Word production: Monitoring, control, and repair

Nazbanou Nozari[1,2]

[1]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA

[2]Cognitive Science Program, Indiana University, Bloomington, Indiana, USA

Corresponding author: Nazbanou Nozari

Address: 1101 E 10th St., Bloomington, IN 47405

Phone number: 812-855-5818

Email: bnozari@iu.edu

Lab website: https://www.nozarilab.com/

Twitter handle: @NozariL

Abstract

Research on word production is concerned with the process of turning a thought into motor movements that produce a word. This has traditionally been studied using two approaches, the psycholinguistic approach and the motor speech approach, which focus on different parts of the word production process. In this paper, I will show how highlighting the strengths of these two approaches, as well as merging them with broader frameworks and theories of action and cognition, can take language production research in novel directions. In doing so, I will discuss processes that complement language production, such as how speakers assess whether production is going smoothly (monitoring), adjust to its difficulties (control), and fix errors (repair). Each proposal combines what we know about language production with insights from other areas of cognition. Through these proposals, I will demonstrate the utility and necessity of a closer integration of broader cognitive frameworks into models of word production, as an important general direction for future research.

Keywords

Language production, monitoring, control, error repair, decision making, learning

Word production, monitoring, control, and repair

Introduction

The study of language production is concerned with how people turn their thoughts into speech that is executed through motor movements. Historically, word production has been studied using two approaches: the psycholinguistic approach and the motor speech approach[1–3]. Generally speaking, the focus of psycholinguistic models is mapping meaning to sound: How do speakers retrieve an arbitrary sound pattern (e.g., /kæt/) to refer to the furry pet? Motor speech models, on the other hand, focus on mapping sound patterns (e.g., /kæt/) to articulatory motor movements[3,4]. This difference in focus has caused the two approaches to remain largely separate. Moreover, with the exception of motor control research, which has heavily influenced motor speech models, neither approach has, traditionally, made close contact with more general theories of action and cognition.

Each approach has its strengths. The main strength of the psycholinguistic approach is its broader view of language production as a process beyond a mere motor act. The emphasis of psycholinguistic models on semantic knowledge (i.e., concepts) as the starting point of language production, naturally pushes these models to address key issues such as consequences of semantic similarity (i.e., overlap in meaning) and mechanisms that select one word among related words (how do I say *cat,* and not *dog,* when talking about the furry pet?). These issues, in turn, provide a natural bridge to theories of cognitive control, which are concerned with the selection of a response among competing alternatives, and theories of decision-making, which address the factors that affect implicit or explicit selection. Motor speech models, on the other hand, have two key advantages over the traditional psycholinguistic models of adult language production; they view the act of production as closely integrated with (a) monitoring and (b) learning. This close integration of production, monitoring, and learning naturally leads to viewing the language production system as a goal-oriented, self-organizing system, which adjusts itself to different situations and task goals, similar to many other cognitive systems[5–7]. By this, I do not mean to imply that learning has never been implemented in psycholinguistic models. There are plenty of models that use learning to explain specific phenomena[8–11]. However, as will be discussed in the next section, learning and monitoring are not integral parts of the main psycholinguistic models of word production.

My goal in the current paper is not to give a comprehensive overview of the models of word production, which exists elsewhere[12–14], but rather to highlight what can be gained by considering both the psycholinguistic and motor speech approaches and linking them with more general theories of action and cognition. I will do so by focusing on mechanisms of monitoring, control, and repair in production. Monitoring assesses whether production is on track to meet the production goals (e.g., communicating a message). This includes catching errors that may hinder communication. Control defines operations that help production proceed smoothly toward its goals. Repair refers to processes that change an utterance (usually an error) to a new utterance. The discussion of these mechanisms will, in turn, bring out the critical role of learning and decision making in the production process. Throughout the paper, I will use the word "learning" in its computational sense of changing the strength of connections between representations in the production system[15]. In this sense, learning is contrasted with changes to the activation of representations without affecting the system's connections. Note that this definition encompasses explicit and implicit learning, and has the feature of being more resilient against the passage of time and interference from unrelated representations than activation-based dynamics[16,17]. I will show that the resulting view offers new perspectives on current hotly debated issues, including the mechanisms underlying word selection and domain-generality/specificity of control processes that regulate language production.

Computational models of word production

Figure 1a is an overview of the representations involved in the entire production process[3]. Production starts from concepts and selects the appropriate word (lemma) to express that concept. If the word needs to be inflected (e.g., help → help<u>ed</u>), it is morphologically encoded and then sent for syllabification and phonological encoding, where its abstract metric and sound structure is put together. Next, it goes through phonetic encoding which prepares it for articulation. Figures 1b and 1c show the rough scope of psycholinguist and motor speech models. Most psycholinguistic models start with concepts but go no further than phonetic encoding. This is not to say that all psycholinguistic models have all these layers; rather, to emphasize that they are rarely concerned with articulatory processes. In contrast, motor speech models usually start with the phonological form or its equivalent in perceptual space and cover the later stages of production. To lay out the foundation for discussing monitoring, control, and repair processes that are the focus of this paper, I will briefly review the main models from the psycholinguistic and motor speech traditions, as well as one model that has attempted to merge the two.

Within the psycholinguistic tradition, mapping meaning to sound entails at least two distinct stages[1] (Fig. 1b): mapping semantic knowledge (represented as semantic features) to lexical items (i.e., lexical retrieval) and mapping lexical items to phonology (i.e., phonological encoding). The first step can be influenced by syntactic demands[1,3]. The second stage can continue on to phonetic encoding but this is skipped in many psycholinguistic models[18–20]. Two main computational models in the psycholinguistic tradition are Levelt et al. (1999)[3] and Dell (1986)[1] and their subsequent versions (see[12,14] for reviews). While the details and scopes of the two models and their variants differ, they mostly agree on the core representations and the main stages of processing. Importantly, both models emphasize the co-activation of related representations within each layer of the production system (e.g., lexical item *dog* for the target *cat*; Fig. 1a). Modeling such coactivation and its consequences has been one of the major strengths of the psycholinguistic tradition and has fostered several theories including competitive/non-competitive selections and domain-generality/specificity of cognitive control, which I will unpack in the following sections[8,21–25]. In addition, these models have been quite successful in explaining language impairment after brain damage or aphasia[19,20,26,27], capturing the development of word production abilities in children[28,29], and mapping language production processes onto brain regions[30–33]. Finally, psycholinguistic models can accommodate different modalities of language production, including written and typed production[34,35]. It is worth noting that despite their similarities, the two main psycholinguistic models differ in one key respect, namely, their assumption regarding modularity (see Box 1).

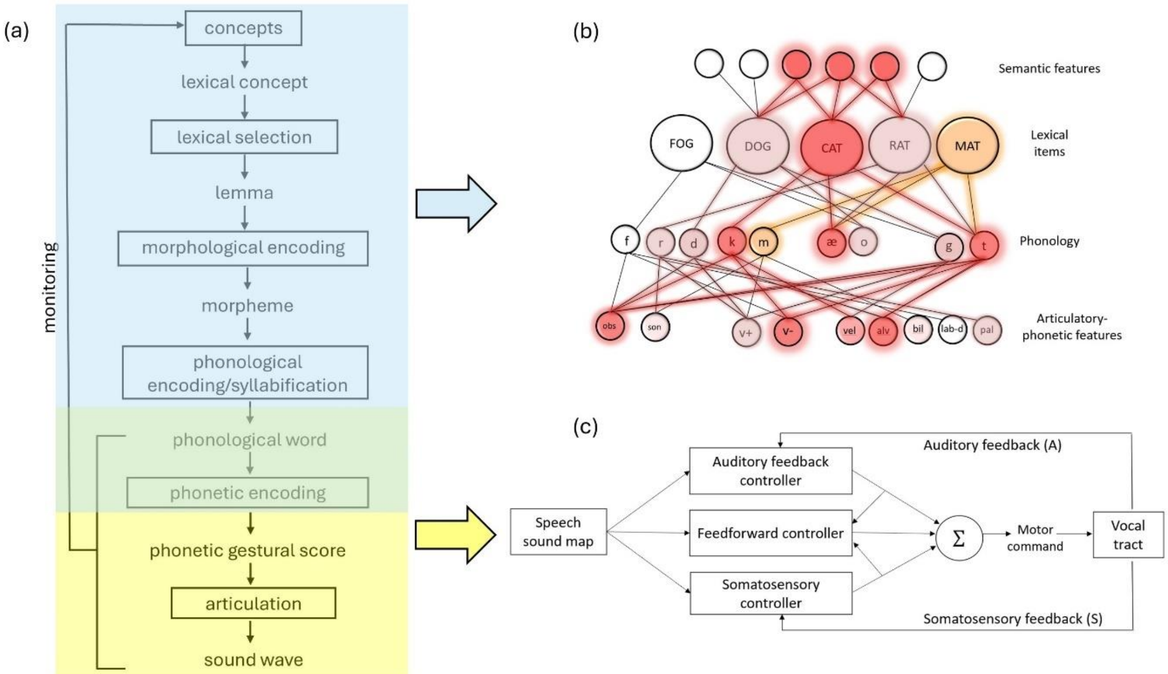

Figure 1. Models of word production. (a) The general architecture of language production from meaning to sound (adopted from[3]). (b) The two-step interactive model[1]. Schematic of spreading activation for a trial with "cat" as the target. Spreading activation to "cat" and competing representations are shown in darker and lighter shades of red, respectively. Phonologically related word "mat" is activated through feedback, and activates its own unique segments through feedforward connections (dynamic shown in orange). For illustrative purposes, only some phonetic features and partial feedback are shown. (c) A simplified version of DIVA (adapted from[4]). See text for model descriptions.

Box 1. Modularity or interactivity within the language production system

Modularity[170] refers to the encapsulation of information within each stage of processing. Psycholinguistic models agree on the multi-stage nature of word processing (Fig. 1a), but they differ on the issue of modularity. This debate concerns two phenomena, cascading and interactivity. In a system with cascading of activation, information from a higher layer leaks into lower layers, before the higher-level processing has been completed. In an interactive system, information that has cascaded to lower layers also feeds back to higher levels and influences processing in those layers. Levelt et al.'s (1999) model[3] does not allow cascading or interactivity and, consequently, is an example of a modular model, whereas Dell's (1986) [modle1] is a prime example of an interactive model. Figure 1a shows principles of cascading and interactivity when the speaker successfully produces the target "cat". Cascading is reflected in the activation of some of the phonology of the competing lexical items that were not ultimately selected (e.g., /d/ for dog)[171]. This is not expected in a model without cascading, because phonological activation only starts after a lexical item has been selected. Interactivity is shown in the activation of the lexical item "mat" through feedback from its segments /æ/ and /t/[96,172,173].

Several behavioral and neural findings now support non-modular models. For example, when participants named a picture as a couch, they were nevertheless faster at reading the following probe "soda", which was related to the alternative label sofa, compared to an unrelated word[171]. This finding shows cascading because to prime "soda", participants must have activated the alternative label (sofa) down to at least its phonological representations, even though they obviously did not select it at the lexical level. Similarly, cascading between phonemes and articulatory-phonetic features was supported by the finding that segmental errors (e.g., /k/ → /g/) had a voice onset time somewhere between target and error phonemes, showing that the articulated product was a blend of the two phonemes[174]. Interactivity has also found support in a number of behavioral findings, including the mixed error effect[1], the lexical bias effect[175] cf.,[176], and the repeated phoneme effect[1,34]. Neural data tell a similar story. Manipulating semantic similarity has effects on later processes, up to articulation, and manipulating segmental similarity affects earlier processes through feedback[60].

Although the evidence cited above supports non-modular systems, they should not be taken as support against distinct processing stages[177] for two reasons. First, due to the mostly arbitrary mapping between semantic features and sounds of words, a single-stage mapping achieved by direct connections from individual semantic features to individual sounds is not possible. Also, even the behavioral and neural evidence that supports cascading and feedback points to a system that retains some modularity[60,164]. The modularity debate is not just of abstract theoretical interest[5]; it has important consequences for the application of theoretical models to new data. For example, interpreting EEG components or fMRI activity as reflecting a specific process based on a timeline derived from a serial (modular) model[31,178] can be problematic, because such an approach neglects the fact that the current component/neural region is influenced by processing in layers before and after it (see[179] for a critical review of this approach and its problems).

The main model from the motor speech tradition is Guenther's directions into velocities of articulators (DIVA) and its later version, gradient order DIVA or GODIVA[2,36,37]. The model is essentially a forward model of motor control (Fig. 1c). Forward models predict the consequences of motor commands through internal simulations of movements[38–40]. The model starts by activating the sound of a word in the speech sound map. A speech motor command is sent to the articulators to produce the word. Simultaneously, the perceptual consequences of this motor command are anticipated in the form of auditory and somatosensory predictions. Once the word has been spoken, the actual perceptual auditory and somatosensory consequences are compared to the anticipated ones. If there are any discrepancies, an error signal is generated, which is used by an inverse model to adjust future motor movements. DIVA and its variants explain a wide range of findings in the acoustic-articulatory part of the language production system. These include learning to produce novel sounds and adjusting speech based on altered auditory or somatosensory feedback, as well as disorders that affect sublexical representations, such as stuttering and apraxia of speech[41–44]. Moreover, these models have allowed for a detailed neural mapping of speech planning and execution processes to both cortical and subcortical regions[4,45,46].

To date, no computational model has preserved the sophistication of GODIVA's account of articulation in a model that also addresses psycholinguistic issues, such as lexical access from meaning, but some models have incorporated elements from both. An example is Hickok's (2012) hierarchical state feedback control model[47]. The model is similar to DIVA in assuming that perception controls production but changes the starting point from the speech sound map to the psycholinguistic concept of lemma[3]. The lemma hierarchically activates syllables and phonemes, each of which have their own perceptual control loops (auditory and proprioceptive, respectively). However, unlike DIVA, the hierarchical state feedback control model does not depend on overt consequences of action. Rather, lemma activation is hypothesized to activate both motor and perceptual representations, with the former suppressing the latter. A correct production of the target extinguishes the activation of the target's perceptual representation, but an error fails to do so, leading to the generation of an error signal. In proposing an internal check process independent of overt perceptual feedback, the model appeals to psycholinguistic notions such as "inner speech"[48], while maintaining a core assumption of motor speech theory, i.e., the reliance of the speech production system on the perceptual system. However, the model does not implement the first stage of processing, i.e., semantic-to-lexical mapping.

As shown in this brief overview, psycholinguistic models pay much attention to the coactivation of representations, while motor speech models highlight the interaction between perception and production systems. In the next section, I will review how this differential emphasis can be leveraged for monitoring and control of the language production system.

Monitoring

As noted earlier, one of the strengths of motor speech models is that monitoring and control are integral to the production process. As seen in Fig. 1c, every act of production is accompanied by parallel activation of sensory representations that act as a "check" on production. Psycholinguistic models, on the other hand, do not pose an integral monitoring mechanism. Instead, for example, in Levelt et al.'s (1999) model, the language comprehension system is in charge of monitoring production. This account, called the "perceptual loop"[49], is elegant in its assumption that the same system used for monitoring the speech of others, is also used for monitoring one's own speech[50]. The proposal has an outer loop and an inner loop. The outer loop is simply speech comprehension, the contribution of which to monitoring can hardly be denied; we hear ourselves all the time and such prominent input can hardly be ignored. The nature of the inner loop has been long debated and criticized[51–53]. More recently, Roelofs (2020)[54] has redefined the perceptual loop as connections between representations in the production system and their corresponding representations in the perceptual system. This redefinition makes the perceptual loop account more similar to that of speech motor models.

Monitoring production through perception works well when production and perceptual representations are distinct. This is certainly true for the lower post-lexical aspects of production (articulatory-phonetic representations) and perception (acoustic representations), and possibly also for phonology[20,55]. At higher layers of the production system, i.e., lemmas and higher (Fig. 1a), there is less motivation for, and evidence of, distinct production and perceptual representations. For example, it is unclear why there would be two sets of lemmas, one for perception and one for production, and even less clear why there would be two sets of semantic or syntactic features[56,57]. Yet, problems can also arise in those layers. One possibility is that higher-level representations are only monitored through the implementation of their sensory-motor representations. This would imply that errors arising during semantic-to-lexical mapping (e.g., *cat* → *dog*) and those arising during lexical-to-phonological mapping (e.g., *cat* → *dat*) are both detected based on the same sensory-motor representations. However empirical evidence suggests a double dissociation in the detection of semantic and phonological/phonetic errors, as well as distinct neural correlates for monitoring these two error types[52,58,59]. Also, EEG studies in both linguistic and non-linguistic tasks have revealed an early negativity (too early to be compatible with sensory-motor comparisons) in trials with a higher, compared to lower, error likelihood[60,61]. It thus seems that speakers are equipped with a mechanism to detect the likelihood of an error early on, perhaps before it has even occurred. One way to model the early detection of errors is the conflict monitoring account.

The conflict monitoring account[52,61,62] proposes that production errors are detected by indexing the level of conflict in the activation of competing representations. When one item is much more activated than others, conflict is low, and so is the probability of an error (Fig. 2). Conversely, when two or more items have comparable activation levels, conflict is high, and errors are more likely. Importantly, conflict monitoring can happen at any stage of production, and is not bound to sensory-motor representations. Also, it can be easily implemented in psycholinguistic activation models, which emphasize the coactivation of competing alternatives.

But what level of conflict should be considered "high"? Answering this question requires using a decision-making framework. Figure 2a shows the application of signal detection theory's (SDT)[63] framework to the distribution of conflict for correct and error trials. SDT has been one of the most influential theories in psychology. In its original form (SDT-I), it contains distributions of signal and noise and models decisions attributing a stimulus to one of these two distributions. Imagine being asked to detect a tone amid noise. Within SDT, your choice is affected by two parameters (Fig. 2a). Discriminability or *d'* determines how separable the two distributions are. The higher the *d'*, the more accurate the choice. Criterion or *c* indicates a threshold (a point somewhere on the distributions) your system defines to label a stimulus as signal or noise. This results in four types of responses. If a stimulus belongs to the signal distribution and is above the criterion, it is correctly labeled as a signal (Hit). If the same is below the criterion, it is erroneously labeled as noise (Miss). Similarly, if a stimulus belongs to the noise distribution and is below the criterion, it is correctly labeled as noise (Correct Rejection). If the same happens to be above the criterion, it is mistakenly labeled as signal (False Alarm).

SDT-II[64] applies the same framework to metacognitive judgments. Now judgments are not on external stimuli but on cognitive choices themselves. In language production, instead of distributions of signal and noise, we have distributions of conflict for error and correct trials. D' depends on the state of the production system and the nature of the task. Healthy mature production systems have a high *d'*. Damaged (e.g., post-stroke), immature (e.g., child), or lightly trained (e.g., L2 speakers) systems have lower *d'*s[52,65,66]. This usually means that correct trials in these systems are also associated with higher amounts of conflict, making the two distributions less distinguishable. Criterion *c* depends on task goals[6]. If accuracy matters, *c* is set conservatively (low misses at the cost of higher false alarms). If speed and fluency matter more, *c* is set liberally (low false alarms at the cost of higher misses). False alarms, although uncommon, can be observed in individuals with aphasia as changing a correct response into an error[52,67], and in neurotypical speakers, as disfluencies[68].

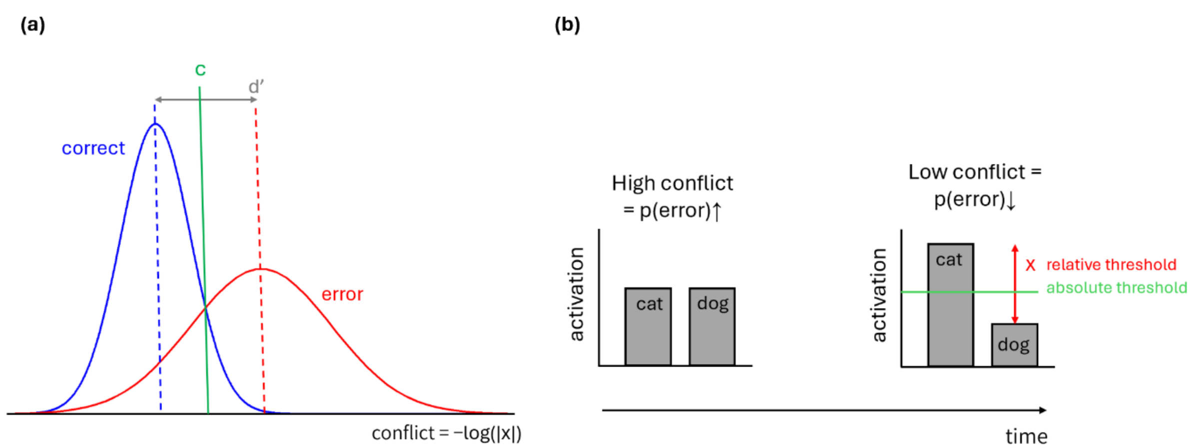

Figure 2. The relationship between conflict and response selection. (a) Hypothetical distributions of conflict for correct and error trials, within the signal detection framework. D' *reflects how cleanly separable the two distributions are. Criterion (c) can be put anywhere on the axis of conflict. Shifting c to the left makes behavior more conservative (fewer misses, more false alarms), whereas shifting c to the right makes behavior more liberal (fewer false alarms, more misses). (b) Change in conflict over time. A high-conflict trial can turn into a low-conflict trial over time (through spreading activation in neural networks or evidence accumulation in decision models). Conflict is inversely related to the difference between the activation of two (or more) items (x in the right-most figure). A competitive selection mechanism critically depends on reaching a certain*

value of x (a relative threshold). Non-competitive selection only needs an absolute threshold, irrespective of x (green).

For selection, the decision rule is: *if conflict < c, then select; otherwise, wait* (until conflict gets smaller with further spread of activation; Fig. 2b). For monitoring, the decision rule is: *if conflict < c, then pass as correct; otherwise, detect as an error*. The selection and monitoring rules only differ in *when* they apply, with selection tending to occur earlier. In an ideal world, conflict monitoring would always operate before selection, preventing all errors. However, this could lead to silences and pauses that speakers may wish to avoid, leading them to sometimes commit to responding before checks are complete. This would manifest as overt errors needing detection and repair, although the rate of overt errors remains low in healthy mature systems[69].

To summarize, applying SDT rules over distributions of conflict allows us to understand selection, monitoring, and error detection within the same framework. In a similar way, SDT can be used to determine the parameters of error detection within other frameworks, although the nature of the distributions depends on the information that those theories consider central to monitoring. Two issues merit further discussion: the first issue is whether selecting the word *cat* depend on how activated *cat* is per se, or how much more activated *cat* is compared to *dog* (Fig. 2b)? This question is at the heart of the mechanisms of lexical selection. The second issue concerns better computational approaches to understanding the first issue. Boxes 2 and 3 address these two points, respectively.

[Boxes 2 and 3 about here]

Towards a more comprehensive model of monitoring

There is an ongoing debate about which monitoring mechanism is the right one[53,54]. My own view is that language production is monitored through a combination of mechanisms. The original perceptual loop model assumes a conscious deliberate monitoring through the comprehension system[49]. Despite attempts at dialing back the role of conscious awareness in the more recent versions of the perceptual loop[54], I think the emphasis on conscious and deliberate processing is actually a great asset of the original theory. It affords monitoring a place for taking into account the speaker's broader knowledge of the world. In fact, the role of deliberate monitoring extends to monitoring the audience for signs of incomprehension and actively using world knowledge and common ground to facilitate communication.

Complementing these conscious processes are implicit mechanisms that have the advantage of being fast and effortless. Forward models are excellent candidates for this purpose at the sublexical level, i.e., for the detection of phonological/phonetic errors or adjusting acoustic properties of speech[70–74]. Classic forward models like DIVA, however, monitor the overt consequences of behavior, precluding them from detection of *potential* errors. Moreover, speakers can detect errors in their silent, unarticulated inner speech, which has no overt perceptual consequences[48]. In these cases, detection mechanisms such as those proposed in the hierarchical state feedback control model or conflict monitoring provide a better explanation. Finally, models that hinge on sensory-motor interactions are not the best candidates for monitoring more abstract representations, while empirical evidence shows that such representations can indeed be monitored. Conflict monitoring proposes a solution for these cases. Note that both forward models and conflict monitoring follow the same fundamental principle of using information to predict the outcome. As such, rather than thought of as opposing theories, they may be better conceived of as complementing theories. Box 4 discusses these theories within the broader umbrella of the reinforcement learning theory[75].

Box 2. Is lexical selection competitive or not?

When we intend to say *cat*, does it matter how activated *dog* is? Much research has shown that increasing semantic similarity, especially taxonomic similarity (cat/dog), between a target and its context interferes with production (see [179,180]for reviews). Many such claims hinge on results from the picture-word interference (PWI) paradigm[181], where participants must name a picture ignoring a distractor word that usually appears in written form on top of the picture. PWI has been criticized[23,182] for its many complexities such as multimodal processing that inevitably involves other systems (e.g., reading) and the need for suppressing a prepotent response, which is not the norm in speaking. However, semantic interference has also been robustly demonstrated in other paradigms, such as blocked cyclic naming[183] where participants repeatedly name a small set of items, and its more ecologically valid variant, continuous naming, where people simply name a sequence of pictures[17,177]. For decades, such interference was taken as evidence for "competitive lexical selection"[3,21,24,184]. The idea is that lexical selection can only proceed after a relative threshold, i.e., a minimum difference between the activation of target and competitor(s), has been reached (x in Fig. 2b). Opposing this idea, some researchers showed that production can be facilitated in semantically similar contexts, especially for thematic relations (bone/dog). This finding was taken as evidence for "non-competitive lexical selection"[23,185,186]. The idea here is that an item can be selected as soon as it passes an absolute threshold, without being affected by the activation of competing representations (see Fig. 2b).

Three points are noteworthy regarding this debate. First, while some computational models explicitly depend on competitive selection to explain interference and facilitation (e.g.,[24] and variations), others do not[8,9,144]. Importantly, the latter type relies on incremental learning mechanisms to explain these effects[8,9,144,172,187]. One such mechanism is error-based learning[9].

When *cat* is the target of production, its shared semantic features with dog also activate the word *dog*. At the end of production, error-based learning mechanisms strengthen the connections between the target (*cat*) and its semantic features while simultaneously weakening the connections between the competitor (*dog*) and the features it shares with the target. This learning mechanism facilitates the retrieval of *cat* when it next becomes the target (i.e., repetition priming[188]), but interferes with a subsequent production of *dog*. A similar mechanism can explain interference in segmentally related contexts[172,187]. Adopting the learning account has two advantages: (a) it explains the longevity of interference induced by contextual similarity[17,183]. Note that other learning models that do not weaken the competitor's connections can explain long-term priming, but do not explain long-term interference without additional mechanisms[188]. (b) When combined with a more accurate representation of semantic relationships, it naturally accommodates the seemingly disparate facilitation and interference effects induced by thematic and taxonomic relations, respectively, which have been previously taken to imply mutually exclusive selection rules. Specifically, combining incremental learning mechanisms with the gradual activation of themes, captures thematic facilitation and taxonomical interference, irrespective of whether the model's selection rule is set to competitive or non-competitive[8].

The second point worth noting is that the models that explain similarity-induced interference via learning do not necessarily deny the existence of mechanisms that impose some degree of competitive selection, such as lateral inhibition. In fact, given the prevalence of inhibitory interneurons and lateral inhibition in cognitive systems[189,190], it would be strange to insist that the language production system is devoid of such connections. Acknowledging lateral inhibition naturally implies some competition in the system, which also motivates and constrains the use of measures such as conflict (i.e., conflict is not just measured between any two arbitrary representations, but between those with mutual inhibitory effects). However, the existence of lateral inhibition does not automatically endorse competitive selection accounts. It is theoretically possible to have lateral inhibition and still select a representation before lateral inhibition has driven down the competition below a fixed value of x (Fig. 2b), which brings me to point 3.

The third noteworthy point is that the debate assumes a dichotomy: Selection is either competitive or not. My own view is that such a binary view is incompatible with the workings of a highly adaptive goal-oriented system such as language production[6,7]. I believe that the concept of a relative threshold x is useful, but this threshold can vary to ignore or highlight competition depending on the situation and task goals[191]. For example, imagine I am talking to a friend about the chance of rain in the afternoon, and the words "likelihood" and "probability" both come to my mind. I would output either one quickly to keep the conversation going, knowing that the slight difference in meaning would not alter my message. This is the equivalent of setting a low value of x for the relative threshold. However, if I am teaching statistics and these two words come to mind, I would pause and deliberate to make sure one is the clear winner for the concept I am about to express. This is the equivalent of setting a high value for x. Importantly, I can change x flexibly as my goals for speed vs. accuracy change. This is the idea of a flexible criterion[6,7,191].

Box 3. Beyond SDT

While SDT provides a simple and useful framework for understanding the relationship between conflict, selection, and error detection, other kinds of decision models can also be applied to language data and have certain advantages over SDT. An example is evidence accumulation models (EAMs)[192,193]. I have dedicated a box to them here because (a) they have already been used to model certain aspects of language processing[167–169], so it is important to be aware of some of their basic features, and (b) they can be an excellent tool for examining competitive vs. non-competitive selection mechanisms (Box 2).

Similar to SDT, they model choice based on evidence, but they also consider time, allowing evidence to accumulate gradually. The parameters of EAMs can be estimated through the quantitative fitting of choice-response time distributions. Similar to *d'* and *c* in SDT, EAM's *drift rate* and *response threshold* parameters are related to the quality of evidence, and the selection criterion, respectively. Mapped onto the lexical selection process, drift rate is akin to how quickly and strongly lexical items gain activation from semantic features. Response threshold is set by the selection rule. Given the close correspondence between EAM's parameters and dynamics of activation and selection, these models appear to be a good tool for examining the mechanisms of lexical selection. But which model is appropriate model for this purpose?

EAMs come in many different flavors. Some, like the linear ballistic accumulator (LBA)[194] and racing diffusion model (RDM)[195], are inherently non-competitive, i.e., evidence for one choice is not evidence against another. In contrast, other variants such as the diffusion decision model (DDM)[196] are inherently competitive, with evidence for one choice always counting as evidence against the other. Fully competitive models like the DDM are often limited to binary choices (although extensions to multiple choices do exist[197]), whereas accumulator and racing diffusion models more naturally accommodate multiple choices. At first glance, it appears that none of these models is well-suited for addressing the issues raised in the selection debate (Box 2): being limited to two choices is incompatible with the many linguistic choices faced by speakers, and utilizing models that naturally accommodate more choices means committing, a priori, to non-competitive selection.

However, while these models are sometimes used in their simplest form, they can also be used as building blocks for more complex models. For example, although LBA is not inherently competitive, competition can be introduced into the model in several ways. One such way is to set up the accumulators to represent relative information, e.g., the presence of one stimulus over another[198]. Another way is to build inhibitory connections between the accumulators[199] (see also leaky competing accumulator[200]). The resulting models can capture competition effects without being restricted to two choices. Customized modeling of competition as inhibitory weights has an additional advantage: the inhibition parameter (i.e., the value(s) of the inhibitory weights) can vary freely and continuously. This continuity allows the model to go beyond a binary competitive/non-competitive distinction, and capture graded competition effects, something that fully competitive models like the DDM cannot do. Such an approach is well-suited for investigating mechanisms of lexical selection and criterion setting, including the degree of competitiveness, its timeline during the evidence accumulation process[201], and its possible fluctuations across individuals, conditions, and even trials.

Despite this clear promise, challenges remain in applying EAMs to language data. For example, EAMs usually require a large amount of data to generate stable and recoverable parameters. Moreover, assumptions such as a constant rate of evidence accumulation within a trial, often made in the LBA, are probably too simplistic for the complex production process. Luckily, many such issues are also of great interest to decision scientists, and new solutions are being continuously investigated. For example, new Bayesian approaches allow for better fits and model-check routines with sparser data[202], and a newer variant of LBA employs a non-stationary process of evidence accumulation[203].

In short, EAMs are promising models for studying a range of issues related to selection, monitoring, and control in language production, however, their application to linguistic data can be complicated. This challenge can be overcome by a closer collaboration between language and decision scientists, and will likely benefit both fields.

Box 4. Reinforcement learning, conflict monitoring, and forward models

Are conflict monitoring and forward models in opposition to one another? In my mind, these are not mutually exclusive theories. In fact, both fit quite well within a broader framework of a more general, well-implemented, and biologically plausible theory of human behavior, the reinforcement learning (RL) theory[75]. RL's main claim is that humans learn from the consequences of their behavior; positive outcomes reinforce a behavior, whereas negative outcomes discourage the behavior. This framework is naturally well-suited to monitoring, as the point of performance monitoring is precisely to ensure that desired outcomes are reached. Hierarchical versions of RL can include a sequence of actions to model complex tasks, e.g., cooking a meal. Such models allow the agent to predict the final outcome from earlier stages. A model-free RL simply links actions with rewards. In a model-based RL, some internal representation of an action is used to predict action outcomes. Below, I will discuss how conflict monitoring and forward models of language monitoring fit the framework of model-based hierarchical RL.

Many of the just criticisms raised against conflict monitoring take issue with "the brain as a conflict detector"[204] (see [81]for a review). A good example of this is the debate between RL and conflict monitoring on the role of the anterior cingulate cortex (ACC)[61,204]. Both theories can explain the ACC-generated EEG marker of error commission, the error-related negativity (ERN). Conflict monitoring also explains an earlier EEG signature (N2)[61] observed in high-conflict trials, regardless of their outcome, but has been criticized for not explaining a late EEG signature in response to external feedback[204]. One way to tackle this is to show that a neurobiologically plausible computational model of RL generates a pattern that is well-aligned with the predictions of conflict monitoring, showing that the two are not in opposition[205]. However, and much more importantly, conflict monitoring, at least as I am using it in this paper, is fundamentally not a theory about neural processing. Nor does it mean to replace memory and learning processes[206]. It simply proposes computations predictive of performance outcomes that are themselves products of learning, memory, and other cognitive processes (see a similar dissociation between quantum computation and quantum brain[207]).

Insofar as conflict computations are good outcome predictors, the theory is useful. For example, we have shown that conflict is a reliable index of a language system's health and maturity, which in turn predicts the probability of speech errors[52,65]. Naturally, if a situation is created in which conflict is irrelevant or conversely related to performance outcomes, there is no good reason to expect conflict monitoring to be useful, but the main point is that such situations are not the norm in cognitive systems. Moreover, conflict monitoring is successful precisely because it presupposes a trained internal model (e.g., semantic → lexical item → phonology → articulatory phonetics, in case of language production), shaped systematically and hierarchically such that information in earlier stages is predictive of the final outcome, exactly as presumed by model-based RLs. However, the theory does not specify how such a hierarchy has come to be.

Forward models provide mechanistic explanations for how hierarchies of actions are shaped through feedback from their sensory consequences (hence my emphasis on the integral role of "learning" in these models). Once formed, the models predict the consequences of motor commands through internal simulations of movements[38,39], as in model-based RL. The idea of internal simulations of performance used as predictions is well-supported in biological sensorimotor control[40]. The key issue in applying these models to language production is implementation. A useful model makes precise and testable predictions. This is the case for DIVA/GODIVA, which I have discussed in some detail in this paper. Other proposals exist for expanding the role of forward modeling to all layers of the production system[208], but this model has been justly criticized for its ambiguity and lack of clear testable predictions[209]. An implementation of such models, if successful, would be a valuable contribution to the language monitoring literature.

In short, conflict monitoring and forward models, as computational rather than neural theories, both hinge on predictive processing and outcome evaluation within a task hierarchy, which is the tenet of the hierarchical model-based RL. Opposition only arises when conflict monitoring is taken as a core mechanism operating regardless of performance outcome, in which case, the theory loses its sense and value. Also, note that the neural circuitry underlying reward processing in RL theories includes regions like the medial prefrontal cortex[210] and cerebellum[211]. As such, uncovering the involvement of these regions in language monitoring is not unique support for forward models, but is compatible with any model that works within the RL framework.

In summary, I believe that language is monitored through multiple mechanisms, including conscious deliberate processes that adjust communication based on speaker's world knowledge and their goals, as well as implicit and largely subconscious processes that use a variety of information at different stages to predict performance outcome. This multi-process monitoring view[53,76] finds support in empirical evidence, such as bimodal distribution of latencies related to error detection[77] and distinct neural signatures for errors detected with full and partial sensory feedback[78]. It is also compatible with computational models of error detection[79], and is aligned with the well-known redundancies in monitoring other functions. An example is gait and posture[80]. The control of normal gait and posture taps into multiple channels of information, including somatosensory, visual, and vestibular sensations. When deviating from the routine, e.g., walking on rough terrain, additional processes are required for gait adjustment. These include cognitive processes that represent the agent's knowledge of their body and motion in space. Importantly, certain parts of the neural circuitry underlying gait and posture control, such as the cerebellum, are involved in all processes that control gait, similar to what has been argued in Box 4 for language monitoring. Luckily, recent advances in the closer integration of monitoring processes with core production processes means that attempts to merge psycholinguistic and motor speech models will also bring together various monitoring mechanisms and nudge us closer to the implementation of the multi-process monitoring view.

Control

Monitoring provides valuable information, but such information is only useful when it can be used to regulate the production system. I refer to this as cognitive control or simply "control". Generally speaking, cognitive control refers to operations that allow humans to behave flexibly in a goal-directed manner[81]. Several taxonomies have been proposed, of which Miyake's division of these functions into prepotent response inhibition, task shifting, and working memory updating is probably the most well-known[82]. But this taxonomy was based on experimental tasks already identified as tapping cognitive control, which leaves many real-life situations undefined. For example, what about the inhibition of highly activated competitors in language production? Do we need cognitive control to suppress "dog" when meaning to say "cat"? Is an agreement error like

"The snake next to the brown lions are green." due to the failure of inhibitory control? Errors resulting from competition of a co-activated representations can be identified at all levels of language production, and there has been a movement in the field to link them to failures of inhibitory control [67,83–87]. This leaves open the question: is there a single "domain-general" inhibitory control ability underlying all tasks and domains? [88–92]. Alternatively, control may be specific to domain and even task [22,93,94]. This is the question I will address in this section.

Is all inhibitory control the same?

Competition between the target and competitors in the language production system can create one of two scenarios: (a) cases where the stimulus-driven (bottom-up) information is sufficient to arrive at the correct response, and (b) cases where such information alone leads to the incorrect response. An example of (a) is naming a picture (e.g., *dog*) after having named a similar picture (e.g., *cat*)[95–98]. Although the similarity creates interference (Box 2), the information provided by the stimulus, i.e., the current picture, is aligned with the task goal of naming that picture. An example of (b) is Stroop-like tasks, where a prepotent response must be suppressed in favor of a less potent response, for example when for social reasons, one must produce a flattering word to describe a disliked object. Here, the stimulus-driven information would lead the speaker to utter a word that is incompatible with the task goal of being polite. Figure 3a shows an example of these two situations, implemented within an experimental paradigm with only two items per block (e.g., cat/dog)[60,97]. In one condition, the task goal is to name the current picture, but the similarity between the current and the previous item creates behavioral interference (see Box 2). In another condition, the task goal is to reverse the names of the two pictures (i.e., say *cat* when you see the dog and vice versa), creating a Stroop-like effect. While both situations can trigger the monitor to detect an increased need for control, there are both theoretical and empirical reasons to doubt that they recruit comparable control mechanisms.

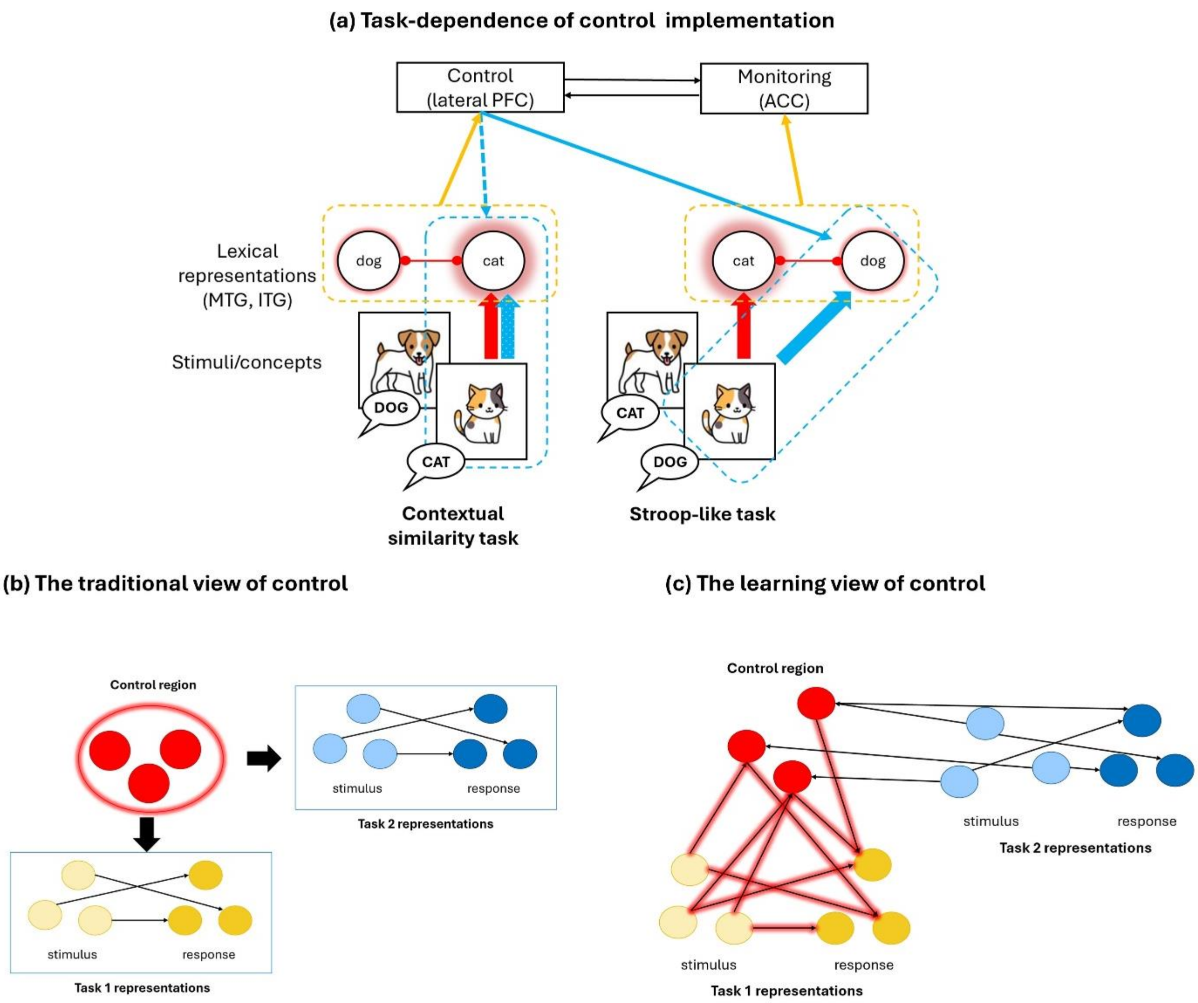

Figure 3. Control implementation. (a) Task-dependency of control. On the left, the task is to name the current picture. External control (blue arrows) is optional because it will push processing in the same direction as bottom-up information (red arrow). On the right, the task is to use the other label for naming the current picture. Since bottom-up information is insufficient to arrive at the correct response external control is necessary. (b) Traditional view and (c) Learning view of cognitive control (adapted from[145]*). Both accounts assume domain-generality in neural correlates*

of central control (red circles), but different mechanisms for control implementation, and therefore different predictions regarding domain-generality in application. See text for details.

Theoretically, external control is not required to achieve the goal when bottom-up information biases processing towards the correct response. When the goal is to name the current picture “cat” (Fig 3a, left), spreading activation from cat’s semantic features activates the word “cat”. As the word “cat” gains more activation, the local inhibitory link to the word “dog” suppresses the latter (lateral inhibition). If “dog” is very active, this suppression may take some time, but has a high likelihood of ultimately ending in the correct response. Now imagine a Stoop-like situation, where the stimulus is the same, but the speaker’s goal is to say “dog” (Fig 3a, right). The dynamics explained above will not achieve this goal; the speaker will end up saying “cat”, as shown in by the red arrow. In such a case external control is required to bias processing based on the task goal[99] (blue arrow mapping the stimulus to the goal-directed lexical representation). In short, the implementation of external control is essential for producing the correct response in one case, but not the other, pointing to possibly distinct control mechanisms in these two situations.

In keeping with this theoretical division, behavioral evidence shows that the magnitude of interference in the contextual similarity task is sensitive to the layer in which competition is induced (semantic vs. phonological), but the magnitude of Stroop-like interference is not[60,86,97]. This finding points to a more local inhibitory mechanism in the former than the latter case. Neural evidence is also generally aligned with this position. Both situations activate medial PFC, such as the anterior cingulate cortex (ACC), which is known to be involved in monitoring[100–104], however, lateral PFC, which is often considered to implement control in collaboration with ACC, is consistently implicated in Stroop-like tasks, but not in tasks that manipulate contextual similarity[103,105,106]. Further highlighting the differences in these control mechanisms are findings of distinct electrophysiological signature for Stroop-like and similarity-induced interference[60].

In short, task demands and their relation to stimulus-driven information determine whether lateral inhibition between local representations (e.g., middle and inferior temporal gyrus; MTG, ITG for lexical representations[32,107,108]) is sufficient to resolve the conflict or whether central control (through regions such as the prefrontal cortex or PFC) is required. This distinction does not mean that external control cannot possibly be involved in situations where bottom-up cues are sufficient. In such cases, external control can still help facilitate faster response generation by strengthening stimulus-response associations in such conditions[91,109–111], but this involvement is not necessary in the same way as in Stroop-like tasks. The optional involvement of external control for optimizing performance in such cases explains why PFC is sometimes implicated in tasks that manipulate semantic similarity, and sometimes not, whereas MTG and ITG are implicated much more consistently[103,106,108,112].

The discussion above showed that not all situations with high competition require external control. But are tasks that do require external control, e.g., varieties of Stroop-like tasks, controlled by a domain-general control mechanism? Here, I find it useful to define "domain-generality" more carefully[113,114]. Domain-generality can have multiple facets. Evidence suggests that processing principles are often domain-general, as long as they are not inherently incompatible with the nature of certain representations[115]. For example, the principle of performance adjustment after encountering a high-conflict trial applies equally well to both linguistic and non-linguistic domains[116–118]. Neural representations can also be domain-general, in the sense that they may carry out the same set of computations over representations in different domains. Association cortex and certain subcortical regions have been implicated as candidates for domain-generality of control

[90,119–124]. The real question, however, is whether domain-generality in computations and neural correlates translates to domain-generality in application.

Domain-generality in application assumes that because the same population of neurons is involved in mediating control across two tasks, activating that population through increasing control demands in one task will lead to the better implementation of control in the other task[88,125]. An extension of this account through merging it with neuroplasticity[126] predicts that training control using one task should improve the implementation of control in other tasks[127,128]. Empirical evidence for domain-generality in application has been mixed, with some papers claiming transfer of control from one task to another, mostly in language comprehension[125,129–131], whereas the bulk of studies from the cognitive control literature point to task and domain-specific control[93,99,132–138].

In short, in terms of general principles and neural correlates, cognitive control shows evidence of domain-generality across linguistic and non-linguistic domains. However, there is an active debate on the domain-generality of cognitive control in application. I will discuss this controversy in light of a new theory of cognitive control, the "learning account", in the next section.

The learning account of control

Before I explain the learning view, I will say a few words on what motivates the learning view of control. As explained earlier, while the main psycholinguistic models of language production do not incorporate learning as a core feature, much evidence suggests that they should. The evidence comes from the success of smaller models in explaining various behavioral effects through incremental learning. For example, several studies have shown that adult speakers rapidly and implicitly learn new phonotactic and orthotactic constraints embedded in strings of nonwords that they recite[11,55,139–142] (see[143] for a review). Integrating learning mechanisms into production also provides an elegant explanation for facilitation and interference effects of contextual similarity[8,9,144] as explained in Box 2. In short, a growing body of evidence suggests that action/production and learning are inseparable. This is the idea behind the learning account of control.

Figures 3b and 3c contrast the traditional and learning views of cognitive control[99,145,146]. Note that both views accept the domain-generality of neural correlates of control[92,123] (red circles). However, they make opposite predictions regarding domain-generality in application. The traditional account assumes that the key element of control is the activation of the control region. Once activated through any task, this activation benefits the current task, but since it is separate from the task itself, it can also benefit any other task, regardless of the overlap between the two tasks[129,131]. In contrast, the learning view posits that the key element in control implementation is the strengthening of the connections between the control center and task-specific representations. Since these connections are, by definition, task-specific (Fig. 3c, red connections), no benefit is expected for a new task. If anything, certain learning rules, such as error-based learning could even predict a disadvantage for the new task (i.e., reverse adaptation)[22], as its connections to the control region could get weaker according to such learning rules (see Box 2 for a parallel explanation of similarity-induced interference).

The learning account, similar to the traditional account, predicts within-task adaptation. This means that increased control demand in the current trial increases control implementation in the subsequent trials of the same task[116]. But the learning account makes two unique predictions: (a) since learning is, by definition, long-lasting, within-task adaptation should be resilient against the passage of time and intervening trials. (b) Recruitment of control for one task should lead to null or reverse cross-task adaptation in a different task. Both findings have empirical support in the cognitive control literature in general[99,147,148] and in language production, in specific[22,149]. It is worth noting that small positive cross-task adaptations[130,131] are not incompatible with the learning account, especially if mechanisms such as Hebbian learning are considered: at the end of a successful trial, all active mappings are reinforced. Therefore, if both sets of stimulus-response mappings across the two tasks and their connections to the control center are simultaneously activated, there is a possibility to observe small positive effects. However, this effect is expected to be inconsistent and weak, as it is usually not possible to keep the mappings of one task fully active while performing another task, hence the ubiquity of switch costs[150].

To summarize, the learning account of cognitive control is both theoretically motivated and well-supported by two critical aspects of data on adaptation, its longevity and its largely domain and task-specific nature. This accounts maintains that, in terms of functional application, control is largely domain-specific. This position is incompatible with brain training programs, which aim to improve a cognitive function (e.g., language after stroke) through training working memory and cognitive control in the context of different tasks and domains[113,151].

Repair

A repair refers to the replacement of one response with another, e.g., "Please pass me the salt… I mean the pepper.". Compared to other aspects of language production, repair processes have received little attention, cf. [66,77,152]. The little empirical research that exists on the topic shows three key properties: (1) At least some repairs can be extremely fast, with as little as 0 ms between stopping the production of the error and initiating the repair[153,154]. (2) Repairs are seen in children as young as 2 years of age who are unable to consciously explain whether and why there was a repair[155,156]. Even in adults, repairs are not always accompanied by conscious awareness[157]. (3) Repairs do not seem to depend on accessing, or even recognizing, the correct target. For example, a speaker may repair one error with another error, showing that access to the correct target is not necessary for repairs. Even more interestingly, when the correct target is part of the repair, it may go unrecognized as the target. For example, in an attempt to name the picture of an "orange", an individual with aphasia produced a string of responses "apple, pineapple, pumpkin, orange, pineapple, peach?", not realizing that the correct target has been produced [66]. In short, repairs can be very fast, and largely independent of a conscious comparison with a correct target.

The mechanisms underlying linguistic repairs have remained largely obscured. Given the close link between monitoring and repair, it is reasonable to expect that repairs reflect some properties of monitoring. For example, it has been shown that error detection follows a bimodal distribution, most likely reflecting the workings of internal vs. external monitoring mechanisms. Correspondingly, repairs also show an early and a late distribution[77]. Moreover, as discussed in the multi-process monitoring view, certain mechanisms are better suited for detecting certain types of errors, causing differences in detection and repair of semantic and phonological/phonetic errors[52,59,79]. But my goal in this section is to focus on a more fundamental issue: where do repairs come from and how do they replace the error?

I will first propose a basic account that captures these properties. I will then augment this basic account with a conflict-based monitoring-control loop[158], to propose an adaptive model of repairs that captures more elaborate findings. The current proposal focuses on lexical repairs (replacing one word with another word), but its general principles can be readily extended to cover other types of linguistic repairs as well.

The basic repair model

An important debate about the nature of the repair process is whether, upon the detection of an error, the production process is started from scratch[159,160], or is continued in some fashion without a restart [66,152]. The latter position is desirable for a number of reasons; it is more easily reconciled with the very fast timeline of some repairs, it is cognitively less demanding, and most importantly, it is compatible with the natural consequences of language production, i.e., the co-activation of target and competitors. As such, a basic repair model can easily harness the basic dynamics of a model of language production. Here, I propose a basic model of repair as a time-based variant[161] of the Foygel and Dell's (2000) model[18]. In this time-based variant, activation dynamics are closer to evidence accumulation models (Box 3), i.e., they start from zero and grow over time with spreading activation, allowing for a more accurate representation of changes over time. Two ideas are behind the basic model. The first is the notion that conflict is generally higher in error vs. correct trials, discussed earlier in the Monitoring section. The second is the notion of continued processing past the point of response[61,162]. The basic repair model works by a simple *respond-&-check* mechanism. At time t, a response is made by selecting the most active representation, but processing continues a little longer, up to time $t+1$, when the system makes a second selection by the same rule. If the selected item matches the original selection, nothing is changed. If it does not, the second selection replaces the first one (an additional utility of post-processing is determining confidence in response[162]).

Figure 4a shows this process. On correct trials (the majority of trials), conflict is consistently low throughout the trial (left panel), therefore, the selected item at the times of response and check is often the same (correct) response, leading to high confidence and no change. On error trials, conflict is, on average, higher. If a response is selected before the system has had time to resolve the conflict, the first selection may be an error (right panel). However, time often resolves conflict in favor of the correct response (see also Fig. 2b), making the second selection the correct response. Repairs implemented through the respond-&-check process can be fast, because the repair has been processed all along[152]. They are also largely subconscious, because the check mechanism is the same implicit activation-based mechanism used for selection. Finally, they are independent of a standard correct response. For example, if the competition is between two or more representations, any one of them can get selected during the check process, regardless of its correctness, allowing the model to explain false alarms of the kind observed in individuals with aphasia, including the example given earlier in this section (see also [52]).

The adaptive repair model

The basic model thus accounts for the three main properties of repairs; it is fast, it does not require conscious processing or a correct target. However, it has no mechanism to adapt to difficult situations. Current findings suggest that when the probability of an error increases, so does the probability of a repair[50,66]. This finding has the same flavor as within-task adaptation, discussed in the Control section. I will thus tackle this problem by situating the basic repair model in a monitoring-control loop triggered by conflict (see also Gauvin & Hartsuiker, 2020, for a similar proposal)[158,161]. Figure 4b shows this adaptive model. The input activates the semantic features, and subsequently, lexical representations, where the respond-&-check mechanism is implemented. Conflict is continuously monitored at this level, providing the system with an average of conflict over all trials, weighted more heavily towards the most recent trial(s). If the average conflict around the time of the first selection (response) surpasses a certain threshold, control is recruited to boost the correct semantic-lexical mapping for the second selection (check). The higher the conflict level, the larger the boost. This results in a higher proportion of repairs in more error-prone situations, compatible with the empirical findings[161].

To recap, the adaptive repair model harnesses the basic dynamics of psycholinguistic models, combines it with conflict-based monitoring and control, and uses insights from the decision-making literature on threshold setting and post-response processing. This allows the model to explain the three basic properties of repairs mentioned earlier. Moreover, the model implements the notion of learning by computing a weighted average of conflict over many trials to be used as the regulating signal. This allows the model to explain the increase in repair rates in the face or higher error rates. Moreover, the learning mechanism can accommodate both sustained control, as well as trial by trial, fluctuation in control needs[163]. I must note that this proposal is quite new, and thus open for testing. Future studies and empirical data will determine how well this model will fare and what needs to be changed. Also, I hope that this proposal inspires alternative proposals that may fare even better than the current model.

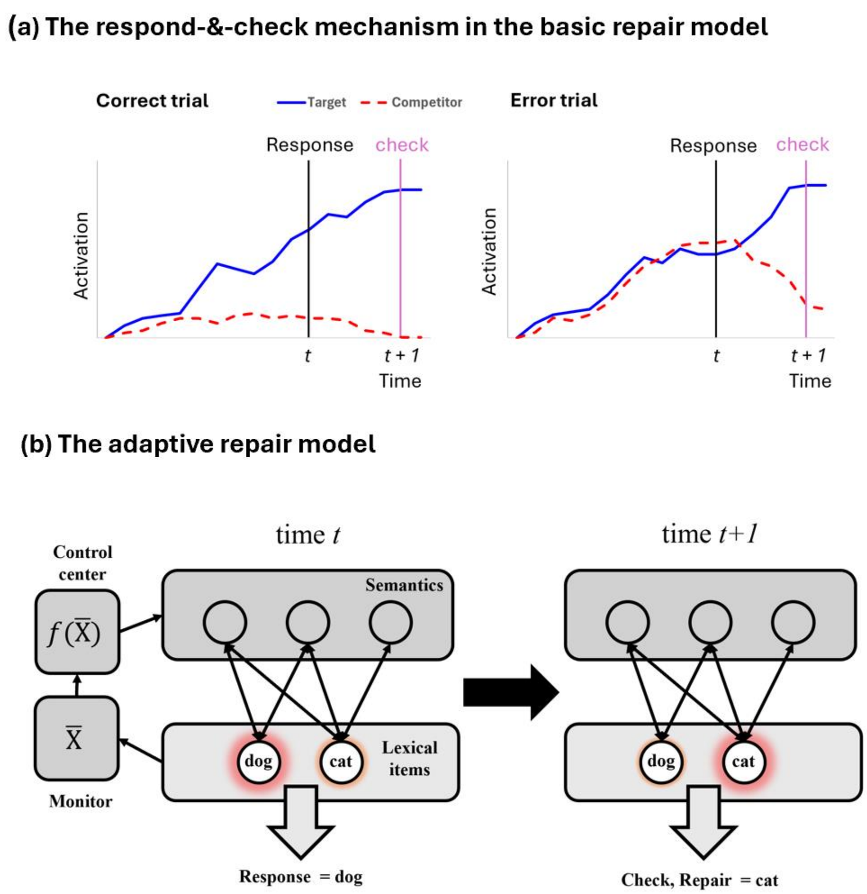

Figure 4. The repair model changing the error "dog" to the correct response "cat". (a) The respond-&-check mechanism in the basic model for a correct trial (left) and an error trial (right). Reselection of the response in the former leads to high confidence and no additional action. In the latter, the changed outcome of selection triggers the replacement of the prior response with a repair. (b) The adaptive repair model. The basic repair process of respond-&-check is situated within a monitoring and control loop triggered by a weighted average of conflict. Red indicates greater activation than orange.

Is there an "ideal" model?

Throughout the paper, I have discussed the differences between psycholinguistic and motor speech models, and how these differences bear on how language production is monitored and regulated. In this final section, I will tackle the question of "an ideal" model. Is an ideal model a "mega-model" that combines the models from the two traditions? As Geroge Box famously said, "All models are wrong, but some are useful.". Therefore, the question of an ideal model may be better phrased as the question of a useful model. Many of the existing models are useful for understanding a target phenomenon. For example, if the purpose is to test whether non-competitive selection is compatible with behavioral interference, a simple model with basic representations and an error-based learning mechanism does the job[9]. However, if the purpose is to explain both taxonomically-induced interference and thematically-induced facilitation within the same system, then the model needs to have a more sophisticated representation of the semantic space[8]. More complex models explain more data but are also harder to implement and understand. Therefore, a "mega-model" is only motivated if one really needs to model the whole production chain from beginning to end.

On the one hand, behavioral and neural data, as well as computational implementations, have pointed to a globally modular, locally interactive system[60,164]. This means that cascading and feedback affect the adjacent layers much more than distant layers. If so, we may not need a mega-model. For example, perhaps going up to the lexical level could be perfectly sufficient for modeling articulatory processes. On the other hand, some aspects of production will be left out by this approach. For instance, focused elements in linguistic messages tend to be acoustically prominent[165]. This finding links semantic representations to articulatory ones at the two ends of the production system. A model with a narrower scope will simply miss some of the relevant representations or must speculate on how processing takes place among those representations. More generally, if we are to understand the regulation of the language production system through hierarchical model-based RL theory (Box 4), we must train models that map the entire process. This approach is critical for assessing whether such a framework applies to a highly generative system like language production. Moreover, it allows us to better understand the kinds of information used in earlier stages to assess a final outcome[166].

In short, I believe some principles like cascading of activation, interactivity, co-activation of similar representations, speaker's goals especially in terms of emphasis on speed vs. accuracy, and implicit learning must always be considered when modeling a certain linguistic phenomenon. On the other hand, the scope of model's representations depends on the purpose of modeling. For certain claims to be verifiable, a mega-model is indeed necessary.

Conclusion

My goal in this paper was to combine insights from different traditions of language production research, as well as advances in other fields of cognition, to propose new directions for research on word production. I will summarize the conclusions in four points, together with their potential contribution to future directions. The first two points are focused on specific mechanisms. The last two points are broader takeaways.

1. A multi-process monitoring view. Motor speech and psycholinguistic traditions differ in some of the mechanisms they propose for monitoring and control of language production. These differences do not necessarily hint at a problem with one approach or the other. Rather, they likely reflect important differences in the nature of representations that are the focus of each tradition. Therefore, instead of choosing the "right" monitoring mechanism, a more fruitful approach could be to view monitoring as a multi-mechanism process. Embracing the multi-process monitoring view shifts the focus of future research to important open questions, such as the relative contribution of various mechanisms to monitoring different aspects of language production, the kind of information used for monitoring at different stages of processing, and better ways of leveraging the complementary role of these mechanisms for rehabilitation after brain damage.

2. A repair mechanism embedded in the monitoring-control loop. The adaptive repair model, proposed here, takes the idea of continued processing past the selection point and embeds it within the same conflict-driven monitoring-control loop that controls the primary production process. This account is parsimonious, because control is always applied the same way, but depending on how quickly it takes effect it may either prevent an error or facilitate a repair. This model, however, is in its nascency. Many issues remain open to investigation such as the timeline of conflict monitoring during a trial, as well as the correct functions for weighting the recency of conflict and scaling of control.

3. Learning and production as one. For reasons stated in this paper, I believe that learning plays a critical role in correctly understanding a range of behavioral findings from rapid adjustment to new constraints to facilitatory and inhibitory effects of contextual similarity, to control and repair processes. Fortunately, psycholinguistic models lend themselves well to the integration of learning mechanisms such as error-based and Hebbian learning. Future work should consider the power of learning accounts in explaining new phenomena, even when such phenomena seem to have little to do with learning in its pedagogical sense.

The implementation of learning mechanisms should not be restricted to the inner workings of the language production system, but should also extend to its interactions with other systems. A broader picture is emerging in the cognitive control literature pointing to the tight link between action and learning. An example is the learning view of cognitive control[94,145,146]. This view predicts that domain-generality in the neural underpinnings of control actually leads to domain-specificity in the application of control, and is supported by evidence from both language production and non-linguistic tasks. Embracing this view and its application to language production is an excellent step towards answering more nuanced questions, such as how the parameters of such learning are set in different populations and under different circumstances.

4. Language production studied within a decision-making framework. Much of the discussion in the current paper draws on principles and mechanisms from fields of cognition other than language. I have shown how applying SDT to distributions of conflict for correct and error responses can introduce useful concepts such as criterion setting into debates of selection and monitoring. This framework can also be applied to information used by other accounts. Other decision making models have also been used and have advantages over SDT[167,168]. But there are also challenges in applying such models to the complex, non-linear process of word production[169] (Box 3). Resolving these challenges is only possible with a closer collaboration between language and decision scientists, but it is an effort with a high payoff for both fields.

Acknowledgments

I would like to thank Gary Dell and Jennifer Trueblood for helpful discussions and comments on an earlier version of this manuscript. This work was supported by NSF award BCS-2317121 to N.N.

Competing interests

The author declares no competing interests.

References

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

This is a staple of psycholinguistic models.

2. Guenther, F. H. A neural network model of speech acquisition and motor equivalent speech production. *Biol. Cybern.* **72**, 43–53 (1994).

3. Levelt, W. J. M., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–38 (1999).

This article lays out the entire architecture of the language production system from a psycholinguistic perspective.

4. Guenther, F. H. *Neural Control of Speech*. (MIT Press, 2016).

This book offers an excellent overview of computational and neural aspects of the motor control of speech production.

5. Nozari, N. How Special Is Language Production? Perspectives From Monitoring and Control. in *Psychology of Learning and Motivation* vol. 68 179–213 (Elsevier, 2018).

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

7. Nozari, N. & Hepner, C. R. To select or to wait? Response to the commentaries. *Cognitive Neuropsychology* **36**, 226–233 (2019).

8. Oppenheim, G. M. & Nozari, N. Similarity-induced interference or facilitation in language production reflects representation, not selection. *Cognition* **245**, 105720 (2024).

9. Oppenheim, G. M., Dell, G. S. & Schwartz, M. F. The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition* 114, 227–252 (2010).

This model is the pioneering model of error-based learning as an alternative to competitive selection.

10. Chang, F., Dell, G. S. & Bock, K. Becoming syntactic. *Psychological Review* **113**, 234–272 (2006).

11. Warker, J. A. & Dell, G. S. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**, 387–398 (2006).

12. Dell, G. S., Nozari, N. & Oppenheim, G. M. Word production: Behavioral and computational considerations. *The Oxford handbook of language production* 88–104 (2014).

13. Kearney, E. & Guenther, F. H. Articulating: the neural mechanisms of speech production. *Language, Cognition and Neuroscience* **34**, 1214–1229 (2019).

14. Roelofs, A. WEAVER++ and other computational models of lemma retrieval and word-form encoding. in *Aspects of language production* 71–114 (Psychology Press, 2013).

15. Rumelhart, D. E., McClelland, J. L. & Group, P. R. *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. (The MIT press, 1986).

16. Hsiao, E. Y., Schwartz, M. F., Schnur, T. T. & Dell, G. S. Temporal characteristics of semantic perseverations induced by blocked-cyclic picture naming. *Brain and language* **108**, 133–144 (2009).

17. Hepner, C. R. & Nozari, N. The dual origin of lexical perseverations in aphasia: Residual activation and incremental learning. *Neuropsychologia* **147**, 107603 (2020).

18. Foygel, D. & Dell, G. S. Models of Impaired Lexical Access in Speech Production. *Journal of Memory and Language* **43**, 182–216 (2000).

19. Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* **104**, 801–838 (1997).

20. Nozari, N., Kittredge, A. K., Dell, G. S. & Schwartz, M. F. Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language* **63**, 541–559 (2010).

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

22. Freund, M. & Nozari, N. Is adaptive control in language production mediated by learning? *Cognition* 176, 107–130 (2018).

This paper presents the first set of empirical evidence in support of the learning theory of cognitive control in language production.

23. Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A. & Caramazza, A. 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**, 503–535 (2007).

24. Roelofs, A. A spreading-activation theory of lemma retrieval in speaking. *Cognition* **42**, 107–142 (1992).

25. Shao, Z., Roelofs, A., Martin, R. C. & Meyer, A. S. Selective inhibition and naming performance in semantic blocking, picture-word interference, and color–word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**, 1806–1820 (2015).

26. Nozari, N. & Dell, G. S. How damaged brains repeat words: A computational approach. *Brain and Language* **126**, 327–337 (2013).

27. Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S. & Sobel, P. A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language* **54**, 228–264 (2006).

28. Budd, M.-J., Hanley, J. R. & Griffiths, Y. Simulating children's retrieval errors in picture-naming: A test of Foygel and Dell's (2000) semantic/phonological model of speech production. *Journal of Memory and Language* **64**, 74–87 (2011).

29. Budd, M.-J., Hanley, J. R. & Nozari, N. Evidence for a Non-Lexical Influence on Children's Auditory Repetition of Familiar Words. *J Psycholinguist Res* **41**, 253–266 (2012).

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

31. Indefrey, P. & Levelt, W. J. M. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).

32. Nozari, N. The neural basis of word production. in *The Oxford Handbook of the Mental Lexicon* (eds. Papafragou, A., Trueswell, J. C. & Gleitman, L. R.) 0 (Oxford University Press, 2022). doi:10.1093/oxfordhb/9780198845003.013.12.

33. Walker, G. M. & Hickok, G. Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM). *Psychon Bull Rev* **23**, 339–352 (2016).

34. Pinet, S. & Nozari, N. "Twisting fingers": The case for interactivity in typed language production. *Psychon Bull Rev* **25**, 1449–1457 (2018).

35. Rapp, B. & Fischer-Baum, S. Uncovering the Cognitive Architecture of Spelling. in *The Handbook of Adult Language Disorders* (Psychology Press, 2015).

36. Bohland, J. W., Bullock, D. & Guenther, F. H. Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience* **22**, 1504–1529 (2010).

37. Tourville, J. A. & Guenther, F. H. The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes* **26**, 952–981 (2011).

38. Bhushan, N. & Shadmehr, R. Computational nature of human adaptive control during learning of reaching movements in force fields. *Biol Cybern* **81**, 39–60 (1999).

39. Mazzoni, P. & Krakauer, J. W. An Implicit Plan Overrides an Explicit Strategy during Visuomotor Adaptation. *J. Neurosci.* **26**, 3642–3645 (2006).

40. McNamee, D. & Wolpert, D. M. Internal Models in Biological Control. *Annual Review of Control, Robotics, and Autonomous Systems* **2**, 339–364 (2019).

41. Chang, S.-E. & Guenther, F. H. Involvement of the Cortico-Basal Ganglia-Thalamocortical Loop in Developmental Stuttering. *Front. Psychol.* **10**, (2020).

42. Civier, O., Tasko, S. M. & Guenther, F. H. Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders* **35**, 246–279 (2010).

43. Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S. & Wallace, M. E. Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Dysfluency: A Theoretical Model of Stuttering. *Contemporary Issues in Communication Science and Disorders* **31**, 105–122 (2004).

44. Miller, H. E. & Guenther, F. H. Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework. *Aphasiology* **35**, 424–441 (2021).

45. Golfinopoulos, E., Tourville, J. A. & Guenther, F. H. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage* **52**, 862–874 (2010).

46. Guenther, F. H. & Vladusich, T. A neural theory of speech acquisition and production. *Journal of Neurolinguistics* **25**, 408–422 (2012).

47. Hickok, G. Computational neuroanatomy of speech production. *Nature Reviews Neuroscience* **13**, 135–145 (2012).

48. Oppenheim, G. M. & Dell, G. S. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* **106**, 528–537 (2008).

49. Levelt, W. J. M. Monitoring and self-repair in speech. *Cognition* 14, 41–104 (1983).

This paper proposes the first psycholinguistic model of speech monitoring, the perceptual loop theory.

50. Levelt, W. J. M. *Speaking: From Intention to Articulation*. (MIT Press, 1989).

51. Vigliocco, G. & Hartsuiker, R. J. The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin* **128**, 442–472 (2002).

52. Nozari, N., Dell, G. S. & Schwartz, M. F. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology* 63, 1–33 (2011).

This paper proposes the conflict model of speech monitoring.

53. Nozari, N. A comprehension-or a production-based monitor? Response to Roelofs (2020). *Journal of Cognition* **3**, (2020).

54. Roelofs, A. On (correctly representing) comprehension-based monitoring in speaking: Rejoinder to Nozari (2020). *Journal of Cognition* **3**, (2020).

55. Warker, J. A., Xu, Y., Dell, G. S. & Fisher, C. Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition* **112**, 81–96 (2009).

56. Patel, T., Morales, M., Pickering, M. J. & Hoffman, P. A common neural code for meaning in discourse production and comprehension. *NeuroImage* **279**, 120295 (2023).

57. Segaert, K., Menenti, L., Weber, K., Petersson, K. M. & Hagoort, P. Shared syntax in language production and language comprehension—an fMRI study. *Cerebral Cortex* **22**, 1662–1670 (2012).

58. Marshall, R. C., Rappaport, B. Z. & Garcia-Bunuel, L. Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language* **24**, 297–313 (1985).

59. McCall, J. D. *et al.* Listening to yourself and watching your tongue: Distinct abilities and brain regions for monitoring semantic and phonological speech errors. *Journal of Cognitive Neuroscience* **35**, 1169–1194 (2023).

60. Pinet, S. & Nozari, N. Different Electrophysiological Signatures of Similarity-induced and Stroop-like Interference in Language Production. *Journal of Cognitive Neuroscience* **35**, 1329–1349 (2023).

61. Yeung, N., Botvinick, M. M. & Cohen, J. D. The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review* **111**, 931–959 (2004).

62. Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychological Review* 108, 624–652 (2001).

This paper proposes the original conflict theory of action monitoring.

63. McNicol, D. *A Primer of Signal Detection Theory*. (Psychology Press, New York, 2004). doi:10.4324/9781410611949.

64. Maniscalco, B. & Lau, H. Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-d′, Response-Specific Meta-d′, and the Unequal Variance SDT Model. in *The Cognitive Neuroscience of Metacognition* (eds. Fleming, S. M. & Frith, C. D.) 25–66 (Springer, Berlin, Heidelberg, 2014). doi:10.1007/978-3-642-45190-4_3.

65. Hanley, J. R., Cortis, C., Budd, M.-J. & Nozari, N. Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology* **142**, 36–47 (2016).

66. Nozari, N., Martin, C. D. & McCloskey, N. Is repairing speech errors an automatic or a controlled process? Insights from the relationship between error and repair probabilities in English and Spanish. *Language, Cognition and Neuroscience* **0**, 1–16 (2019).

67. Nozari, N. The dual origin of semantic errors in access deficit: activation vs. inhibition deficit. *Cognitive Neuropsychology* **0**, 1–23 (2019).

68. Postma, A. & Kolk, H. The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech, Language, and Hearing Research* **36**, 472–487 (1993).

69. Levelt, W. J. M. Accessing words in speech production: Stages, processes and representations. *Cognition* **42**, 1–22 (1992).

70. Katseff, S., Houde, J. & Johnson, K. Partial Compensation for Altered Auditory Feedback: A Tradeoff with Somatosensory Feedback? *Lang Speech* **55**, 295–308 (2012).

71. Niziolek, C. A. & Guenther, F. H. Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *J. Neurosci.* **33**, 12090–12098 (2013).

72. Runnqvist, E. *et al.* Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production. *Cortex* **81**, 203–214 (2016).

73. Runnqvist, E. *et al.* Cerebellar and cortical correlates of internal and external speech error monitoring. *Cerebral Cortex Communications* **2**, tgab038 (2021).

74. Shiller, D. M., Rvachew, S. & Brosseau-Lapré, F. Importance of the auditory perceptual target to the achievement of speech production accuracy. *Canadian Journal of Speech-Language Pathology & Audiology* **34**, (2010).

75. Sutton, R. S. & Barto, A. G. *Reinforcement Learning, Second Edition: An Introduction*. (MIT Press, 1988).

This is the reference for a well-laid-out reinforcement learning account.

76. Lind, A. & Hartsuiker, R. J. Self-monitoring in speech production: comprehending the conflict between conflict-and comprehension-based accounts. *Journal of Cognition* **3**, (2020).

77. Nooteboom, S. G. & Quené, H. Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language* **95**, 19–35 (2017).

78. Pinet, S. & Nozari, N. Electrophysiological correlates of monitoring in typing with and without visual feedback. *Journal of cognitive neuroscience* **32**, 603–620 (2020).

79. Hartsuiker, R. J. & Kolk, H. H. J. Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory. *Cognitive Psychology* **42**, 113–157 (2001).

80. Takakusaki, K. Functional Neuroanatomy for Posture and Gait Control. *J Mov Disord* **10**, 1–17 (2017).

81. Brown, J. W. Beyond Conflict Monitoring: Cognitive Control and the Neural Basis of Thinking Before You Act. *Curr Dir Psychol Sci* **22**, 179–185 (2013).

82. Miyake, A. *et al.* The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology* **41**, 49–100 (2000).

83. Trude, A. & Nozari, N. Inhibitory control supports referential context use in language production and comprehension. in *Proceedings of the 39th Annual Conference of the Cognitive Science Society* 1218–1224 (Cognitive Science Society, 2017).

84. Nozari, N. & Omaki, A. Syntactic production is not independent of inhibitory control: Evidence from agreement attraction errors. in *Proceedings of the 40th Annual Conference of the Cognitive Science Society Austin, TX: Cognitive Science Society* (2018).

85. Shao, Z., Roelofs, A. & Meyer, A. S. Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The Quarterly Journal of Experimental Psychology* **65**, 1927–1944 (2012).

86. Shao, Z., Meyer, A. S. & Roelofs, A. Selective and nonselective inhibition of competitors in picture naming. *Mem Cogn* **41**, 1200–1211 (2013).

87. Crowther, J. E. & Martin, R. C. Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. *Front. Hum. Neurosci.* **8**, (2014).

88. Ness, T., Langlois, V. J., Kim, A. E. & Novick, J. M. The State of Cognitive Control in Language Processing. *Perspect Psychol Sci* 17456916231197122 (2023) doi:10.1177/17456916231197122.

89. Novick, J. M., Trueswell, J. C. & Thompson-Schill, S. L. Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience* **5**, 263–281 (2005).

90. Nozari, N. & Thompson-Schill, S. L. Chapter 46 - Left Ventrolateral Prefrontal Cortex in Processing of Words and Sentences. in *Neurobiology of Language* (eds. Hickok, G. & Small, S. L.) 569–584 (Academic Press, San Diego, 2016). doi:10.1016/B978-0-12-407794-2.00046-8.

91. Nozari, N., Trueswell, J. C. & Thompson-Schill, S. L. The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychon Bull Rev* **23**, 1942–1953 (2016).

92. Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K. & Farah, M. J. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *PNAS* **94**, 14792–14797 (1997).

93. Straub, E. R., Schiltenwolf, M., Kiesel, A. & Dignath, D. Generalizability of control across cognitive and emotional conflict. *Journal of Experimental Psychology: Human Perception and Performance* **50**, 2–22 (2024).

94. Braem, S., Chai, M., Held, L. K. & Xu, S. One cannot simply 'be flexible': regulating control parameters requires learning. *Current Opinion in Behavioral Sciences* **55**, 101347 (2024).

95. Belke, E., Meyer, A. S. & Damian, M. F. Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A* **58**, 667–692 (2005).

96. Breining, B., Nozari, N. & Rapp, B. Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychon Bull Rev* **23**, 500–506 (2016).

97. Nozari, N., Freund, M., Breining, B., Rapp, B. & Gordon, B. Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience* **31**, 886–903 (2016).

98. Schnur, T. T. *et al.* Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *PNAS* **106**, 322–327 (2009).

99. Verguts, T. & Notebaert, W. Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review* 115, 518–525 (2008).

This paper proposes an implementation of cognitive control as Hebbian learning.

100. Botvinick, M. M., Cohen, J. D. & Carter, C. S. Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences* **8**, 539–546 (2004).

101. Gauvin, H. S., De Baene, W., Brass, M. & Hartsuiker, R. J. Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage* **126**, 96–105 (2016).

102. Pardo, J. V., Pardo, P. J., Janer, K. W. & Raichle, M. E. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences* **87**, 256–259 (1990).

103. Piai, V., Roelofs, A., Acheson, D. J. & Takashima, A. Attention for speaking: domain-general control from the anterior cingulate cortex in spoken word production. *Front. Hum. Neurosci.* **7**, (2013).

104. Riès, S., Janssen, N., Dufau, S., Alario, F.-X. & Burle, B. General-Purpose Monitoring during Speech Production. *Journal of Cognitive Neuroscience* **23**, 1419–1436 (2011).

105. Piai, V., Riès, S. K. & Swick, D. Lesions to Lateral Prefrontal Cortex Impair Lexical Interference Control in Word Production. *Front. Hum. Neurosci.* **9**, (2016).

106. de Zubicaray, G. I., Hansen, S. & McMahon, K. L. Differential processing of thematic and categorical conceptual relations in spoken word production. *Journal of Experimental Psychology: General* **142**, 131–142 (2013).

107. Poeppel, D. & Hickok, G. Towards a new functional anatomy of language. *Cognition* **92**, 1–12 (2004).

108. Riès, S. *et al.* Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *PNAS* **114**, E4530–E4538 (2017).

109. Martin, R. C. & Cheng, Y. Selection demands versus association strength in the verb generation task. *Psychonomic Bulletin & Review* **13**, 396–401 (2006).

110. Nozari, N., Mirman, D. & Thompson-Schill, S. L. The ventrolateral prefrontal cortex facilitates processing of sentential context to locate referents. *Brain and Language* **157–158**, 1–13 (2016).

111. Wagner, A. D., Paré-Blagoev, E. J., Clark, J. & Poldrack, R. A. Recovering Meaning: Left Prefrontal Cortex Guides Controlled Semantic Retrieval. *Neuron* **31**, 329–338 (2001).

112. de Zubicaray, G. I. & McMahon, K. L. Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience* **9**, 260–269 (2009).

113. Nozari, N. & Martin, R. Is working memory domain-general or domain-specific? Implications for theory and practice. (2024).

114. Nozari, N. & Novick, J. Monitoring and Control in Language Production. *Curr Dir Psychol Sci* 26, 403–410 (2017).

This brief review discusses a taxonomy for the domain-generality of cognitive control and applies it to language processing.

115. Hepner, C. R. & Nozari, N. Resource allocation in phonological working memory: Same or different principles from vision? *Journal of Memory and Language* **106**, 172–188 (2019).

116. Gratton, G., Coles, M. G. H. & Donchin, E. Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General* **121**, 480–506 (1992).

117. Freund, M., Gordon, B. & Nozari, N. Conflict-based regulation of control in language production. in *Proceedings of the 38th annual conference of the cognitive science society* 1625–1630 (2016).

118. Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N. & Notebaert, W. The congruency sequence effect 3.0: a critical test of conflict adaptation. *PloS one* **9**, e110462 (2014).

119. January, D., Trueswell, J. C. & Thompson-Schill, S. L. Co-localization of Stroop and Syntactic Ambiguity Resolution in Broca's Area: Implications for the Neural Basis of Sentence Processing. *Journal of Cognitive Neuroscience* **21**, 2434–2444 (2009).

120. Kan, I. P. & Thompson-Schill, S. L. Selection from perceptual and conceptual representations. *Cognitive, Affective, & Behavioral Neuroscience* **4**, 466–482 (2004).

121. Nozari, N., Arnold, J. E. & Thompson-Schill, S. L. The Effects of Anodal Stimulation of the Left Prefrontal Cortex on Sentence Production. *Brain Stimulation* **7**, 784–792 (2014).

122. Nozari, N. & Thompson-Schill, S. L. More attention when speaking: Does it help or does it hurt? *Neuropsychologia* **51**, 2770–2780 (2013).

123. Thompson-Schill, S. L. *et al.* Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *PNAS* **95**, 15855–15860 (1998).

124. Wessel, J. R. & Anderson, M. C. Neural mechanisms of domain-general inhibitory control. *Trends in Cognitive Sciences* **28**, 124–143 (2024).

125. Kan, I. P. *et al.* To adapt or not to adapt: The question of domain-general cognitive control. *Cognition* **129**, 637–651 (2013).

126. Tang, H. *et al.* Prefrontal cortical plasticity during learning of cognitive tasks. *Nat Commun* **13**, 90 (2022).

127. Hussey, E. K. *et al.* Memory and language improvements following cognitive control training. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **43**, 23–58 (2017).

128. Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I. & Bunting, M. F. Clearing the garden-path: improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience* **29**, 186–217 (2014).

129. Hsu, N. S. & Novick, J. M. Dynamic Engagement of Cognitive Control Modulates Recovery From Misinterpretation During Real-Time Language Processing. *Psychol Sci* **27**, 572–582 (2016).

130. Ovans, Z. *et al.* Cognitive control states influence real-time sentence processing as reflected in the P600 ERP. *Language, Cognition and Neuroscience* **37**, 939–947 (2022).

131. Thothathiri, M., Asaro, C. T., Hsu, N. S. & Novick, J. M. *Who* did *what*? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition* **178**, 162–177 (2018).

132. Akçay, Ç. & Hazeltine, E. Domain-specific conflict adaptation without feature repetitions. *Psychon Bull Rev* **18**, 505–511 (2011).

133. Egner, T., Delano, M. & Hirsch, J. Separate conflict-specific cognitive control mechanisms in the human brain. *NeuroImage* **35**, 940–948 (2007).

134. Kim, C., Chung, C. & Kim, J. Conflict adjustment through domain-specific multiple cognitive control mechanisms. *Brain Research* **1444**, 55–64 (2012).

135. Kunde, W., Augst, S. & Kleinsorge, T. Adaptation to (non)valent task disturbance. *Cogn Affect Behav Neurosci* **12**, 644–660 (2012).

136. Kunde, W. & Stöcker, C. A Simon effect for stimulus-response duration. *The Quarterly Journal of Experimental Psychology Section A* **55**, 581–592 (2002).

137. Kunde, W. & Wühr, P. Sequential modulations of correspondence effects across spatial dimensions and tasks. *Memory & Cognition* **34**, 356–367 (2006).

138. Wendt, M., Kluwe, R. H. & Peters, A. Sequential modulations of interference evoked by processing task-irrelevant stimulus features. *Journal of Experimental Psychology: Human Perception and Performance* **32**, 644–667 (2006).

139. Atilgan, N., Chu, E. & Nozari, N. Statistical learning of orthotactic constraints: new insights from typing. *Proceedings of the Annual Meeting of the Cognitive Science Society* **44**, (2022).

140. Smalle, E. H. M., Daikoku, T., Szmalec, A., Duyck, W. & Möttönen, R. Unlocking adults' implicit statistical learning by cognitive depletion. *Proceedings of the National Academy of Sciences* **119**, e2026011119 (2022).

141. Warker, J. A., Dell, G. S., Whalen, C. A. & Gereg, S. Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **34**, 1289 (2008).

142. Warker, J. A. & Dell, G. S. New phonotactic constraints learned implicitly by producing syllable strings generalize to the production of new syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**, 1902 (2015).

143. Dell, G. S., Kelley, A. C., Hwang, S. & Bian, Y. The adaptable speaker: A theory of implicit learning in language production. *Psychological review* **128**, 446 (2021).

144. Oppenheim, G. M. The paca that roared: Immediate cumulative semantic interference among newly acquired words. *Cognition* **177**, 21–29 (2018).

145. Braem, S. & Egner, T. Getting a grip on cognitive flexibility. *Current Directions in Psychological Science* **27**, 470–476 (2018).

146. Egner, T. Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Front. Psychol.* **5**, (2014).

147. Schiltenwolf, M., Kiesel, A. & Dignath, D. No temporal decay of cognitive control in the congruency sequence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **49**, 1247–1263 (2023).

148. Verguts, T. & Notebaert, W. Adaptation by binding: a learning account of cognitive control. *Trends in Cognitive Sciences* **13**, 252–257 (2009).

149. Zhu, D. *et al.* Cognitive control is task specific: Further evidence against the idea of domain-general conflict adaptation. Preprint at https://doi.org/10.31234/osf.io/kgrvd (2024).

150. Rogers, R. D. & Monsell, S. Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General* **124**, 207–231 (1995).

151. Ganesan, K. *et al.* Cognitive control training with domain-general response inhibition does not change children's brains or behavior. *Nat Neurosci* 1–12 (2024) doi:10.1038/s41593-024-01672-w.

152. Nooteboom, S. G. & Quené, H. Repairing speech errors: Competition as a source of repairs. *Journal of Memory and Language* **111**, 104069 (2020).

153. Blackmer, E. R. & Mitton, J. L. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition* **39**, 173–194 (1991).

154. Nooteboom, S. & Quené, H. Temporal aspects of self-monitoring for speech errors. *Journal of Memory and Language* **105**, 43–59 (2019).

155. Clark, E. V. Awareness of Language: Some Evidence from what Children Say and Do. in *The Child's Conception of Language* (eds. Sinclair, A., Jarvella, R. J. & Levelt, W. J. M.) 17–43 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1978). doi:10.1007/978-3-642-67155-5_2.

156. Karmiloff-Smith, A. From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition* **23**, 95–147 (1986).

157. Pinet, S. & Nozari, N. Correction without consciousness in complex tasks: evidence from typing. (2021).

158. Gauvin, H. S. & Hartsuiker, R. J. Towards a new model of verbal monitoring. *Journal of Cognition* **3**, (2020).

159. Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H. & Pickering, M. J. Concurrent processing of words and their replacements during speech. *Cognition* **108**, 601–607 (2008).

160. Seyfeddinipur, M., Kita, S. & Indefrey, P. How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition* **108**, 837–842 (2008).

161. Burgess, J. L. & Nozari, N. A conflict-based model of speech error repairs in humans. *Proceedings of the Annual Meeting of the Cognitive Science Society* **44**, (2022).

162. Pleskac, T. J. & Busemeyer, J. R. Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review* **117**, 864–901 (2010).

163. Braver, T. S., Reynolds, J. R. & Donaldson, D. I. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron* **39**, 713–726 (2003).

164. Rapp, B. & Goldrick, M. Discreteness and interactivity in spoken word production. *Psychological Review* 107, 460–499 (2000).

This paper provides evidence for the limits of interactivity within the language production system.

165. Arnold, J. E. & Watson, D. G. Synthesising meaning and processing approaches to prosody: performance matters. *Language, Cognition and Neuroscience* **30**, 88–102 (2015).

166. Botvinick, M. & Weinstein, A. Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369, 20130480 (2014).

This paper provides a good overview of hierarchical model-based reinforcement learning.

167. Anders, R., Riès, S., van Maanen, L. & Alario, F.-X. Evidence accumulation as a model for lexical selection. *Cognitive Psychology* **82**, 57–73 (2015).

168. Anders, R., Riès, S., Maanen, L. V. & Alario, F.-X. Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology* **34**, 1–20 (2017).

169. Anders, R. & Alario, F.-X. Challenges to developing time-based signal detection models for word production. *Cognitive Neuropsychology* **0**, 1–4 (2019).

170. Fodor, J. A. *The Modularity of Mind: An Essay on Faculty Psychology*. (A Bradford Book / MIT Press, Cambridge, Mass, 1983).

171. Peterson, R. R. & Savoy, P. Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **24**, 539–557 (1998).

172. Breining, B. L., Nozari, N. & Rapp, B. Learning in complex, multi-component cognitive systems: Different learning challenges within the same system. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **45**, 1093–1106 (2019).

173. Sadat, J., Martin, C. D., Costa, A. & Alario, F.-X. Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology* **68**, 33–58 (2014).

174. Goldrick, M. & Blumstein, S. E. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* (2007) doi:10.1080/01690960500181332.

175. Nozari, N. & Dell, G. S. More on lexical bias: How efficient can a "lexical editor" be? *Journal of Memory and Language* **60**, 291–307 (2009).

176. Hartsuiker, R. J., Corley, M. & Martensen, H. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language* **52**, 58–70 (2005).

177. Strijkers, K., Costa, A. & Pulvermüller, F. The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage* **163**, 206–219 (2017).

178. Indefrey, P. The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Front. Psychol.* **2**, (2011).

179. Nozari, N. & Pinet, S. A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics* **53**, 100875 (2020).

180. Bürki, A., Elbuy, S., Madec, S. & Vasishth, S. What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language* **114**, 104125 (2020).

181. Schriefers, H., Meyer, A. S. & Levelt, W. J. M. Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* **29**, 86–102 (1990).

182. Oppenheim, G. M. & Balatsou, E. Lexical competition on demand. *Cognitive Neuropsychology* **0**, 1–4 (2019).

183. Schnur, T. T., Schwartz, M. F., Brecher, A. & Hodgson, C. Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language* **54**, 199–227 (2006).

184. Abdel Rahman, R. & Melinger, A. When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **33**, 604–614 (2007).

185. Alario, F.-X., Segui, J. & Ferrand, L. Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A* **53**, 741–764 (2000).

186. Costa, A., Alario, F.-X. & Caramazza, A. On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review* **12**, 125–131 (2005).

187. Qu, Q., Feng, C. & Damian, M. F. Interference effects of phonological similarity in word production arise from competitive incremental learning. *Cognition* **212**, 104738 (2021).

188. Howard, D., Nickels, L., Coltheart, M. & Cole-Virtue, J. Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition* **100**, 464–482 (2006).

189. Coultrip, R., Granger, R. & Lynch, G. A cortical model of winner-take-all competition via lateral inhibition. *Neural Networks* **5**, 47–54 (1992).

190. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci* **5**, 793–807 (2004).

191. Singer, M. & Wixted, J. T. Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition* **34**, 125–137 (2006).

192. Donkin, C. & Brown, S. D. Response times and decision-making. *Stevens' handbook of experimental psychology and cognitive neuroscience* **5**, 349–377 (2018).

193. Heathcote, A. & Matzke, D. Winner Takes All! What Are Race Models, and Why and How Should Psychologists Use Them? *Curr Dir Psychol Sci* **31**, 383–394 (2022).

194. Brown, S. D. & Heathcote, A. The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology* **57**, 153–178 (2008).

195. Leite, F. P. & Ratcliff, R. Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics* **72**, 246–273 (2010).

196. Ratcliff, R. & McKoon, G. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation* **20**, 873–922 (2008).

197. Ratcliff, R. & Starns, J. J. Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review* **120**, 697–719 (2013).

198. van Ravenzwaaij, D., Brown, S. D., Marley, A. A. J. & Heathcote, A. Accumulating advantages: A new conceptualization of rapid multiple choice. *Psychological Review* **127**, 186–215 (2020).

199. Strickland, L., Loft, S., Remington, R. W. & Heathcote, A. Racing to remember: A theory of decision control in event-based prospective memory. *Psychological review* **125**, 851 (2018).

200. Usher, M. & McClelland, J. L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review* **108**, 550–592 (2001).

201. Teodorescu, A. R. & Usher, M. Disentangling decision models: From independence to competition. *Psychological Review* **120**, 1–38 (2013).

202. Heathcote, A. *et al.* Dynamic models of choice. *Behav Res* 51, 961–985 (2019).

This paper provides a good overview of decision-making models.

203. Holmes, W. R., Trueblood, J. S. & Heathcote, A. A new framework for modeling decisions about changing information: The Piecewise Linear Ballistic Accumulator model. *Cognitive psychology* **85**, 1–29 (2016).

204. Holroyd, C. B. & Coles, M. G. H. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review* **109**, 679–709 (2002).

205. Cockburn, J. & Frank, M. Reinforcement learning, conflict monitoring, and cognitive control: an integrative model of cingulate-striatal interactions and the ERN. (2011).

206. Schmidt, J. R. Evidence against conflict monitoring and adaptation: An updated review. *Psychon Bull Rev* **26**, 753–771 (2019).

207. Busemeyer, J. & Bruza, P. D. *Quantum Models of Cognition and Decision*. (Cambridge University Press, 2012).

208. Pickering, M. J. & Garrod, S. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* **36**, 329–347 (2013).

209. de Ruiter, J. P. & Cummins, C. Forward modelling requires intention recognition and non-impoverished predictions. *Behavioral and Brain Sciences* **36**, 351 (2013).

210. Zarr, N. & Brown, J. W. Hierarchical error representation in medial prefrontal cortex. *NeuroImage* **124**, 238–247 (2016).

211. Kostadinov, D. & Häusser, M. Reward signals in the cerebellum: Origins, targets, and functional implications. *Neuron* **110**, 1290–1303 (2022).