The results of the current study also suggest that, under the constraint of time, the production system does not strive to find the best word but may opt for words that, while being errors from a semantic perspective, are considered "good enough" approximations of the intended message. These words most commonly correspond to items that are members of the same semantic category as the target. Carota et al. (2022, 2023) showed that information about an item's semantic category is accessed very early during the word planning process, specifically within 150–200 ms post picture presentation. This early categorization might explain why people often choose words from the same category as the target, rather than other semantically related or completely unrelated words. Words in the same category as the target likely receive strong activation early on via their shared semantic features with the target. Under time pressure, and when context allows, the system may overlook semantic inaccuracies and select the most readily accessible word that fits the category, even if it does not optimally express the intended message.

The present study also adds to the literature that highlights the complex interaction between executive control (i.e., attention) and language-specific processes in word production (e.g., San José et al., 2021; Todorova et al., 2020) and extends this to erroneous word production in neurotypical speakers (for aphasic speakers, see also Roelofs, 2023b). Specifically, we suggest that when the executive control (i.e., attention to producing a word) operates efficiently, the selected errors are categorically related to the target, and lexical selection does not require additional time to complete compared to targets. In contrast, compromised executive control (e.g., when speakers experience attentional lapses) leads to semantically unrelated errors, the selection of which is time consuming. Therefore, the interplay between linguistic and executive function is crucial, as it does not only account for when speakers produce a word, but also what word they produce.

Finally, our findings suggest that existing models that explain the timing of correct responses (at least in time-constraining contexts) could also be applicable to coordinate errors, given that they seem to follow the same behavioral patterns, both in timing and in the way in which word-intrinsic properties influence their selection. There is no need to assume distinct mechanisms to explain their selection and/or timing. The results of the present study do not allow for a discrimination between the different word production models (e.g., competitive vs. non-competitive models of lexical selection). Yet, by looking at both correct and incorrect word choices together, this study offers a better understanding into the complex processes underlying word production, and advocates for a unified approach of how we select words, be it correct ones or semantically close incorrect ones.

In sum, this research aimed to better understand the origin of lexical errors by assessing whether they arise from a hasty selection and premature decision to speak (*premature selection hypothesis)* or from momentary attentional disengagement from the task (*attentional lapse hypothesis).* We found that lexical errors were slower than targets in the tail, but not in the normal part of the RT distribution, with the tail effect primarily resulting from errors that were not semantic coordinates, i.e., members of the target's semantic category. We therefore concluded that coordinate errors occur due to a premature selection but intact attentional control, following the same lexical constraints as targets, while other errors, given the variability in their nature, may vary in their origin, with one potential source being lapses of attention.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*(5), 713–734. https://doi.org/10.1080/01690960802597250

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. https://doi.org/10.3758/BF03195559

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495–523. https://doi.org/10.1016/j.jml.2007.10.004

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

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. The Quarterly Journal of Experimental Psychology Section A, 58(4), 667–692. https://doi.org/10.1080/02724980443000142

Best, W. (1996). When Racquets Are Baskets But Baskets Are Biscuits, Where Do the Words Come From ? A Single Case Study of Formal Paraphasic Errors in Aphasia. *Cognitive Neuropsychology*, *13*(3), 443–480. https://doi.org/10.1080/026432996381971

Boersma, P., & Weenink, D. (2022). *Praat: Doing Phonetics by Computer [Computer program]*. *(Version 6.1.38).* http://www.praat.org/

Bormann, T., Kulke, F., & Blanken, G. (2008). The influence of word frequency on semantic word substitutions in aphasic naming. *Aphasiology*, *22*(12), 1313–1320. https://doi.org/10.1080/02687030701679436

Brown, S., & Heathcote, A. (2003). QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 485–492. https://doi.org/10.3758/BF03195527

Carota, F., Schoffelen, J.-M., Oostenveld, R., & Indefrey, P. (2022). The Time Course of Language Production as Revealed by Pattern Classification of MEG Sensor Data. Journal of Neuroscience, 42(29), 5745–5754. https://doi.org/10.1523/JNEUROSCI.1923-21.2022

Carota, F., Schoffelen, J.-M., Oostenveld, R., & Indefrey, P. (2023). Parallel or sequential? Decoding conceptual and phonological/phonetic information from MEG signals during language production. Cognitive Neuropsychology, 40(5–6), 298–317. https://doi.org/10.1080/02643294.2023.2283239

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(2), 345–361. https://doi.org/10.1037/0278-7393.25.2.345

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. Cognition, 81(3), B77–B86. https://doi.org/10.1016/S0010-0277(01)00135-4

Del Viso, S., Igoa, J. M., & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*(3), 161–185. https://doi.org/10.1007/BF01067213

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production.

Psychological Review, *93*(3), 283. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. (1990). Effects of Frequency and Vocabulary Type on Phonological Speech Errors. *Language and Cognitive Processes*, *5*(4), 313–349. https://doi.org/10.1080/01690969008407066

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. Psychological Review, 104(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dhooge, E., & Hartsuiker, R. J. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language*, *66*(1), 163–176. https://doi.org/10.1016/j.jml.2011.08.004

Fieder, N., Wartenburger, I., & Abdel Rahman, R. (2019). A close call: Interference from semantic neighbourhood density and similarity in language production. *Memory & Cognition*, *47*(1), 145–168. https://doi.org/10.3758/s13421-018-0856-y

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a New Model of Verbal Monitoring. *Journal of Cognition*, *3*(1), 17. https://doi.org/10.5334/joc.81

Gerhand, S., & Barry, C. (2000). When Does a Deep Dyslexic Make a Semantic Error? The Roles of Age-of-Acquisition, Concreteness, and Frequency. *Brain and Language*, *74*(1), 26–47. https://doi.org/10.1006/brln.2000.2320

Glaser, W. R., & Düngelhoff, F.-J. (1984). The Time Course of Picture-Word Interference. Journal of Experimental Psychology: Human Perception and Performance, 10(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, S1364661322000201. https://doi.org/10.1016/j.tics.2022.01.005

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, *82*(2), 113–145. https://doi.org/10.1016/S0093-934X(02)00001-9

Harley, T. A., & MacAndrew, S. B. G. (2001). Constraints Upon Word Substitution Speech Errors. *Journal of Psycholinguistic Research*, *30*(4), 395–418. https://doi.org/10.1023/A:1010421724343

Hodgson, C., & Lambon Ralph, M. A. (2008). Mimicking aphasic semantic errors in normal speech production: Evidence from a novel experimental paradigm. Brain and Language, 104(1), 89–101. https://doi.org/10.1016/j.bandl.2007.03.007

Jacobs, C. L., & MacDonald, M. C. (2023). A chimpanzee by any other name: The contributions of utterance context and information density on word choice. *Cognition*, *230*, 105265. https://doi.org/10.1016/j.cognition.2022.105265

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://psycnet.apa.org/doi/10.1037/0278-7393.20.4.824

Jongman, S. R. (2017). Sustained Attention Ability Affects Simple Picture Naming. *Collabra: Psychology*, *3*(1), 17. https://doi.org/10.1525/collabra.84

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. https://doi.org/10.1080/17470218.2014.964736

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive*

Neuropsychology, *25*(4), 463–492. https://doi.org/10.1080/02643290701674851

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. https://doi.org/10.1177/09567976221089603

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lampe, L. F., Hameau, D. S., Bürki, A., Fieder, N., Sowman, P. F., & Nickels, L. (2019). Investigating effects of semantic variables on word production. Preregistration on the Open Science Framework. Osf.io/yw6ma. Preregistration on the Open Science Framework. Osf.Io/Yw6ma. https://osf.io/yw6ma/

Lampe, L. F., Hameau, S., & Nickels, L. (2021). Semantic variables both help and hinder word production: Behavioral evidence from picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(1), 72. https://doi.org/10.1037/xlm0001050

Lampe, L. F., Hameau, S., & Nickels, L. (2023). Are they really stronger? Comparing effects of semantic variables in speeded deadline and standard picture naming. *Quarterly Journal of Experimental Psychology*, *76*(4), 762–782. https://doi.org/10.1177/17470218221103356

Lampe, L. F., Zarifyan, M., Hameau, S., & Nickels, L. (2024). Why is a flamingo named as pelican and asparagus as celery? Understanding the relationship between targets and errors in a speeded picture naming task. *Cognitive Neuropsychology*, *0*(0), 1–33. https://doi.org/10.1080/02643294.2024.2315822

Lee, C., Lew-Williams, C., & Goldberg, A. E. (2022). Accessibility factors that lead to good-enough language production. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*, 8. https://escholarship.org/uc/item/828625bm

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01). https://doi.org/10.1017/S0140525X99001776

Lloyd-Jones, T. J., & Nettlemill, M. (2007). Sources of error in picture naming under time pressure. *Memory & Cognition*, *35*(4), 816–836. https://doi.org/10.3758/BF03193317

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. Cognitive, Affective, & Behavioral Neuroscience, 11(1), 32–43. https://doi.org/10.3758/s13415-010-0009-7

Moses, M. S., Nickels, L. A., & Sheard, C. (2004). "I'm sitting here feeling aphasic!" A study of recurrent perseverative errors elicited in unimpaired speakers. Brain and Language, 89(1), 157–173. https://doi.org/10.1016/S0093-934X(03)00364-X

Noteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 87–95). Academic Press Inc.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. https://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., & Hepner, C. R. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, *36*(5–6), 193–207. https://doi.org/10.1080/02643294.2018.1476335

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

Oppenheim, G., & Nozari, N. (2024). Similarity-induced interference or facilitation in language production reflects representation, not selection. *Cognition*, *245*, 105720. https://doi.org/10.1016/j.cognition.2024.105720

Papoutsi, C., Tourtouri, E., Piai, V., Lampe, L. F., & Meyer, A. (2024). *Fast and slow errors: What naming latencies of errors reveal about the interplay of attentional control and word planning in speeded picture naming*. Open Science Framwork. https://osf.io/s4eq6/

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*(6), 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461. https://doi.org/10.1037/0033-2909.86.3.446

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (2004). Error Biases in Spoken Word Planning and Monitoring by Aphasic and Nonaphasic Speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, *111*(2), 561. https://doi.org/10.1037/0033-295X.111.2.561

Roelofs, A. (2005). Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 42–63). Hove: Psychology press. https://doi.org/10.1016/j.cognition.2004.07.001

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 353–368. https://doi.org/10.1037/0278-7393.34.2.353

Roelofs, A. (2023a). Modeling the attentional control of vocal utterances: From Wernicke to WEAVER++/ARC. In *The Routledge International Handbook of Psycholinguistic and Cognitive Processes* (2nd ed.). Routledge.

Roelofs, A. (2023b). What semantic errors restricted to either speaking or writing in aphasia tell us about lemmas. *Language, Cognition and Neuroscience*, *38*(3), 277–287. https://doi.org/10.1080/23273798.2022.2108548

San José, A., Roelofs, A., & Meyer, A. S. (2021). Modeling the distributional dynamics of attention and semantic interference in word production. *Cognition*, *211*, 104636. https://doi.org/10.1016/j.cognition.2021.104636

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. https://doi.org/10.1016/0749-596X(90)90011-N

Todorova, L., Neville, D. A., & Piai, V. (2020). Lexical-semantic and executive deficits revealed by computational modelling: A drift diffusion model perspective. *Neuropsychologia*, *146*, 107560. https://doi.org/10.1016/j.neuropsychologia.2020.107560 van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vitevitch, M. S. (1997). The Neighborhood Characteristics of Malapropisms. *Language and Speech*, *40*(3), 211–228. https://doi.org/10.1177/002383099704000301

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(4), 664–680. https://doi.org/10.1037/0278-7393.17.4.664

Vitkovitch, M., Humphreys, G. W., & Lloyd-Jones, T. J. (1993). On naming a giraffe a zebra: Picture naming errors across different object categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2), 243–259. https://doi.org/10.1037/0278-7393.19.2.243

Appendix A

Response Classification

Table A1

Response Classification With Examples

Response accuracy	Broad response type	Detailed response type	Example
Targets	Correct	Correctly produced target word	"Avocado" for *avocado*
		Correct response with determiner	"An anchor" for *anchor*
		Disfluency on the initial phoneme/syllable	"ho hose" for *hose*
		Hesitation	"Er… bat" for *bat*
		Elaboration	"Baseball bat" for *bat*
		Abbreviation	"croc" for *crocodile*
		Synonym	"Trousers" for *pants*

Errors			
	Semantic	Coordinate (category-based)	"Goose" for *duck*
		Superordinate (category-based)	"Bird" for *seagull*
		Subordinate (category-based)	"Pound" for *coin*
		Associate	"Camp" for *tent*
		Other response with semantic connection to target	"Staircase" for *basement*
		Part-whole	"Frame" for *mirror*
		Incomplete response	"Telesc" for *microscope*
		Semantic-then-phonological (words and nonwords)	"Plute" for *clarinet* (mediated via flute)
	Other	Unrelated	"Umbrella" for *celery*
		Visual	"Stick" for *cigar*

	Nonword	"Kersh" for *prawn*
	Phonological (word or nonword)	"Balloon" for *spoon* "Prana" for *banana*
	Compound	"Pipe" for *bagpipe*
	False start on unrelated phoneme	"uuu zh" for *slippers*
	False start on target related phoneme	"Tr" for *tricycle*
	Disfluency with self-correction	"P kettle" for *kettle*
	Morphological-semantic-phonological	"Toast" for *toaster*
Omissions	Omission	NA
	Expression of failure to respond	"Thing" for *turtle*

Note. NA = not applicable. In the "Example" column, participant responses are shown in quotation marks, while the correct words of the items are presented in italics.

Appendix B

Exploratory analysis: Inspecting name repetition effects from the previous standard naming round

Participants named the same pictures once or twice in a standard naming task before the speeded picture naming task. In similar picture naming tasks, familiarizing participants with the experimental items is considered common practice, aiming to reduce naming difficulties related to the visual identification of the items. However, in the standard naming round(s) preceding the speeded round in this study, participants were not provided with feedback on their response accuracy. Instead, when exposed to the items in the standard naming round, they assigned a specific label to each picture (without it necessarily being correct), which would then be highly accessible in the speeded naming task. Such priming could facilitate naming in the speeded naming round, or hinder it, for example if participants decide to select a different label. Thus, we investigated if participants' RTs in the speeded naming task were influenced by whether they named the pictures using the same or different labels from the ones they used in the previous standard naming round. Half of the participants performed one standard naming round and the other half performed two standard naming rounds before the speeded naming round. Here, we focused on the data of the participants (N = 38) who had completed only one standard naming round before the speeded naming round.

Data analysis

In order to test whether RTs in the speeded naming round were affected by name repetition from the previous standard naming round, we classified the speeded naming responses based on whether they matched the response of the first standard naming round (e.g., first standard response: *apple*; speeded response: *apple*; category: matched) or not (e.g., first standard response: *orange*, speeded response: *apple*; category: mismatched).

Data points for four items (*board, bridge, racquet, pie, cup*) and one participant were excluded from the analysis, as these values were missing from the first standard naming dataset. This led to an overall of 10,280 datapoints.

Speeded naming responses that used exactly the same word as the standard naming responses or that were part of the standard naming responses (e.g., standard naming response: *apple, red apple* or *pah...apple*, speeded response: *apple*) were classified as *matched*. This category also includes trials where the speeded response was part of the standard naming response, but they were not semantically similar (e.g., standard naming response: *eggplant*, speeded response: *egg)*. These were overall 14 trials (out of 10,280). We decided to keep these trials in the *matched* category, as they could be considered a truncated version of the previous standard naming response. Other speeded naming responses were classified as *mismatched*. Table B1 shows the number of observations per *Response Accuracy* (target and error response) and *Response Matching* (matched and mismatched speeded response) condition. Matched targets are correct responses that were produced in both standard and speeded round. Matched errors are erroneous responses[10] that were produced in both standard and speeded round. Mismatched targets correspond to correct speeded responses that did not match the standard response. Note that out of the standard naming responses that mismatched the speeded target responses, 47% were errors, 4% were synonyms or abbreviations, and 49% were omissions. Moreover, out of the

[10] As in the target-error classification for the primary analyses and following Lampe et al's (2023) classification of the same data, errors are responses that do not match with the standardized names in McRae et al.'s (2005) database.

standard naming responses that mismatched the speeded error responses, 69% were correct responses, 13% were errors, and 18% were omissions.

Table B1

Number of Observations and Mean Response Times per Response Accuracy Type in the Speeded Naming Task and Response Matching Between the Speeded and the Previous Standard Naming Task.

Response Accuracy in Speeded Naming	Response Matching between Speeded and Standard Naming	N	M
Target	Match	7,804	599 (23)
Target	Mismatch	424	747 (36)
Error	Match	513	643 (31)
Error	Mismatch	1,539	665 (32)

Note: N = number of observations; *M* = mean response time; Mean response times are given in milliseconds. Standard errors of the mean are shown in parentheses.

In order to statistically evaluate RT differences between conditions of Response Accuracy and Response Matching with the previous naming round, we performed a linear mixed effects models analysis, with log-transformed naming latencies as the dependent variable. The factors *Response Accuracy* in the speeded naming task (effect-coded: *target* = -0.5, *error* = +0.5) and *Response Matching* between the speeded and the previous standard naming task (effect-coded: *Match* = -0.5, *Mismatch* = +0.5) as well as their interaction were included in the model as fixed effects. The model also included by-item and by-participant random intercepts and by-participant random slopes for Response Accuracy and Response Matching. Finally, we visually examined the cumulative RT distribution for each of the conditions following the same procedure as reported in the primary analysis.

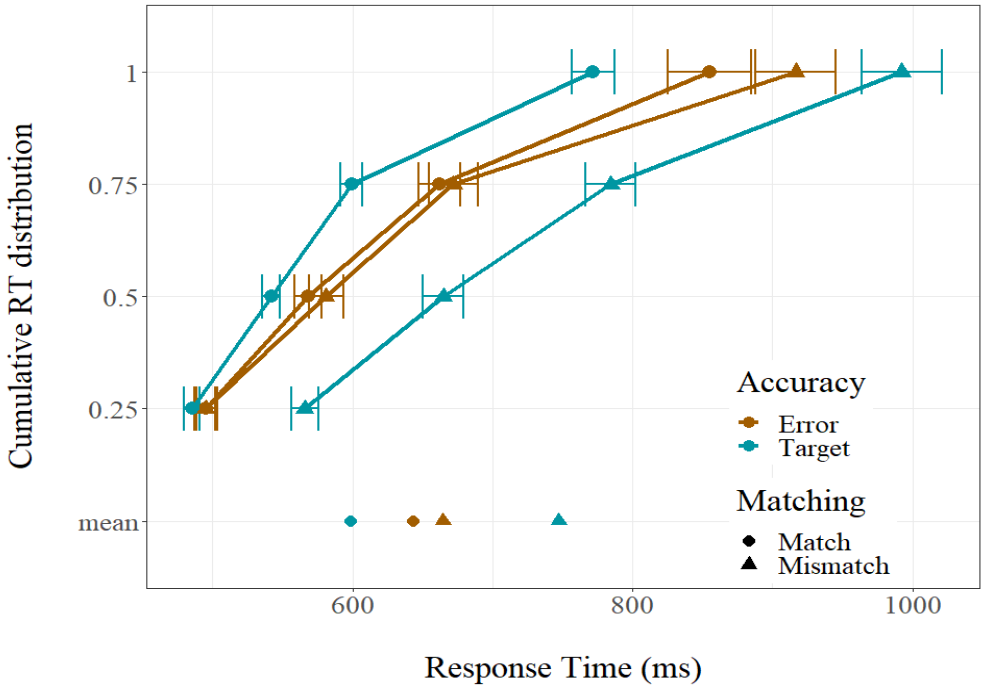
Results and Discussion

The cumulative RT distributions are shown in Figure B1. Group-level mean RTs and standard errors for the different response types are shown in Table B1. The figure shows a noticeable RT difference between the distributions for the matched target words and the mismatched target words. However, RTs of the matched errors and the mismatched errors are similarly distributed.

These observations were confirmed by a linear mixed effects models analysis, which revealed a main effect of Response Matching, suggesting that, irrespective of Response Accuracy, speeded responses that mismatched participants' responses in the standard naming round (M = 686 ms, SE = 34 ms) were produced more slowly relative to those that matched their responses in the previous (standard naming) round (M = 602 ms, SE = 24 ms; β = .14, $S.E.$ = .012, t = 11.671, p < .001). Interestingly, we found no main effect of Response Accuracy (β = .01, $S.E.$ = .01, t = 1.593, p = .11). Crucially, we found a significant interaction between Response Matching and Response Accuracy (β = -0.14, $S.E.$ = .013, t = -10.234, p < .001). Given our research question for this analysis (Does Response Matching differentially influence the RTs of targets and errors?), we performed simple-effects analyses in order to test whether RTs of targets and errors separately are predicted by Response Matching. The results showed that the effect of Response Matching was present only in target responses (β = .13, $S.E.$ = .012, t = 10.99, p = <.001), but not in error responses (β = -.006, $S.E.$ = .013, t = -.494, p = .62).

Figure B1

Cumulative RT Distribution for Target and Erroneous Responses of the Speeded Naming Round that Match or Mismatch the Responses of the Previous Standard Naming Round



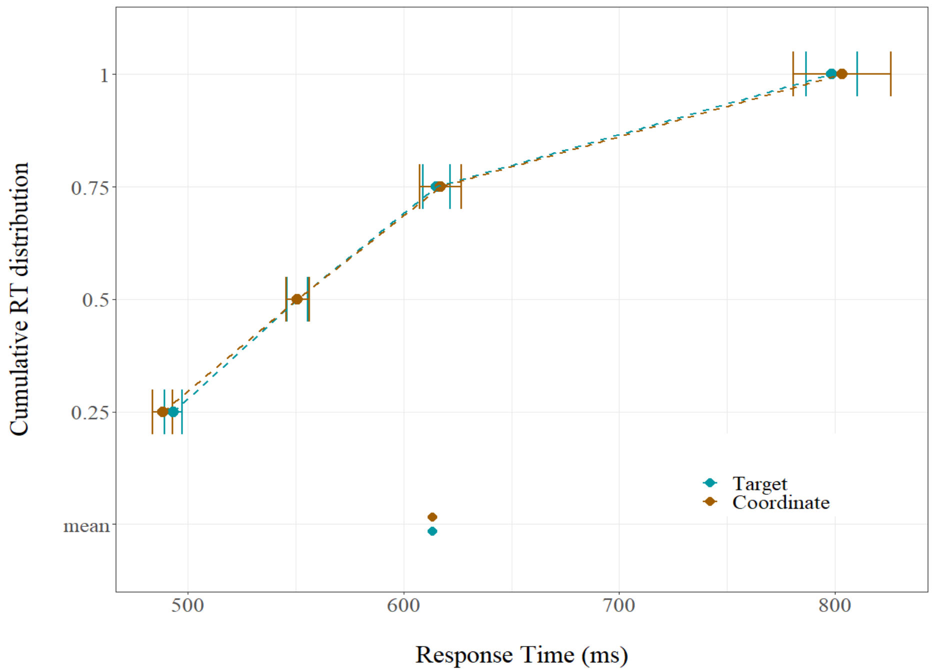
Note. Mean RTs are shown at the bottom of the figure. Error bars indicate the standard error for each RT bin. RT = response time; ms = milliseconds.

In sum, we explored whether participants' naming latencies in the speeded naming task were affected by using the same or different labels as they did in the previous standard naming round. The main effect of Response Matching indicates that speeded responses that were previously produced for the given items were generally faster than speeded responses that were different from those in the standard naming round, most likely due to being highly accessible for retrieval. The simple effects analysis showed that this effect is driven by one level of response accuracy, namely the target responses. Specifically, mismatched target responses were produced more slowly than matched target responses. Mismatched target responses represent cases where participants either correct a previous error or change their response from the standard naming round. This change requires additional cognitive processes: : participants either recognized and corrected a mistake they made at an earlier round or—if they had not responded at all in the earlier round—they were more deliberate in providing a correct response in the speeded round. Both scenarios involve more complex cognitive operations compared to simply retrieving a previously used, correct label, thereby resulting in slower RTs.

Interestingly, we found that matched and mismatched errors were not produced at significantly different naming latencies. Matched errors—repeated incorrect responses—suggest persistent visual uncertainty or misidentification. Mismatched errors—new incorrect responses or corrections of previous correct responses—indicate difficulties under time pressure or with visual identification. The lack of difference in RTs between these error types suggests that the cognitive processes involved in producing errors, whether consistent or not, are similarly demanding.

Appendix C

Cumulative RT Distribution for Target Responses and Coordinate Errors



Note. Mean RT for each of the two response types is shown at the bottom of the figure. Error bars indicate the standard error for each RT bin. RT = response time; ms = milliseconds.