To select or to wait? Response to the commentaries

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

Abstract

In Nozari and Hepner (this issue), we proposed a theoretical framework for reconciling two seemingly irreconcilable theories of lexical selection: competitive vs. non-competitive selection. By superposing a decision framework (signal detection theory) onto the distributions of conflict derived from the core dynamics of mapping semantic features to lexical representations, we argued that a flexible selection criterion could accommodate patterns predicted by both competitive and non-competitive models of lexical selection. Five excellent commentaries posed various questions regarding the necessity, applicability, and scope of the proposed framework. This paper addresses those questions.

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

To select or to wait? Response to the commentaries

In memory of Albert Costa, who always sparked a good debate.

In Nozari and Hepner (this issue), we proposed a theoretical framework for reconciling two seemingly irreconcilable theories of lexical selection: competitive vs. non-competitive selection. This approach views the word production system as comprising two separate (but related) models: a) a model of the internal dynamics of mapping one layer of representation to another, a process that leads to the *activation* of representations, and b) a model of decision-making based on the information generated by the first system, a process that leads to *selection* (see Nozari, 2019, for neuropsychological evidence of the separation of activation and selection processes). In our framework, we have deliberately avoided modelling the internal dynamics of production because excellent models of these dynamics already exist (e.g., Foygel & Dell, 2000; Rapp & Goldrick, 2000). While we encourage combining them with a model of decision-making, this has not been the purpose of the current proposal. Instead, our goal here has been to show that the dynamics of activation (as implemented by such models) naturally generate information (i.e., conflict) that is useful for decision-making processes in the production system. Examples of such decision-making processes are those which determine when to mark a response as an error (Hanley, Cortis, Budd, & Nozari, 2016; Nozari, Dell, & Schwartz, 2011) or, as Nozari and Hepner (this issue) have argued to be a natural extension, when to delay selection in order to decrease the chance of an error in the first place.

In five insightful commentaries, Anders and Alario (this issue; A&A), Costa (this issue), Mahon and Navarrete (this issue; M&N), Melinger and Abdel Rahman (this issue; M&A), and Oppenheim and Balatsou (this issue; O&B) provided a wide range of comments on the utility, feasibility, and relevance of the framework, as well as suggestions for future applications. In the following two sections, we first tackle the issues of utility and feasibility. Many of the comments by M&A, M&N, and O&B are addressed here. The third section deals specifically with the comments of A&A on the link between evidence accumulation (EA) models and the current proposal. The fourth section addresses the remaining comments. We end by highlighting some of the unanswered questions encouraged by the current framework.

Does the proposed framework add anything?

Two of the commentaries (M&A and M&N) questioned the usefulness of the proposed framework, for exactly opposite reasons. M&A see the problem of lexical selection as already solved in favor of the competitive view, while M&N view competitive selection as altogether illusory, an artifact of specific experimental paradigms. In a recent review of the behavioral, neuroimaging, and EEG evidence of co-activation of multiple representations during production, we evaluated both of these claims (Nozari & Pinet, under review). In keeping with the claims of M&A, the empirical evidence for interference in word production is robust. Paradigm-specific idiosyncrasies cited as creating illusory interference (e.g., Navarrete, Del Prato, Peressotti, & Mahon, 2014) have been addressed with further empirical studies (e.g., Belke, Shao, & Meyer, 2017), and manipulation of semantic distance has shown that closer semantic relationships induce greater interference (Rose, Aristei, Melinger, & Abdel Rahman, 2019). There is, thus, little doubt, given the current state of the empirical evidence, that semantic similarity can, and often does, induce real interference in production.

Robustness of interference, however, does not automatically translate into competitive lexical selection. It has been shown that such interference can in fact be generated by other mechanisms, such as incremental learning (Oppenheim, 2018; Oppenheim, Dell, & Schwartz, 2010), without appealing to competitive selection. Similarly, ERP differences between semantically related and unrelated words during picture naming, which have often been taken as evidence for competitive lexical selection, might find an alternative—and potentially more powerful—explanation in the dynamics of spreading activation (see Nozari & Pinet, under review). It is true that, to date, non-competitive mechanisms like incremental learning have not been applied to findings from the PWI paradigm, whereas such findings can be easily accommodated using a competitive selection mechanism (Roelofs, 2018). How much weight should be given to findings from the PWI paradigm, which entails processes other than those routinely engaged in word production (e.g., reading or auditory processing of a distractor), is currently a topic of debate (see Oppenheim & Balatsou, this issue, vs. Roelofs, 2018).

To summarize, empirical evidence of interference arising from a non-strategic process is strong. However, in most cases, this interference can be explained by both competitive and non-competitive models of lexical selection. There are two ways out of this: 1) To continue adding to the very large body of behavioral and neural evidence, and interpreting them within the existing theoretical frameworks. It seems to us that this approach has reached the limits of its explanatory potential. 2) To entertain the possibility of a framework within which both selection mechanisms are possible. This framework, signal detection theory (SDT), has been immensely successful in testing the predictions of fixed threshold vs. flexible criterion models in a wide range of perceptual and cognitive tasks (see Wixted, in press, for an outstanding review of the history of SDT), and can speak to *why* performance looks competitive in one case and non-competitive in another. The purpose of Nozari and Hepner (this issue) was to advocate the adoption of this approach for understanding lexical selection.

Does the proposed approach work?

Two of the commentaries (M&A and A&A) raised concerns about the functionality of the proposed framework. Before we address these concerns, we feel that potential ambiguity relating to our use of the term "task goals" must be resolved. To make things clearer, we will draw a distinction between a *task demand* and a *performance goal*. By *task demand*, we mean the activity that must be performed. Using M&A's example, naming all of the pictures and naming a subset based on a certain rule would constitute activities with different *task demands* (within the same PWI paradigm). By *performance goal*, we mean the prioritization of either speed or accuracy. Thus in M&A's example, there are *two* different task demands (name all of the pictures vs. name a subset) but only *one* performance goal.

Are the *task demand* and the *performance goal* truly separable parameters? Yes. These two factors can be manipulated orthogonally. One could, for example, hold the task constant (e.g., name all of the pictures) while using different instructions to manipulate performance goals ("Respond as quickly as possible." vs. "Make sure you do not make a mistake."). Conversely, one could hold the performance goal constant (prioritize speed) while changing the task (name all of the pictures vs. name only the subset of pictures that matches rule X). The criterion would only be expected to change in the latter example, when the *performance goal* changes, and not when the *task demand* changes. In keeping with this prediction, Aristei and Abdel Rahman (2013) did not observe a change in the pattern of performance in their experiment because only the *task demand*, and not the *performance goal*, had changed.

Model fitting provides empirical evidence for the independence of these two parameters. Figure 1a shows data from 60 participants in a recent word typing study by our lab. Participants heard a word and typed it under time pressure without immediate visual feedback. After each trial, they were asked whether they had made an error or not. A signal detection model was fitted to each individual's data in order to estimate two parameters: *d′* and *c*. The *d′* parameter indicates the distance between the distributions of conflict for correct and error trials, which is determined by the task demands and the state of the production system. The other parameter, *c*, indicates the position of the criterion (in terms of the number of standard deviations from the criterion of an "ideal observer"—i.e., one placed at the intersection of the two distributions). As shown in the figure, *d′* and *c* can be completely independent from each other (Pearson's *r* = 0.03). Figures 1b-c show how this can be the case. Data from two participants with identical *d′* (1.91) are plotted: one (Figure 1b) had a hit rate of 55% and a false alarm rate of 3%, while the other (Figure 1c) had 37% and < 1%. The two participants had set different performance goals, and thus different selection criteria *c*, for themselves: minimizing misses and minimizing false alarms, respectively.

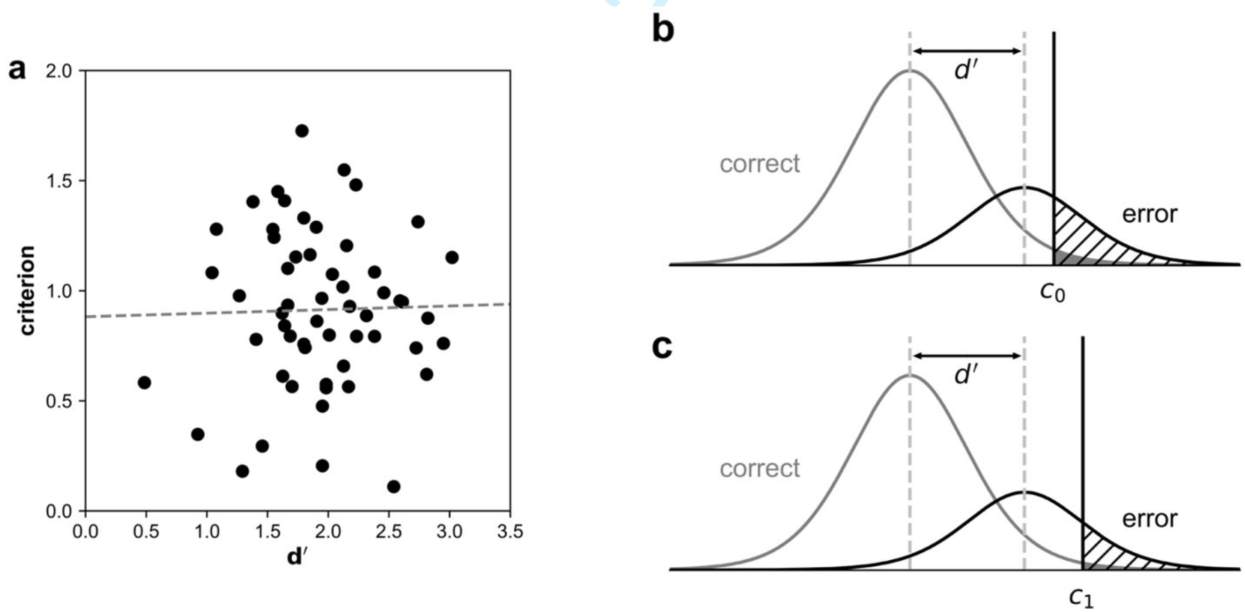

Figure 1. Separation of d' and criterion (c), parameters corresponding to the internal dynamics of the production system and the decision-making framework, respectively. (a) Scatterplot of criterion (c) and d' in a sample of 60 participants. (b-c) Schematics of model fitting for two

participants from the sample: participants had identical d′, but different criteria (see text for details).

The example above shows that the parameters indexing the internal dynamics of the production system (*d′*) and the placement of the criterion (*c*) are theoretically and empirically separable. This is also true of the analogous parameters in the EA models, including Drift Diffusion Models (DDMs), discussed by A&A. In an EA model, the internal dynamics of the production system are captured by the drift rate, while the point at which a response is made is determined by the response boundaries, similar to the selection criterion. Even simpler versions of the DDM (e.g., the EZ-diffusion model; Wagenmakers, Van Der Maas, & Grasman, 2007), which reduce the number of parameters by making simplifying assumptions, still retain these two parameters as separate (see Nozari & Dell, 2009, for an application of this model). This is not to imply that the two parameters never covary. When performance goals must be maintained despite changing task demands that cause a decrease in *d′* (reflecting closer distributions of conflict), a shift in the criterion towards a more conservative value is expected (see Figure 2 in Nozari & Hepner, this issue). To the degree the decrease in *d′* due to the changing task demands and the adjustment of the criterion to adhere to certain performance standards are similar across participants, a correlation between *d′* and *c* is expected (see Anders, Riès, van Maanen, & Alario, 2015, for a similar finding using a Shifted Wald model). Similarly, cases in which the criterion is pushed to an extreme (e.g., by instructions that strongly stress speed) could reduce drift rates in EAs (e.g., Starns, Ratcliff, & McKoon, 2012). The key point, however, is that the two parameters represent conceptually and empirically separable processes and are by no means redundant.

The link to EA models

The application of SDT and EA models to word production indeed shares a common goal: both models highlight the importance of viewing word production as a goal-oriented task with a decision-making component (see Nozari, 2018, for elaboration of this view). However, A&A go further to claim that our proposed approach "reduce[s] [neural network model] components to a framework of accumulation (of inverse conflict)", and that DDMs will emerge as superior models because they are "a standard above current fitting procedures". This section evaluates these claims.

Recall that the purpose of the proposed framework has been to demonstrate that two seemingly opposing theories (competitive vs. non-competitive selection) can be accommodated within a unified model. This claim does not itself require data fitting. The next step is to determine which model is the correct model of word production: purely competitive, purely non-competitive, or a hybrid (the current proposal). This is what must be verified with empirical data. The current framework makes it clear that the key issue is whether there is a fixed threshold or a flexible criterion. With regard to model fitting, neither EA nor SDT models are immune to the challenges inherent in working with production data (e.g., dealing with omissions). Those challenges aside, both DDMs and SDT models can be fitted to individual data (see example above), can discriminated between fixed threshold and flexible criterion models, and can model trial-by-trial shifts of the criterion (e.g., Brown & Steyvers, 2005). There are thus no inherent shortcomings in the SDT framework as far as model fitting is concerned. In light of this, we will compare the suitability of the EA models advocated by A&A over the current proposal for the purpose at hand.

EA and SDT are closely related frameworks for modelling decision-making behavior, but they have one prominent difference. The former models the process of evidence accumulation and decision-making from stimulus presentation to response generation, whereas the latter explains decision-making on a slice of data from a particular point in time. In other words, SDT is agnostic to the processes before the point at which a decision is made. Each framework has had success in explaining certain kinds of behavior. EA models have been very successful in modelling RTs in forced-choice tasks with rapid stimulus-response mapping (Ratcliff, Smith, Brown, & McKoon, 2016), while SDT models have been very successful in modelling memory and metacognitive judgments (e.g., Wixted, 2007). The suitability of the model thus depends on the area in which it is applied. Which one would be more useful in the context of word production?

If we agree that the purpose of modelling is to understand the underlying cognitive processes, it follows that a good model is a model that (a) has parameters that can be clearly linked to cognitive processes, and (b) captures the complex patterns in the data with as few parameters as possible (almost any model can capture almost any data pattern if enough parameters are added, resulting in overfitting). A neural network model (NNM) like Dell, Schwartz, Martin, Saffran, and Gagnon (1997) is an example of a useful cognitive model because it reduces the complex, nonlinear process of mapping meaning to sound to two cognitively meaningful parameters, corresponding to the strengths of semantic-to-lexical and lexical-to-phonological mappings. By contrast, EA models often have many more parameters. Ratcliff's original (1978) model had seven parameters, and some variations have more. More importantly, the interpretability of these parameters depends on the complexity of the task. As Ratcliff et al. (2016) aptly put it, the scope of most EA models is limited to modelling decisions that are "made rapidly and at a low level cognitively" (p. 260). It is possible to extend these models to more complex tasks successfully, but this requires merging them with other types of models like NNMs (e.g., the LCA model; Usher & McClelland, 2001).

When simple EA models are applied to cognitively complex processes like word production, non-trivial problems arise. The RT on each production trial is the product of a complex, non-linear process in a system with cascading and interactivity (see Dell, Nozari, & Oppenheim, 2014, for a review of the evidence for these). An EA model must unpack all of this information from the RTs in a meaningful way. This is a problem that does not exist for the type of quick, low-level judgments usually modelled using EA (Ratcliff et al., 2016). In addition, assumptions that may be valid for forced-choice tasks in simple serial systems, like positing a "non-decision" component which is independent of the "decision" components, are questionable for a complex non-serial process like word production. For example, EA models of word production have included motor processing in the "non-decision" time after a word has been selected (e.g., Anders et al., 2015; Anders, Riès, van Maanen, & Alario, 2017). However, Baese-Berk and Goldrick (2009) found longer voice onset times for the initial consonants of words that had a minimal pair with a voiced initial consonant, showing that articulatory-motor processes are not independent of the prior stages of processing.

These and similar problems lead to results which are sometimes sharply at odds with other types of models (e.g., NNMs). For example, the finding of *higher* accumulation rates in individuals with brain damage compared to healthy controls (Anders et al., 2017) is the opposite of what NNMs of damaged systems predict (e.g., Dell et al., 1997; Dell, Schwartz, Nozari, Faseyitan, & Branch Coslett, 2013; Nozari & Dell, 2013). If we accept, based on their ability to fit RTs, that EA models are superior to other model types, we must also accept that damaged systems transmit information *faster*. This would in turn require an entirely new cognitive theory of damage that, in addition to the new findings, would also account for all of the data that was compatible with the exact opposite theoretical perspective. More broadly, EA models succeed in fitting the data because they have many parameters, some of which (like those relating to variability across trials) naturally allow the models to accommodate a wide range of patterns in the data. However, the solutions obtained in this way are not necessarily unique. As discussed extensively in this literature, the same empirical pattern can often be explained in more than one way. For example, the relative speeds of correct and error trials could be explained in terms of assumptions about variability across trials, but they could also be explained by collapsing bounds (see Ratcliff et al., 2016, and references therein). These models are thus by no means immune to model mimicry.

In short, EA models do fit the entire distribution of RTs, but this may not be desirable in this case. Production RTs mark the end of a multi-stage process in a non-linear system with both cascading and interactivity, but the goal is to understand something about the middle of this process, i.e., lexical selection. By proposing that SDT be combined with the output of models that more appropriately capture the process of semantic to lexical mapping, we have attempted to avoid the above-mentioned problems. It is important to note that we do not mean to imply that EA models have no place in word production. However, it would be incorrect to claim that they are superior to other models because of the fit they provide to RT distributions.

Remaining comments

In this section, we discuss three remaining comments, relating to issues that we believe are critical for the correct application of the proposed framework—specifically, its scope and limitations.

1) The scope of a decision-making framework. We continue to emphasize that the dynamics of spreading activation in a semantic-lexical space are a big part of lexical retrieval. These reflect the speaker's knowledge structure; i.e., which features connect to each other to form a concept, which concepts are linked to each other through their features, which lexical items are linked to which features and to each other, etc. No generalized decision-making framework, including SDT, can or should replace specialized cognitive models designed to capture the dynamics of semantic-to-lexical mapping. Indeed, when seeking to understand the differences in behavior across tasks with different materials/demands, these dynamics should be the first line of investigation. Strong motivation for this approach comes from studies of concept processing, which have shown that the pattern of neural activity for a concept differs depending on task demands and context (e.g., Hargreaves, White, Pexman, Pittman, & Goodyear, 2012; Musz & Thompson-Schill, 2015). It is thus quite reasonable, for example, to expect verbs and nouns to be activated differently depending on which one must be produced in a task (M&N). There is, however, a longstanding tradition of separating language processing from other goal-oriented tasks. Current models of word production that include syntactic processing do not treat task demands as a key factor in shaping the activation in semantic-lexical space. They assume that nouns and verbs are activated equally up to the point at which selection occurs (Gordon & Dell, 2003). We suggest that this assumption may need to be revisited in the light of the evidence that task demands introduce early constraints on processing often way before decision points.

The scope of a decision-making framework, in contrast, is limited to explaining how selection decisions are optimized based on performance goals (i.e., prioritizing either speed or accuracy). For example, explaining a change from producing lots of quick commission errors to lots of omission errors in the same picture naming task within the same individual with aphasia would fall within the scope of such a framework. Similarly, two individuals with aphasia who produce lots of omission vs. commission errors in the same task may have differences in their decision-making criteria. On the other hand, explaining why the same individual responds better to verb primes than noun primes in a picture naming task, or why some individuals are better helped by noun vs. verb primes, is not a question for a decision-making model. As a rule of thumb, decision-making models apply when different patterns of performance must be explained *within the same task*. Appealing to these models to explain differences in performance between two different tasks would add a useless parameter with little or no explanatory power, and we discourage the use of the framework in this way.

2) Measure of conflict. M&A maintain that conflict measures derived from one vs. many competitors should behave differently. Our computational simulations (Nozari et al., 2011) do not support this claim. The reason is that co-activation in a large semantic cohort results in both higher activation of the closest competitor and higher total activation across all competitors. Both one- and many-competitor measures of conflict would thus be similarly affected, with the simpler measures being less affected by noise. We emphasize that our proposed framework is not committed to a specific measure of conflict. Any measure that has a consistently higher value when co-activated items have closer activation levels would work (but see Nozari, under review, for a criticism of one commonly used measure of conflict, Hopfield energy, which does not meet this criterion). We believe that our position is actually aligned with recent findings from M&A's group showing that the difference in activation levels, not the number of activated representations, is the determining factor in competition (Rose et al., 2019).

3) Criterion flexibility. M&N criticized the proposal for underestimating how flexible the criterion would need to be. While the example they cited is not the right one (for reasons described in the first point above), they are correct in stating that a criterion may change from trial to trial. Indeed, if language production is subject to the same kind of decision-making process as other cognitive systems like visual processing and memory, as both we and others have argued (Anders et al., 2015, 2017; Nozari, 2018; Nozari & Hepner, this issue), the selection criterion is highly likely to vary from trial to trial (e.g., Benjamin, Diaz, & Wee, 2009). These adjustments would be perfectly compatible with an incremental learning process, examples of which have been found in other production processes (e.g., Breining, Nozari, & Rapp, 2018; Oppenheim et al., 2010; Warker & Dell, 2006).

Conclusions and outstanding questions

Current behavioral data suggest that similarity-induced interference in picture naming is more than an illusion. The corresponding neural data show that the production system is indeed sensitive to the co-activation of similar representations. Neither of these findings, however, provides conclusive evidence for competitive selection models. We suggest investigating whether a system with a flexible criterion, which has been shown to provide a better explanation for performance in other cognitive domains, could also be more successful in explaining selection in word production. By no means does this imply that current models of word production should be replaced with a decision-making model. Rather, a decision-making framework can be superposed on the dynamics of such models. If a flexible-criterion model can provide a better account for the data than a fixed-threshold model, then the issue of competitive vs. non-competitive selection gives way to new questions: “How often and under what circumstances does the selection criterion change?”, “How do differences in learning affect criterion shifts?”, or “What conditions impair the adjustment of the criterion based on performance goals?”. This may, in turn, allow us to differentiate between problems stemming from impaired transmission of information between layers of representations, and those resulting from setting a maladaptive criterion for selection in individuals with aphasia.

Acknowledgments

This work was supported in part by the Therapeutic Cognitive Neuroscience Fund endowed to the Cognitive Neurology division of the Neurology Department at Johns Hopkins University. We would like to thank Rasha Abdel Rahman, Evangelia Balatsou, Albert Costa, Alissa Melinger, Brad Mahon, Eduardo Navarrete, and Gary Oppenheim for their thoughtful commentaries. We would also like to thank John Wixted for his insightful comments about the application of SDT to word production.

References

Anders, R., & Alario, F.-X. (this issue). Challenges to developing time-based signal detection models for word production. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2019.1581603

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

Aristei, S., & Abdel Rahman, R. (2013). Semantic interference in language production is due to graded similarity, not response relevance. *Acta Psychologica*, *144*(3), 571–582. https://doi.org/10.1016/j.actpsy.2013.09.006

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

Belke, E., Shao, Z., & Meyer, A. S. (2017). Strategic origins of early semantic facilitation in the blocked-cyclic naming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1659–1668. https://doi.org/10.1037/xlm0000399

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*(1), 84–115. https://doi.org/10.1037/a0014351

Breining, B. L., Nozari, N., & Rapp, B. (2018). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000630

Brown, S., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 587–599. https://doi.org/10.1037/0278-7393.31.4.587

Costa, A. (this issue). A flexible criterion of response selection: When is it needed? Comment on Nozari and Hepner (2018). *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2018.1532404

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). 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

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*(3), 380–396. https://doi.org/10.1016/j.cognition.2013.05.007

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

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, *27*(1), 1–40. https://doi.org/10.1207/s15516709cog2701_1

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*, 36–47. https://doi.org/10.1016/j.jecp.2015.09.008

Hargreaves, I. S., White, M., Pexman, P. M., Pittman, D., & Goodyear, B. G. (2012). The question shapes the answer: The neural correlates of task differences reveal dynamic semantic processing. *Brain and Language*, *120*(1), 73–78. https://doi.org/10.1016/j.bandl.2011.10.004

Mahon, B. Z., & Navarrete, E. (this issue). Adjudicating conflict in speech production – Do we need a central selection mechanism? *Cognitive Neuropsychology*.

Melinger, A., & Abdel Rahman, R. (this issue). To re-tune or not to re-tune: Comments on the flexible criterion. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2018.1562886

Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*, *76*, 41–51. https://doi.org/10.1016/j.neuropsychologia.2014.11.029

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

Nozari, N. (under review). A comprehension- or a production-based monitor? Response to Roelofs (2019).

Nozari, N. (2018). How special is language production? Perspectives from monitoring and control. In K. D. Federmeier & D. G. Watson (Eds.), *Psychology of Learning and Motivation* (pp. 179–213). https://doi.org/10.1016/bs.plm.2018.08.006

Nozari, N. (2019). The dual origin of semantic errors in access deficit: Activation vs. inhibition deficit. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2019.1587397

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a “lexical editor” be? *Journal of Memory and Language*, *60*(2), 291–307. https://doi.org/10.1016/j.jml.2008.09.006

Nozari, N., & Dell, G. S. (2013). How damaged brains repeat words: A computational approach. *Brain and Language*, *126*(3), 327–337. https://doi.org/10.1016/j.bandl.2013.07.005

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. (this issue). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2018.1476335

Oppenheim, G. M. (2018). The paca that roared: Immediate cumulative semantic interference among newly acquired words. *Cognition*, *177*, 21–29. https://doi.org/10.1016/j.cognition.2018.02.014

Oppenheim, G. M., & Balatsou, E. (this issue). Lexical competition on demand. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2019.1580189

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

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59–72. https://doi.org/10.1016/j.cognition.2017.12.007

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

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *64*(1), 1–34. https://doi.org/10.1016/j.cogpsych.2011.10.002

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. https://doi.org/10.3758/BF03194023

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 387–398. https://doi.org/10.1037/0278-7393.32.2.387

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Wixted, J. T. (in press). The Forgotten History of Signal Detection Theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*