To select or to wait? The importance of criterion setting in debates of competitive lexical selection

Nazbanou Nozari[a,b]* and Christopher R. Hepner[a]

[a]*Department of Neurology, Johns Hopkins University, Baltimore, USA;* [b]*Department of Cognitive Science, Johns Hopkins University, Baltimore, USA*

Correspondence should be addressed to:
Nazbanou Nozari, MD, PhD
Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, USA
Phone: 443-287-1712
Fax: 410-955-0188
E-mail: nozari@jhu.edu

Contact information for Christopher R. Hepner:
Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, USA
Phone: 443-287-1706
E-mail: chepner3@jhu.edu

To select or to wait? The importance of criterion setting in debates of competitive lexical selection

Competitive accounts of lexical selection propose that the activation of competitors slows down the selection of the target. Non-competitive accounts, on the other hand, posit that target response latencies are independent of the activation of competing items. In this paper, we propose a signal detection framework for lexical selection and show how a flexible selection criterion affects claims of competitive selection. Specifically, we review evidence from neurotypical and brain-damaged speakers and demonstrate that task goals and the state of the production system determine whether a competitive or a non-competitive selection profile arises. We end by arguing that there is conclusive evidence for a flexible criterion in lexical selection, and that integrating criterion shifts into models of language production is critical for evaluating theoretical claims regarding (non-)competitive selection.

Keywords: lexical selection, competitive selection, non-competitive selection, signal detection theory, response criterion, criterion setting, language production, monitoring, lateral prefrontal cortex, semantic interference

Most models of word production agree that producing a word involves several stages, the first of which is mapping semantic features onto lexical representations (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999). During this process, both the target word and other words that share semantic features with the target are activated through spreading activation. It is also generally accepted that this step is followed by a selection process in which a single activated lexical item is selected for further processing. One of the major debates in the word production literature centers on the nature of this selection process. The “competitive” accounts propose that the activation of (non-target) competitors slows down the selection of the target (e.g., WEAVER++; Roelofs, 1992). The “non-competitive” accounts posit that response latencies for producing the most highly activated item are independent of the activation of competing items (e.g., Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). A third group of accounts are agnostic with regard to competitive selection—as defined above—because they do not address questions of timing (e.g., Dell, 1986; see Dell, Nozari, & Oppenheim, 2014 for a review).

The main evidence in favor of the competitive account was first provided by Picture Word Interference (PWI) paradigms (e.g., Schriefers, Meyer, & Levelt, 1990) in which speakers showed longer response times (RTs) when they named a picture while ignoring a distractor word. More recently, the discovery of longer RTs when naming pictures in the context of semantically-related pictures has added to the evidence in favor of the competitive account (e.g., Belke, Meyer, & Damian, 2005; Schnur et al., 2009). This evidence has been countered by alternative explanations for paradigm-specific findings (Mahon et al., 2007; Navarrete, Del Prato, Peressotti, & Mahon, 2014; Navarrete, Mahon, & Caramazza, 2010). Whether lexical selection is competitive or not thus remains under dispute. In this article, we propose that a key element is missing from this debate. This element is the issue of a flexible criterion for selection, which depends on factors like task goals and the likelihood of successfully resolving conflict between competing responses (Lind, Hall, Breidegard, Balkenius, & Johansson, 2014)

(Semantic-lexical) mapping vs. selection

The first step in producing a word is activating that word from the relevant semantics. This process is referred to as *semantic-lexical mapping*, and depends on the activation of the semantic features and the strength of the connections between those features and the associated lexical representations (e.g., Dell, 1986). Words that are related in meaning (e.g., “cat” and “dog”) share a subset of their semantic features (e.g., *animal*, *furry*, *pet*, *four-legged*, etc.). Since semantic features may be associated with multiple lexical representations, activation of a single semantic feature can, and often does, lead to the activation of more than one word. However, in most cases, only one word is the target for production. A process must therefore be in place to select a single word from the set of activated words. This is referred to as the *selection process*. While mapping might have some spatiotemporal overlap with selection in the brain (e.g., Riès et al., 2017), the two processes can be distinguished computationally: a selection process cannot select representations that have not been activated, and a mapping process cannot adjudicate between multiple activated representations without a selection rule.

The dynamics of semantic-lexical mapping have been studied extensively and laid out in detail in a number of computational models of language production (e.g., Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Foygel & Dell, 2000; Goldrick & Rapp, 2007; Oppenheim, Dell, & Schwartz, 2010; Rapp & Goldrick, 2000; Roelofs, 1992). This research has contributed substantially to our understanding of the factors affecting the probabilities of different error types in neurotypical and damaged systems, as well as the facilitation and interference effects induced by representational similarity. For example, Dell et al. (1997) demonstrated that the pattern of increased lexical errors produced by individuals with certain types of brain damage could be simulated successfully by weakening the connection weights between semantic and lexical representations (and thereby decreasing the signal to noise ratio in the semantic-lexical mapping). Similarly, Oppenheim et al. (2010) showed that a model in which the production of each word causes incremental changes to the weights of the connections between semantic and lexical representations can explain the greater difficulty often observed in the production of semantically-related items (see below). Much less attention has been paid to the details of the selection process. Note, however, that all of the models listed above still include at least a simple selection rule separate from the proposed dynamics of semantic-lexical mapping. For example, the interactive two-step model of word production (Dell, 1986) and its later variants implement the following selection rule: at a fixed time step *t*, inspect the activations of all lexical items and select the most highly activated one.

To highlight the importance of separating the dynamics of mapping and selection, we use the blocked cyclic naming paradigm (e.g., Belke et al., 2005; Schnur, Schwartz, Brecher, & Hodgson, 2006) and a computational model which explains the main findings obtained from it (Oppenheim et al., 2010) as an example. In this paradigm, participants repeatedly name a small set of pictures in semantically related and unrelated blocks. The basic finding is interference in semantically-related blocks, but not unrelated blocks, after the first cycle, in the form of longer reaction times (RTs) and/or more errors (e.g., Schnur et al., 2006, 2009). Importantly, the magnitude of this interference is proportional to the semantic distance between the items (Vigliocco, Vinson, Damian, & Levelt, 2002). This clearly links the interference effect to the dynamics of the semantic-lexical mapping process (as opposed to selection): the more semantic features the items share, the stronger the effect. To explain this effect, Oppenheim and colleagues proposed that naming an item (e.g., "dog") causes small but persistent changes in the connection weights between semantic features and lexical items: the connections between the item produced (e.g., "cat") and its semantic features (e.g., *animal, pet*) are strengthened, while the connections between those features and competing lexical items (e.g., "dog", "rabbit") are weakened. Because of this, naming a picture "dog" would be more difficult after naming a picture "cat" (as opposed to naming a semantically-unrelated item like "pen"), because "dog" would receive less activation from its semantic features. As a result, there would be less difference in activation between "cat" and "dog"; i.e., greater competition. Changes to the mapping part of the process can thus provide a sufficient explanation for the increased competition in related blocks in the blocked cyclic naming task, and potentially other tasks that require naming a picture in the presence of semantically related items (e.g., continuous naming without repetition; e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006; cf. Navarrete et al., 2014). However, lexical retrieval is not yet complete. Which response will be selected? When will it be selected? How much more activation than other potential responses must it have to be selected? These are the questions for a theory of the *selection process* in lexical retrieval.

Put differently, the mapping process determines how many items would be activated in a given production situation and to what degree. It can thus make a prediction about the level of difficulty involved in a given production task. The behavioral consequences of this difficulty depend on the selection process. For example, in the Oppenheim et al. (2010) model, if the selection rule was "at a fixed time *t*, pick the lexical representation with highest activation", the model would predict a higher rate of semantically related errors in related blocks than in unrelated blocks (on occasions when the activation of a competitor surpasses that of the target), but it would not predict a higher proportion of trials in which no response is produced, i.e., omission errors. However, two other behavioral outcomes have been reported in the blocked cyclic naming paradigm in addition to the increased rate of semantic errors: longer RTs and an increased rate of omission errors (e.g., Schnur et al., 2006, 2009; see Oppenheim et al., 2010 for a comprehensive review of the empirical evidence). A viable selection theory must explain all three outcomes. To this end, Oppenheim et al. (2010) proposed a "booster" mechanism that operates when more than one lexical item is activated, repeatedly amplifying the activation of all of the lexical representations until the difference between the activation of the most highly activated lexical representation and its competitors reaches a certain criterion. Semantic errors arise when a semantic competitor's activation is higher than the target by at least the amount set as the criterion. Longer RTs emerge when more boosts are required to reach the criterion, and omission errors happen when the booster "times out" before the criterion is reached (Oppenheim et al., 2010, p. 231).

Oppenheim et al.'s (2010) model employs a production-external process (the booster) to implement the requirement of the difference criterion, and thus determine the point at which a lexical item is selected. An example of such a process is a competition-biasing operation attributed to the lateral prefrontal cortex (LPFC; e.g., Nozari & Thompson-Schill, 2013; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Thompson-Schill et al., 1998; see Nozari & Thompson-Schill, 2015 for a review; see also Riès, Karzmark, Navarrete, Knight, & Dronkers, 2015 for a discussion of when the LPFC may or may not be involved in lexical selection). But a difference criterion can also be enforced without appealing to production-external processes (e.g., Roelofs, 1992). Either way, the question remains: how does the production system determine this difference criterion? This question, which we refer to as the problem of criterion setting, is the main question that a theory of selection must address, and is the focus of this article.

What is a criterion?

A criterion is a psychological threshold above which a behavior is exhibited. The concept of a criterion has been an integral part of memory research for decades (e.g., Treisman & Williams, 1984). In the context of signal detection theory, old and new items in memory probes were thought to be associated with two overlapping distributions of familiarity (Banks, 1970). To determine the level of familiarity at which an item should be declared "old", a criterion had to be placed somewhere on these distributions. If the criterion is placed such that a high level of familiarity is required before an item is labeled as "old", all new items—which have low familiarity—will be labelled as "new", but so will some of the old items that have not been well registered. On the other hand, if the criterion is placed more liberally, such that even a little familiarity is deemed enough to declare an item "old", all old items will be recognized as old, but so will some of the new items by mistake. Importantly, criterion setting is a dynamic process; not only does the placement of the criterion differ from person to person, but it can also shift within individuals (Cox & Shiffrin, 2012; Singer & Wixted, 2006).

Discussion of a criterion and how it shifts is also relevant to language production. Nozari, Dell, and Schwartz (2011) have reported a clear case of an adjustable criterion for detecting error responses. A key concept in the monitoring account proposed by these authors is *conflict*, which can be defined as the inverse of the difference between the activation of the most highly activated representation and the next highest[1]. Using computational simulations, the authors showed that the amount of conflict is predictive of the probability of making an error, verifying the intuition that the activation of competitors makes production more difficult. They then proposed a signal detection framework for monitoring, similar to the one for recognition memory except that the two distributions represented the amount of conflict associated with correct and error responses instead of the familiarity of new and old items. Whether a response is detected as an error or not depends on the placement of the criterion. A

[1] Other functions that take into account more competitors yield the same result.

response associated with conflict levels above the criterion generates an error signal. Otherwise, no error signal is generated and the response passes as correct. This monitor works efficiently in neurotypical mature production systems, in which the two distributions are far apart (Figure 1a); it has a high hit rate (correctly detecting an error; light pink area) with a reasonably low false alarm rate (mistakenly detecting a correct response as an error; Nozari et al., 2011). If the criterion is placed farther to the left, the speaker can detect even more of their errors, but there will inevitably be more false alarms (see Nozari et al., 2011 for a report of such a case in an individual with aphasia). Importantly, monitoring can also intercept errors before they emerge (Hartsuiker & Kolk, 2001) and thus directly impact the primary production process. When instructed to avoid errors, neurotypical speakers reduced their error rates substantially (from 1519 to 687 in a tongue-twister task; Postma, Kolk, & Povel, 1990). Critically, however, this decrease in error rates was accompanied by an increase in silent and filled pauses that, at least in part, represent false alarms (from 169 to 243; Postma et al., 1990), compatible with a leftward criterion shift.

In this article, we apply the framework previously used to explain monitoring to the process of selection itself and discuss its utility in explaining behavioral response patterns and adjudicating between theories of selection.

Response criterion in lexical selection

A monitoring framework like the conflict-based monitor (Hanley, Cortis, Budd, & Nozari, 2016; Nozari et al., 2011; Nozari & Novick, 2017) can easily be integrated into the selection process by using the same critical information, namely the amount of conflict. The selection process within this framework is described in the following example, with "cat" as the target, and "dog" as the closest competitor: At time *t*, each of these lexical representations has an activation level; for example, 0.02 and 0.01, respectively. The monitor compares this level of conflict (i.e., the inverse of the difference between the two activation levels; in this case, 100) to a criterion (e.g., 110). If the level of conflict is below the criterion, selection proceeds immediately. In neurotypical systems under normal (low conflict) circumstances, the majority of these responses will be correct (Figure 1, light blue area), with a few errors in which the activation of "dog" was significantly above that of "cat" (Figure 1, dark purple area).

If, on the other hand, the activation levels of "cat" and "dog" are closer together, e.g., 0.019 and 0.015 the level of conflict is above the criterion (e.g., 250), and the monitor delays selection in favor of further processing. This is helpful, because the response that would be produced if selection were to proceed immediately on these trials would have had a high likelihood of being an error due to noise (Figure 1, light pink area) and only a small chance of being the correct response (Figure 1, dark blue area). In healthy production systems, the additional processing time during the delay imposed by the criterion often leads to a better signal to noise ratio, either by allowing activation to spread further within the production system, or through the help of production-external processes like prefrontally-mediated biasing mechanisms such as the booster in Oppenheim et al.'s (2010) model (see above). Either way, the delay often leads to the correct response gaining substantially more activation than its competitors (Figure 1b, activation graphs), decreasing the level of conflict over time (t_1, t_2, etc.) to the criterion or below at t_n (Figure 1b, conflict graph). The higher the initial level of conflict, the more time would be necessary for the level of conflict to drop below the criterion, which would result in longer delays.

The critical question is what causes this criterion to be set at, for example, 110 and not at 80 or 140. Naturally, a more stringent selection criterion (one far to the left) would decrease the chance of errors because most of the trials that could potentially lead to an error are delayed for further processing, but it would also impose longer delays on the majority of trials for the same reason, ultimately leading to more accurate speech with longer delays. On the other hand, a more liberal criterion (placed far to the right) would lead to more fluent (i.e., less often delayed) but more error-prone speech. Finally, the criterion may be flexible, changing from situation to situation. To situate the competitive selection account in this framework, we first present the most straightforward implementation of competitive selection, a *fixed-criterion model*. We review the model's predictions and assess their compatibility with empirical data. We then present an alternative *flexible-criterion model* and discuss the empirical evidence in light of this model.

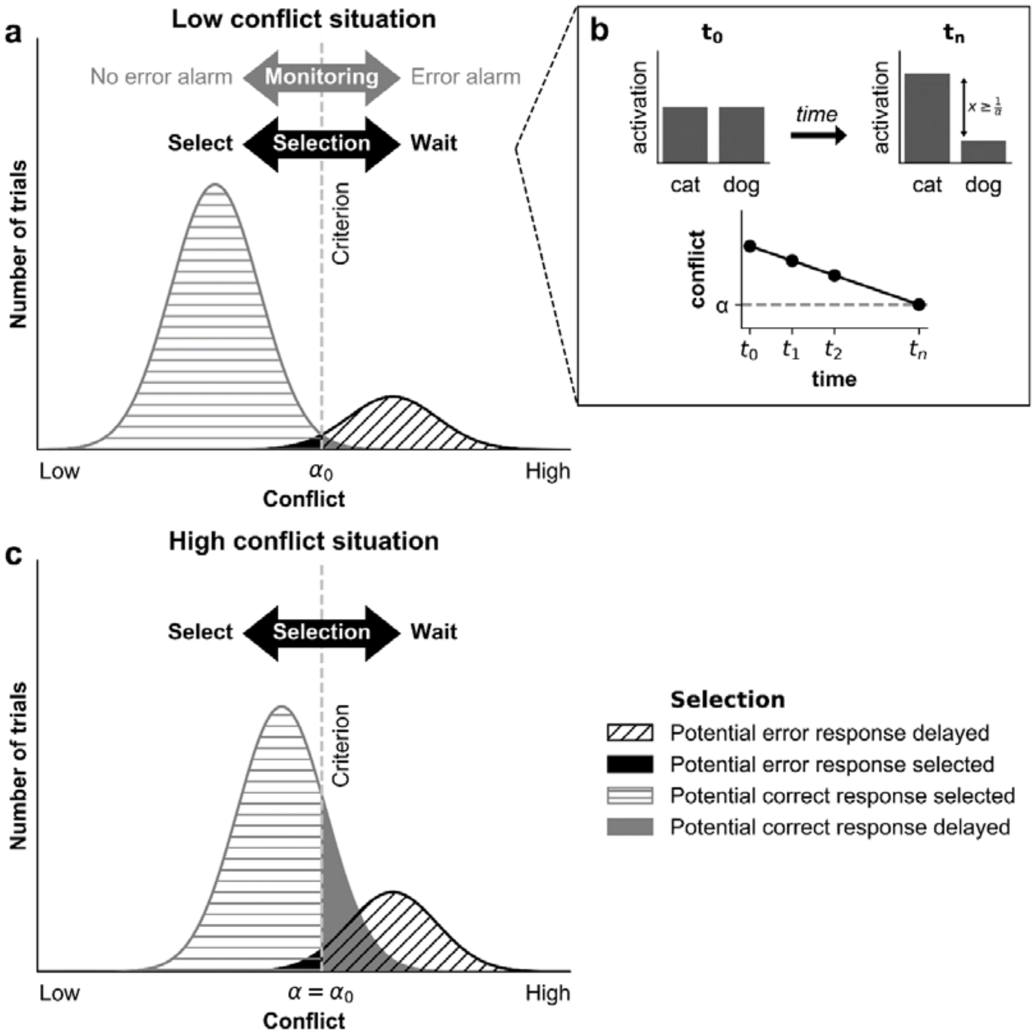

Figure 1. The fixed-criterion model of selection. a) Under normal (low-conflict) circumstances, the production system aims for a criterion, α_0, that would optimize performance (reach an acceptable balance between accuracy and latency). The intersection between the conflict distributions of potential correct and error responses is a reasonable location. Responses with conflict levels below α_0 are selected without delay (light blue and dark purple areas). Responses with conflict levels above α_0 are delayed (dark blue and light purple). b) Delayed responses undergo further processing until conflict drops to or below α_0. The amount of time to reach α (t_0 to t_n) determines response latency. c) Situations of high conflict shift the distribution of responses towards higher conflict values and bring the distributions of correct and error trials

closer together. The fixed-threshold model requires the criterion to be maintained at α_0, which inevitably leads to more delayed responses (larger dark blue area). Unless α is very large, the predictions of the fixed threshold model will always align with that of a competitive model of selection.

The fixed-criterion model of selection

The fixed-criterion model of selection is summarized in Figure 1. In this model, the production system aims to optimize performance under normal (low conflict) circumstances, meaning that it aims for a criterion that would provide reasonable accuracy with acceptable amounts of production delay (to avoid overly disfluent speech). A criterion placed at the intersection between the two distributions often provides a reasonable solution to the optimization problem in signal detection theory, so we adopt that position here as well. Decisions to proceed with, or to delay, selection follow the same dynamics explained above. The critical feature of the fixed-criterion model is that, once this criterion has been determined, it is held constant at the same conflict level α_0 for all situations, including high-conflict ones. To fully understand the ramifications of maintaining a fixed criterion, we must first review the changes that take place in the system under high-conflict situations, e.g., semantically-related blocks in the blocked cyclic naming paradigm. As stated earlier, these changes have to do with mapping part of the lexical retrieval process, and are independent of the selection process. For the discussion at hand, we focus on semantic-lexical mapping, though similar principles are likely to apply to other parts of the production system, e.g., lexical-phonological mapping (Breining, Nozari, & Rapp, 2016, under review; Nozari, Freund, Breining, Rapp, & Gordon, 2016).

A comparison between panels (a) and (c) of Figure 1 highlights two changes that take place when the system moves from a low-conflict to a high-conflict state: (1) the overall level of conflict increases, and (2) the distributions of conflict for error and correct trials get closer together because of the increased noise on all trials. This means that the activation levels of “cat” and “dog” will be closer (e.g., 0.019 and 0.015; conflict = 250) in a high-conflict situation than a low-conflict situation (e.g., the unrelated blocks in blocked cyclic naming), regardless of whether the trial will end in an error or a correct response. Thus, over many trials, the distributions of conflict for correct and error trials will be closer together (see Nozari et al., 2011 for simulations demonstrating this effect). Given these changes, maintaining the criterion at the fixed level α_0 (as predicted by the fixed-criterion model) helps keep the error rate (dark purple area) low, but inevitably leads to more delayed responses (larger dark blue area). Unless α_0 is very large, the predictions of the fixed-criterion model will always align with a competitive model of selection: longer RTs when competition is higher, a pattern of behavior commonly observed when neurotypical adults produce words in situations of high competition (e.g., Schnur et al., 2009; Schriefers et al., 1990). Since the criterion is set so as to ensure reasonable accuracy, a fixed-criterion model predicts that, in relative terms, higher conflict will have a larger effect on latency than on accuracy. However, there is at least some neuropsychological data which suggest this may not always be the case. For example, Riès, Greenhouse, Dronkers, Haaland, and Knight (2014) found that, individuals with LPFC damage did not show greater increases in latency in the semantically-related blocks of a blocked cyclic naming task than neurotypical adults. However, they had significantly greater increases in error rates. Below, we discuss a flexible-criterion model of selection, and demonstrate how the model can accommodate this finding while still explaining a competitive selection profile like the one predicted by the fixed-criterion model.

The flexible-criterion model of selection

Figure 2 shows the schema of a flexible-criterion model of selection. Under normal (low-conflict) circumstances, the system estimates the position of the criterion by attempting to optimize performance as described in the previous section (Figure 2a). However, instead of keeping this criterion (α_0) constant across all situations, the system flexibly reconfigures itself under new circumstances. If the optimization process gives equal priority to accuracy and speed, it will reposition the criterion at the intersection between the two distributions (Figure 2b). Note that because of the higher levels of conflict and the reduced distance between the two distributions, the new criterion α is higher than α_0. Moreover, both latencies for correct responses (dark blue area) *and* error rates (dark purple area) are increased compared to the low-conflict situation.

Since there is no commitment to a fixed criterion in the flexible-criterion model, the criterion can also shift further to the right (Figure 2c) or to the left (Figure 2d). If it shifts to the right, the result is a more liberal criterion even farther away from the original α_0, manifesting behaviorally as selecting correct responses with relatively preserved speed at the cost of producing more overt errors (much larger purple area). If the criterion shifts to the left (Figure 2d), the result is a more conservative criterion closer to the original α_0. Behaviorally, this manifests as longer latencies on correct responses (larger dark blue area) while error rates are kept relatively constant. Note that this scenario has the same behavioral consequences as a fixed-criterion model and is perfectly aligned with the predictions of a competitive model of selection. Thus a flexible-criterion model *can* be competitive, but it does not have to be. As such, a flexible-criterion model can accommodate both the response patterns predicted by competitive selection models (see above), as well as patterns like those reported in Riès et al. (2014), in which the relative increase in error rates is larger than the relative increase in latencies.

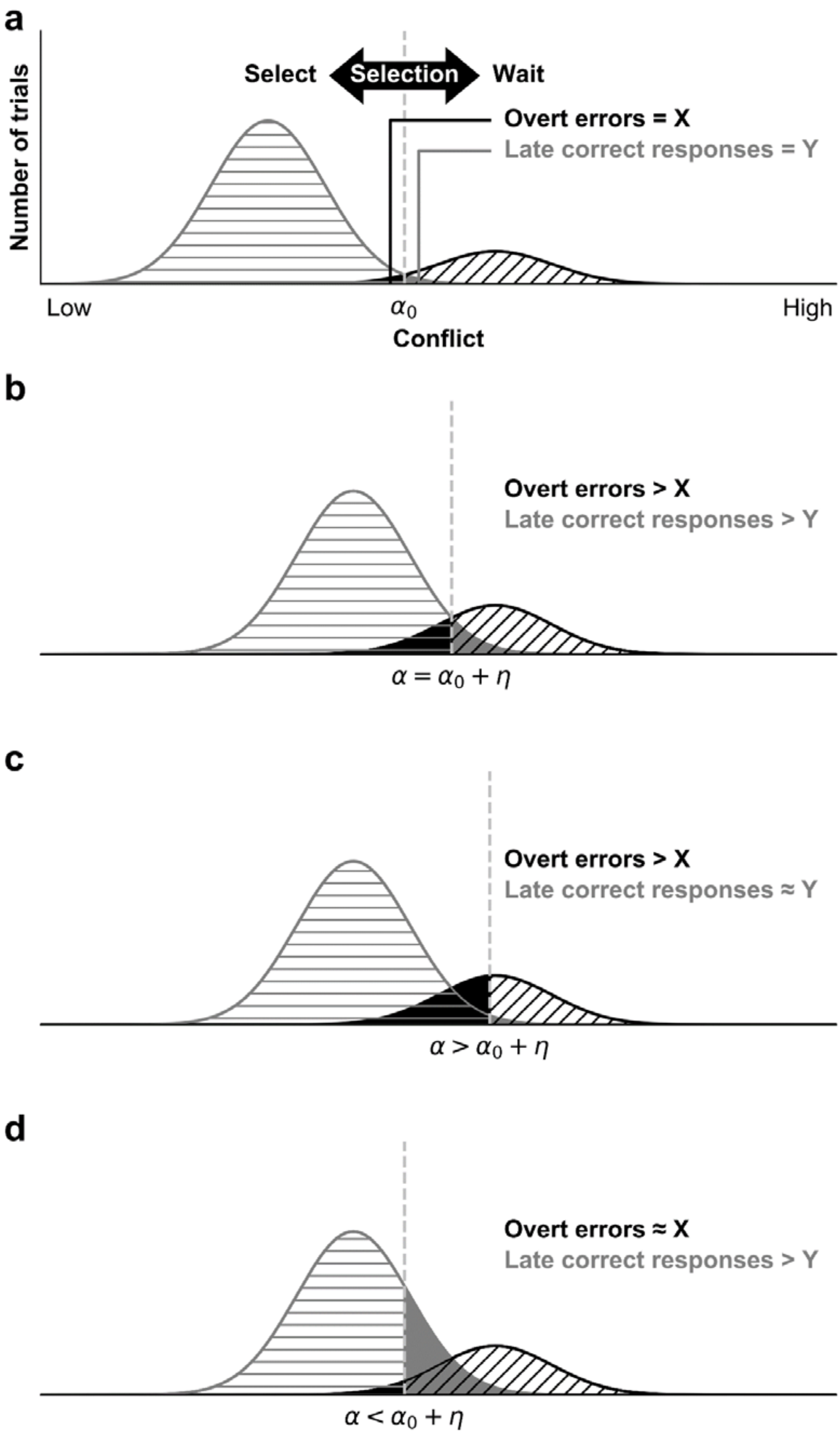
a
Number of trials
Select
Selection
Wait
Overt errors = X
Late correct responses = Y
Low
α_0
High
Conflict
b
Overt errors > X
Late correct responses > Y
$\alpha = \alpha_0 + \eta$
c
Overt errors > X
Late correct responses ≈ Y
$\alpha > \alpha_0 + \eta$
d
Overt errors ≈ X
Late correct responses > Y
$\alpha < \alpha_0 + \eta$

Figure 2. The flexible-criterion model of selection. a) In low-conflict situations, the system sets the criterion at α_0 as described in Figure 1. Unlike the fixed-criterion model, the system searches for a new optimal criterion α under high-conflict situations. b) If the criterion is placed at the intersection of the two distributions, both error rates (dark purple area) and latencies (dark blue area) will increase. c) If the criterion shifts to the right and away from the original α_0, latencies can be kept low, but the rate of overt errors increases. d) If the criterion shifts to the left towards the original α_0, the rate of overt errors will remain low, but latencies will increase. The flexible-criterion model can thus predict a pattern compatible with competitive selection (d), but also one compatible with non-competitive selection (c).

Flexible criterion, state of the production system, and task goals

Why a flexible criterion? The answer is to optimize performance under different circumstances based on task goals. In most cases, the goal is to produce speech with reasonable accuracy and fluency. In the framework presented here, there is always a trade-off between these two aspects of production. In healthy systems under low-conflict situations, it is possible to keep error rates low without delaying production too much. As conflict increases, this becomes more difficult, and speakers face a choice between sacrificing accuracy or sacrificing fluency (i.e., accepting longer delays). Prioritizing accuracy is usually the better choice because, in healthy production systems, a little bit of additional delay often allows the system to arrive at the correct response (Figure 1b). Thus, it is not surprising to observe a profile of competitive selection in neurotypical adults when conflict is high.

The situation, however, is not the same in individuals with brain damage. Since the focus of the current paper is lexical selection, we will limit our discussion to lesions that affect lexical access (i.e., semantic-lexical mapping), excluding damage to the core semantic system (e.g., semantic dementia; e.g., Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000), or problems in other parts of the production system (e.g., lexical-phonological mapping). Damage to the semantic-lexical mapping process, regardless of the specific mechanism (e.g., weaker connections, faster decay rates; see Dell et al., 1997 for a comparison), leads to increased conflict between lexical representations on all trials (e.g., Piai, Riès, & Swick, 2016), leading to a reduced distance between the conflict distributions of errors and correct responses (see Nozari et al., 2011 for a computational demonstration of this). In other words, a damaged system effectively becomes a system in a perpetual high-conflict state. When placed under high-conflict circumstances (e.g., semantically-related blocks in blocked cyclic naming) the situation resembles the depiction in Figure 3. The distribution of conflict for potential correct and error responses may have such a large degree of overlap that performance optimization (i.e., finding an acceptable balance between accuracy and fluency) becomes a serious challenge (Figure 3a).

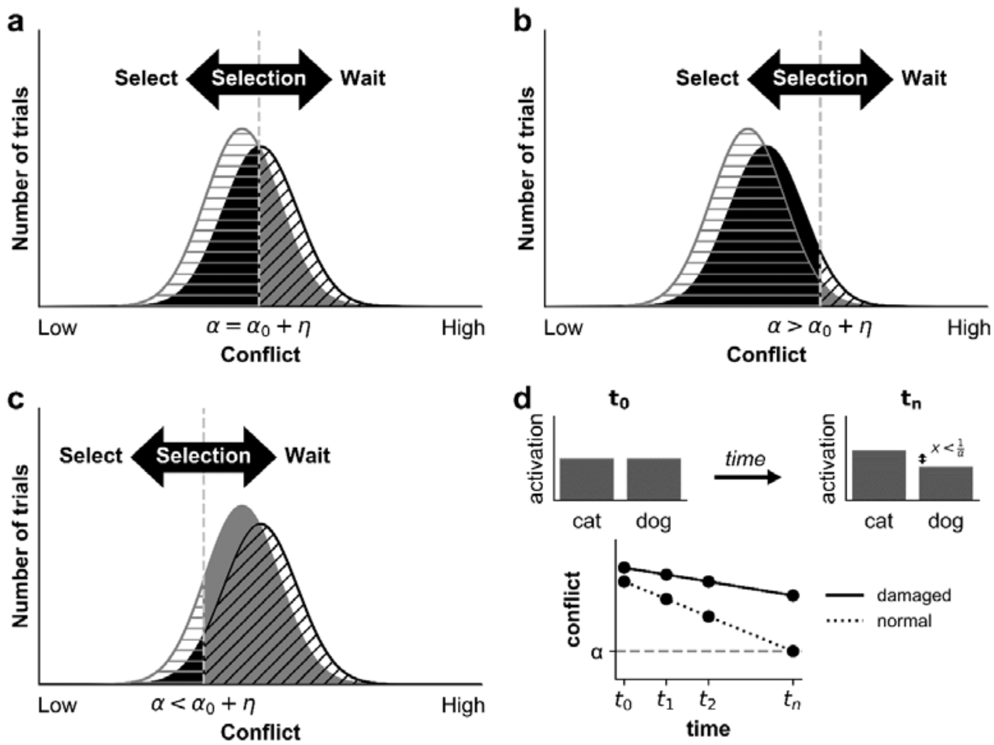

Figure 3. A high-conflict situation in a production system with damage to the semantic-lexical mapping process. The distributions of conflict for potential correct and error responses overlap significantly, so performance optimization is difficult, no matter where the criterion is placed. a) The criterion is placed at the intersection of the two distributions. b) The criterion is shifted to the right, away from the original α_0*, resulting in a large number of commission errors (mostly semantic) but few delays. c) The criterion is shifted to the left, towards the original* α_0*, resulting in a large number of delayed responses. d) Critically, the delayed responses are unlikely to reach the desired criterion in a reasonable time because of the poor state of semantic-lexical mapping, leading to a large number of omission errors and, in severe cases, near-mutism. The absence of a competitive selection profile (c) in many individuals with aphasia may reflect the system's optimization to escape this near-mute state.*

Under these circumstances, if the speaker behaves like a neurotypical individual and tries to preserve accuracy by shifting the criterion to the left, closer to the original α_0, the number of potentially correct trials with delays will be massive, so fluency will be considerably reduced. In addition, the delay that gives neurotypical systems time to convert potential errors (light pink area) to correct responses is not nearly as effective in damaged systems (Figure 3d). The reason for this is that whatever mechanism led to the increased conflict in the damaged system in the first place, e.g., lower semantic-lexical connection weights, will also affect processing over the delay period. Since this mechanism is impaired, there is a high probability that the desired criterion α_0 will not be reached within the same time frame as in neurotypical systems (t_0 to t_n), or within a reasonable time frame at all. The consequence is a very large number of omission errors. This profile has indeed been reported in individuals with lexical access problems. For example, Schnur et al. (2006) reported that their participants with aphasia made significantly more omission errors than controls in the semantically-related blocks of a blocked cyclic naming task. Similarly, Robinson, Shallice, and

Cipolotti (2005) reported a significant increase in the rate of omission errors in completing unconstrained sentences (e.g., "There is nothing wrong with the…"), compared to those in which contextual cues strongly biased competition towards a certain response thus reducing conflict (e.g., "Water and sunshine help plants..."; Robinson et al., 2005). Finally, the transient mutism resulting from damage to the Broca's area, part of the lateral prefrontal cortex that is hypothesized to be involved in conflict resolution (e.g., Thompson-Schill et al., 1998; see Nozari & Thompson-Schill, 2015 for a review), may reflect an attempt to keep the criterion as close as possible to the original α_0 (Levine & Mohr, 1979). These findings, together with the general profile of disfluent speech and severe word finding difficulties often linked to LPFC lesions (e.g., Buckner, Corbetta, Schatz, Raichle, & Petersen, 1996), indicate that some individuals with damage to the semantic-lexical mapping process still maintain a criterion close to the original α_0, in line with the prediction of the competitive account of selection.

Critically, however, many do not. If the criterion does not shift towards the original α_0 (Figure 3a/b), the result is a large increase in the number of overt errors. If the task manipulates semantic similarity, the majority of these errors will be semantic errors (e.g., "dog" instead of "cat"). In line with this prediction, Blanken, Dittmann, and Wallesch (2002) reported an individual with a stroke in the Middle Cerebral Artery (most likely leading to LPFC damage) who made more semantic errors when naming pictures with more close semantic neighbors (e.g., "spoon"; neighbors: "fork", "knife") than pictures with fewer close semantic neighbors (e.g., "glasses"). Similarly, Schnur et al. (2006) showed that a group of individuals with LPFC damage produced reliably more semantic errors when naming pictures in the context of other semantically-related pictures than in a mixed context. As mentioned above, the same study also reported significantly more omission errors in the related context, although it is unclear whether different individuals showed different profiles, or whether different trials within the same individual were associated with different error types. Either way, the existence of both profiles requires a model in which the criterion does not have to be kept in the same position regardless of the level of conflict in the system. Finally, as alluded to earlier, Riès et al. (2014) also reported individuals with LPFC damage who showed exaggerated costs compared to controls in the semantically-related condition in terms of increased rate of semantic errors, but not response latencies. In an attempt to model the RTs using an evidence-accumulation model, Anders, Riès, Maanen, and Alario (2017) found that one of the main differences between LPFC and control groups was the response criterion parameter: the control group had a higher response threshold (i.e., a lower or more conservative criterion closer to the original α_0 in the current framework) in the semantically-related condition, but the LPFC group did not ($\alpha > \alpha_0$). While the interpretation of this finding requires assumptions about the relationship between errors and RTs and consideration of the impact of the excluded omission trials on parameter estimations (see the Open Questions section below), it shows that the selection criterion is not always held constant. In summary, neuropsychological evidence shows that, under high-conflict circumstances, individuals with impaired semantic-lexical mapping may show a profile consistent with competitive selection (i.e., lots of omissions and delays) if they choose to keep the criterion close to the original α_0. However, they may also show profiles that are better aligned with non-competitive selection (i.e., lots of semantic errors with relatively little cost to latencies in high-conflict situations) if they allow the criterion to move far above the original α_0.

But is non-competitive selection only possible in damaged systems? To answer this question, we must revisit the concept of performance optimization based on task goals. Recall that the framework discussed here predicts a trade-off between accuracy and fluency. When conflict is very high and the two distributions are very close together, maintaining accuracy could result in very long delays and many omission errors, to the degree that the speaker may seem to be mute. In such cases, raising the criterion to allow faster responses at the cost of producing some commission errors is actually beneficial because it shows the speaker's intent to meet the task goal, i.e., to engage in the conversation, or to do the task required of them. In the same vein, the task goal may modulate criterion setting in neurotypical individuals. The classic example is the prioritization of either speed or accuracy; in forced-choice tasks like lexical decision, this is modeled by a change in the response threshold, i.e., the criterion (e.g., Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). Note that the task goal may also determine which of the possible responses count as correct, changing the definition of accuracy itself. A good example is the use of specific labels for objects. Certain contexts might demand that an object be named using a specific label, e.g., "settee", while under other circumstances it might be perfectly reasonable to refer to the same object as "couch", "sofa", "settee", or similar labels. In terms of semantic-lexical mapping, both situations lead to the activation of multiple lexical competitors. However, resolving competition in favor of a particular response only matters in the former case. In other words, what is considered "accurate" is different based on the task goal, and there is at least some suggestion in the literature that this difference may affect criterion selection and the consequent behavioral profile. For example, Alario et al. (2004) found longer RTs for naming pictures with low name agreement, suggesting that activation of several close alternatives slows down selection, as posited by competitive accounts of selection (criterion moved closer to the original α_0; Figure 2d). On the other hand, Oppenheim (2017) found the opposite pattern: he collected new timed picture naming norms to test the idea that strong alternatives necessarily delay dominant name retrieval. He showed that RTs for the dominant label (e.g., "couch") were *faster* when the probability of the second most common label (e.g., "sofa") was *higher*, suggesting that, in contrast to the predictions of competitive selection accounts, the activation of multiple competitors actually made production faster (a criterion moved far to the right compared to the original α_0; Figure 2c). Both studies used simple picture naming studies that were well controlled for other factors affecting lexical retrieval, but there was one critical difference between the two. In Alario et al. (2004), a specific label was suggested to the participants for each picture: after they named each picture in Experiment 1, they were given the preferred label to be memorized for Experiment 2. In Experiment 2, they were explicitly instructed to use the preferred labels from Experiment 1. In both cases, participants were made aware that the task goal involved the production of one preferred label. By contrast, Oppenheim (2017) specifically avoided assigning any preference to particular labels during the study. Consequently, his participants approached the task with a slightly different goal; namely, producing any label that was suitable for the picture.

Task goals are also directly relevant to the interpretation of neuropsychological data. An example is a debate around the role of ventrolateral PFC (VLPFC). Martin and Cheng (2006) demonstrated that an individual with damage to the ventrolateral PFC (VLPFC) showed no impairment in generating a verb in response to the probe noun "door", even though "door" is associated with two verbs ("open" and "close") which, under a competitive account, must be competing strongly with one another for selection. Based on this evidence, the authors concluded that the hypothesis that VLPFC has a role in resolving conflict (e.g., Thompson-Schill et al., 1997, 1998) must be incorrect; otherwise, damage to this region should have caused great difficulty in responding to trials with probes such as “door”. However, as in Oppenheim’s (2017) study, *any* of the activated responses are equally acceptable given the goals of the verb generation task. Thus, instead of assuming that selection must be competitive and concluding that VLPFC must not have a role in conflict resolution, an alternative interpretation of this finding is that selection can be non-competitive if the task goal does not require conflict resolution in favor of a single preferred response. Consistent with this interpretation, when the task goal does require selection of a particular response among two highly activated competitors, individuals with lesions encompassing VLPFC do have difficulty: for example, Noonan, Jefferies, Corbett, and Lambon Ralph (2010) showed that participants with such lesions were impaired when they were asked to name a picture (e.g., *lion*) while being presented with the onset of a competitor (“T” for *tiger*).

In summary, any word production task is subject to two main constraints: time and accuracy. Task goals may prioritize either one (i.e., a speed-accuracy trade-off), or change the nature of the constraints altogether—for example, by changing the definition of an “accurate” response from a single preferred label (e.g., only “couch” is correct) to any of several acceptable alternatives (e.g., both “couch” and “sofa” are acceptable). The dynamics of selection, in particular the position of the criterion, may be determined by these task goals.

Is selection competitive or not?

Adding a criterion to the complex debates regarding the cognitive and neural architecture of activation and selection (Riès et al., 2017) may seem undesirable, especially given the fact that the criterion varies between individuals and even within the same individual in different situations. But the evidence presented in this paper suggests that it is the key to settling the question of competitive vs. non-competitive selection. We have shown that a fixed-criterion model almost always predicts a profile consistent with competitive selection. We then reviewed evidence that cannot be accommodated by this model. As a solution, we proposed a flexible-criterion model of selection which can accommodate both competitive and non-competitive selection. Note that, while the flexibility of the criterion allows the model to simulate different response profiles, these profiles are not unconstrained. For example, unless task goals are changed, it is not possible for a speaker to produce faster *and* more accurate responses in high-conflict situations compared to low-conflict situations.

In summary, a flexible criterion is necessary to explain the different response profiles observed in production tasks. Where the criterion is placed in each situation is determined by (1) the general state of the production system under normal (low-conflict) situations, (2) the current level of conflict in the system, and (3) whether task goals prioritize fluency (i.e. speed) or accuracy. When high levels of accuracy can be achieved at little cost to fluency, a competitive selection profile arises. When, on the other hand, the cost to fluency is unacceptable to the speaker, or multiple competitors (e.g., "couch", "sofa", etc.) are acceptable as accurate responses, a non-competitive selection profile is observed.

Open questions

The proposal of a flexible criterion can be implemented in a connectionist model, an evidence accumulation model (such as a drift diffusion model), or any other type of model with a selection mechanism that requires a criterion (Anders, Riès, van Maanen, & Alario, 2015; Roelofs, 1992; van Maanen & van Rijn, 2007). The common question, regardless of the specific model, is exactly which factors influence the position of the criterion under different circumstances. We have proposed a criterion that is sensitive to the level of conflict between activated representations, and have shown that using this information is sufficient to explain various response patterns in production tasks. But it is important to note that a selection criterion may be based on information other than conflict. For example, Lupker, Brown, and Colombo (1997) have proposed a "time criterion" for reading tasks. This time criterion can change based on factors like word frequency and determines when to begin articulation. A potential question for future research is whether a time criterion is more successful as explaining the empirical data in lexical selection than our proposed conflict-based criterion.

Another critical question is the time course of changes in the criterion. For example, do speakers move their criterion based on the outcomes of previous trials (i.e., the rates of omission and commission errors) over the course of a task even when task goals and conflict levels are stable? While behavioral reports from neuropsychological studies have pointed towards both competitive and non-competitive selection profiles (e.g., Schnur et al., 2006), it is unclear whether the same individual can show both profiles within the same task or not. To test this in single word production, one can study individuals' response profiles over the course of a task for changes that may indicate a criterion shift; for example, if an individual shows a predominance of omissions and long latencies in the first half of the task, but many more semantic errors and shorter latencies in the second half of the task, one can conclude that the selection criterion has changed within this individual during this task.

The example above, as well as various arguments throughout the paper, shows the special place that neuropsychological data have in furthering our understanding of the selection process in language production. The framework we have proposed also highlights the critical importance of examining error types (especially omission vs. commission errors), disfluencies (a manifestation of delay), and the interdependence between errors and RTs. This is particularly important when selecting data to use as the basis for computational simulations. For example, Anders et al. (2017) modeled RTs in a subset of the data from participants with PFC lesions reported in Riès et al. (2014). The exclusion criteria for this analysis included omission errors (44% of all errors), as well as "verbal dysfluencies (e.g., stuttering, utterance repairs: 4%) and hesitations (e.g., if the experimenter perceived the production of the possessive pronoun ["my" required on every response,] to be abnormally lengthened or separated from the production of the noun by a pause: 29%)." (p. 219). A model of RTs that excludes these categories is likely to arrive at a biased measurement of the criterion, since these response categories are directly relevant to the placement of the criterion.

We also emphasized the importance of task goals and performance optimization in criterion placement. While there is need for more solid empirical evidence for the influence of goals on criterion setting—Oppenheim (2017) and Alario et al. (2004) may have differed in other aspects as well as task goals—it is important to practice caution in interpreting the cause of a potential criterion shift (or lack thereof). For example, simulations by Anders et al. (2017) showed that, unlike neurotypical controls, patients with PFC damage did not appear to have changed their criterion in the high-conflict condition, leading the authors to conclude that the PFC patients may have a criterion setting deficit. While this is a viable possibility, the current framework suggests that, in a system with damage affecting semantic-lexical mapping, the absence of a leftward criterion shift towards the original α_0 may reflect the speaker's goal of avoiding a large number of omissions (and thus looking uncooperative or mute). Claiming that this profile truly reflects a deficit requires testing the individuals under circumstances in which a competitive selection profile is better aligned with the task goals; i.e., the individual is told that accuracy is the most important factor, even at the cost of producing few words. If the individual continues to show the non-competitive profile, one can then conclude that a criterion setting deficit may exist.

Finally, it is important to keep in mind that, while understanding the criterion setting mechanisms can shed light on various response profiles, many questions about interference and facilitation may be better suited for a theory of the mapping than a theory of selection. Examples of such questions may include the sensitivity of semantic interference effects to the passage of time and the number of intervening items (e.g., Schnur, 2014), or the role of item repetition in generating interference (e.g., Navarrete et al., 2014). Generally speaking, a selection theory does not explain the level of conflict (this is the province of a mapping theory) or provide a mechanism for eliminating the costs associated with higher conflict; it simply reflects the speaker's choice of the form this cost takes (speed or accuracy). A flexible criterion allows speakers to choose either less fluent speech with fewer commission errors or more fluent speech with more commission errors, unless the task goal fundamentally changes the constraints by allowing a larger set of response alternatives.

To conclude, based on the current framework, we highly encourage researchers to report error breakdowns (e.g., Schnur et al., 2006) in empirical studies of language production. We also note the value of computational models of response selection in word production, and emphasize the importance of considering the interdependencies between errors and RTs in modeling the process of lexical selection. The current framework makes general qualitative predictions, but ultimately such predictions must be verified by implementing and testing quantitative models.

Acknowledgments

This work was supported in part by the Therapeutic Cognitive Neuroscience Fund, and in part by a Science of Learning Initiative grant awarded to NN.

References

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

Anders, R., Riès, S. K., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57–73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Anders, R., Riès, S. K., van Maanen, L., & Alario, F.-X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology*, *34*(1–2), 1–20. https://doi.org/10.1080/02643294.2017.1282447

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81–99. https://doi.org/10.1037/h0029531

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

Blanken, G., Dittmann, J., & Wallesch, C.-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters*, *325*(1), 72–74. https://doi.org/10.1016/S0304-3940(02)00227-6

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*(9), 1207–1215. https://doi.org/10.1016/S0028-3932(00)00034-8

Breining, B., Nozari, N., & Rapp, B. (under review). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system.

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*(2), 500–506. https://doi.org/10.3758/s13423-015-0900-x

Buckner, R. L., Corbetta, M., Schatz, J., Raichle, M. E., & Petersen, S. E. (1996). Preserved speech abilities and compensation following prefrontal damage. *Proceedings of the National Academy of Sciences*, *93*(3), 1249–1253. https://doi.org/10.1073/pnas.93.3.1249

Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, *4*(1), 135–150. https://doi.org/10.1111/j.1756-8765.2011.01177.x

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: Behavioral and computational considerations. In M. A. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 88–104). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199735471.013.014

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

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*(2), 182–216. https://doi.org/10.1006/jmla.2000.2716

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260. https://doi.org/10.1016/j.cognition.2005.12.010

Hanley, J. R., Cortis, C., Budd, M.-J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, *142*(Supplement C), 36–47. https://doi.org/10.1016/j.jecp.2015.09.008

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*(2), 113–157. https://doi.org/10.1006/cogp.2000.0744

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/s0140525x99001776

Levine, D. N., & Mohr, J. P. (1979). Language after bilateral cerebral infarctions: Role of the minor hemisphere in speech. *Neurology*, *29*(7), 927–927. https://doi.org/10.1212/WNL.29.7.927

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 570–590. https://doi.org/10.1037/0278-7393.23.3.570

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Martin, R. C., & Cheng, Y. (2006). Selection demands versus association strength in the verb generation task. *Psychonomic Bulletin & Review*, *13*(3), 396–401. https://doi.org/10.3758/BF03193859

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, *76*, 253–272. https://doi.org/10.1016/j.jml.2014.05.003

Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, *134*(3), 279–289. https://doi.org/10.1016/j.actpsy.2010.02.009

Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2010). Elucidating the nature of deregulated semantic cognition in semantic aphasia: Evidence for the roles of prefrontal and temporo-parietal cortices. *Journal of Cognitive Neuroscience*, *22*(7), 1597–1613. https://doi.org/10.1162/jocn.2009.21289

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., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*(7), 886–903. https://doi.org/10.1080/23273798.2016.1157194

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403–410. https://doi.org/10.1177/0963721417702419

Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt? *Neuropsychologia*, *51*(13), 2770–2780. https://doi.org/10.1016/j.neuropsychologia.2013.08.019

Nozari, N., & Thompson-Schill, S. L. (2015). Left ventrolateral prefrontal cortex in processing of words and sentences. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 569–584). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00046-8

Oppenheim, G. M. (2017, September). *Strong competitors facilitate target name retrieval in simple picture naming*. Poster session presented at Architectures and Mechanisms of Language Processing 2017, Lancaster, UK.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Piai, V., Riès, S. K., & Swick, D. (2016). Lesions to lateral prefrontal cortex impair lexical interference control in word production. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00721

Postma, A., Kolk, H., & Povel, D.-J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, *33*(1), 19–29. https://doi.org/10.1177/002383099003300102

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499. https://doi.org/10.1037/0033-295X.107.3.460

Riès, S. K., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., … Knight, R. T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences*, *114*(23), E4530–E4538. https://doi.org/10.1073/pnas.1620669114

Riès, S. K., Greenhouse, I., Dronkers, N. F., Haaland, K. Y., & Knight, R. T. (2014). Double dissociation of the roles of the left and right prefrontal cortices in anticipatory regulation of action. *Neuropsychologia*, *63*(Supplement C), 215–225. https://doi.org/10.1016/j.neuropsychologia.2014.08.026

Riès, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T., & Dronkers, N. F. (2015). Specifying the role of the left prefrontal cortex in word selection. *Brain and Language*, *149*, 135–147. https://doi.org/10.1016/j.bandl.2015.07.007

Robinson, G., Shallice, T., & Cipolotti, L. (2005). A failure of high level verbal response selection in progressive dynamic aphasia. *Cognitive Neuropsychology*, *22*(6), 661–694. https://doi.org/10.1080/02643290442000239

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

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, *75*, 27–44. https://doi.org/10.1016/j.jml.2014.04.006

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, *106*(1), 322–327. https://doi.org/10.1073/pnas.0805874106

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

Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition*, *34*(1), 125–137. https://doi.org/10.3758/BF03193392

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797. https://doi.org/10.1073/pnas.94.26.14792

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences*, *95*(26), 15855–15860. https://doi.org/10.1073/pnas.95.26.15855

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*(1), 68–111. https://doi.org/10.1037/0033-295X.91.1.68

van Maanen, L., & van Rijn, H. (2007). An accumulator model of semantic interference. *Cognitive Systems Research*, *8*(3), 174–181. https://doi.org/10.1016/j.cogsys.2007.05.002

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*(3), B61–B69. https://doi.org/10.1016/S0010-0277(02)00107-5

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159. https://doi.org/10.1016/j.jml.2007.04.006