How special is language production?
Perspectives from monitoring and control

Nazbanou Nozari[a,b]

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

Correspondence should be addressed to:

Nazbanou Nozari, MD, PhD

Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, USA

Phone: 443-287-1712

Fax: 410-955-0188

E-mail: nozari@jhu.edu

Contents

Abstract

In his seminal essay, "The Modularity of Mind", Fodor (1983) presents arguments in favor of language comprehension as a special module along with other input processing systems. His view on language production is less clear. In this chapter, I first demonstrate that language production and comprehension are quite similar when evaluated in light of Fodor's criteria for modules: both meet a subset of those criteria in that their behavior resembles automatic processing; neither, however, is informationally encapsulated. This partial conformity with the criteria for specialized modules leaves the question "How special is language production?" unanswered. I then propose that this question can be answered by re-examining the origin of what resembles the behavior of an automatic system. I will argue that language production is, in fact, an efficiently monitored and controlled system, and that the monitoring and control mechanisms are shared between language production and other systems. These domain-general mechanisms, however, operate on domain-specific representations, creating specialized monitoring-control loops that can be selectively trained and selectively damaged.

1. The special place of language

As a uniquely human ability, language processing has always had a special place in cognition; so special in fact that it has been considered by many to be altogether separate from cognition. Most of the philosophical arguments, as well as empirical research, on the special nature of language processing has concerned language *comprehension*, though. My goal in this chapter is to evaluate how special language *production* is with regard to the rest of cognition. I will start with an overview of Fodor's seminal essay, "The Modularity of Mind" (Fodor, 1983), to introduce a theoretical framework within which it would be possible assess how "special" a cognitive system is, and also to apply the same framework that has been employed to evaluate the specialness of language comprehension to language production. In doing so, I will re-evaluate some of Fodor's assumptions and claims, and discuss a possible extension of his views to language production. I will then evaluate those views in light of the evidence from empirical studies of language production.

1.1. Functional taxonomy of cognitive processes: where is language?

Fodor recognizes two general families of cognitive systems: *modules* and *central processes*. This taxonomy is function-based: the function of modules is input analysis, whereas the function of central processes is the fixation of belief. In Fodor's view, what we believe depends on the input we receive plus information about how reliable that input is, but how we process an input is largely insensitive to what we believe. Where does language fall in this taxonomy? There are two options: (1) To group language with central processes of thought, and contrast that with perceptual processes, e.g., vision, audition, etc. (2) To group language with perceptual systems, consider them all input systems, and contrast them with central processes such as thought and problem solving.

Since the representational nature of language and its similarity to thought in that sense can hardly be denied, there would have to be a strong motivation for choosing the second classification. Fodor argues that such motivation exists. Language comprehension and other perceptual processes form a natural kind, he argues, because they have something fundamental in common: they are all modules.

1.2. What is a module?

According to Fodor, a system can be considered modular if it has all (or most) of the key characteristics that follow. Modular systems (1) are fast and mandatory and allow only limited central access to mental representations, i.e., they meet the criteria for automaticity; (2) are informationally encapsulated; and (3) are domain-specific.

1.2.1. Fast and mandatory processing, with limited central access to mental representations

These three properties were discussed separately by Fodor, but I have grouped them together as indicators of the "automaticity" of a system. Fodor argues that language comprehension is mandatory because we cannot help but understand the meaning of what is spoken in our native language, even when we do not intend to listen to a conversation or when we are instructed to focus on the acoustic properties of the input instead of its meaning (Marslen-Wilson & Tyler, 1981). On this ground, Fodor has compared language comprehension to a reflex, i.e., a system with no reliance on cognitive control operations. Our first interpretation of a sentence may not always be correct, but Fodor's critical point is that at least some meaning is always extracted online, quickly, and independent of listeners' intentions. This position is compatible with evidence showing that listeners quickly fixate not only the picture whose name they just heard (e.g., lock), but also semantically-related pictures (e.g., key), showing quick and involuntary access to the (extended) semantics of heard words (Yee & Sedivy, 2006).

By "limited central access to mental representations", Fodor means that conscious awareness of intermediate cognitive operations is limited. For example, upon hearing an utterance, a host of phonological, lexical, syntactic, and semantic processes are at work, but we are only aware of the final result: comprehension of the meaning of the sentence. We do not have access to the acoustic features of the words we heard, and we cannot accurately report how many lexical items or syntactic trees have been activated during the parsing of the sentence. This lack of awareness is a desirable property for a cognitive system, because, as Fodor explains, "the world is, in general, considerably more stable than are its projections onto the surfaces of transducers. Constancies correct for this, so that in general percepts correspond to distal layouts better than proximal stimuli do. But, of course, the work of the constancies would be undone unless the central systems which run behavior were required largely to ignore the representations which encode unconnected proximal information." (Fodor, 1983; p. 60).

In summary, language comprehension is fast, mandatory, and mostly without conscious awareness of the intermediate processes which map linguistic input to meaning. These three properties make language comprehension a reasonable candidate for an automatic process.

1.2.3. Informational encapsulation

When is a system informationally encapsulated? In a nutshell, when it does not have access to information from other systems. The tricky question here is what constitutes "other systems". For example, Fodor does not see the McGurk effect (perceiving the same auditory stimulus as different speech sounds depending on the accompanying visual information from mouth movements) as evidence of the influence of information from one system (visual) on another (auditory), because the phenomenon is contained within the "language domain". Similarly, he dismisses evidence of top-down processing within the same system as evidence of the penetrability of modules. Consider the phoneme restoration effect (Warren, 1970): listeners "perceive" phonemes that have been spliced out of familiar words and replaced with noises like coughs. For example, even though they never actually hear an "s" in /legi(cough)lature/, they report having heard "legislature" with a cough in the background. This finding has been taken as evidence of top-down feedback from lexical items to lower-level representations like phonemes. Fodor argues that even if this is true—a claim he disputes on the grounds that guessing is an equally valid explanation—it would not constitute evidence against modularity, because both lexical items and phonemes are part of the language module. In later sections when drawing parallels between comprehension and production, I will discuss evidence showing that lexical and phonological domains are governed by very different principles and operations. Thus evidence of an interaction between the two speaks against their modularity, although such evidence can only speak to the issue within the language domain. But for now, let us ask what kind of evidence Fodor would accept as evidence against the modularity of language processing.

An example of such evidence would be syntactic parsing guided by semantic context or real-world background (Fodor, 1983, p. 77). A large body of research in the '80s and '90s addressed exactly this question. Against a theory of encapsulated syntactic processing—which claims that the first-pass parsing of incoming utterances is purely syntactic and impermeable to other sources of information (e.g., Ferreira, & Clifton, 1986; Rayner, Carlson, & Frazier, 1983)—Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) showed that even the earliest moments of sentence processing were sensitive to the visual context. They presented participants with sentences such as "Put the apple on the towel in the box" and recorded their eye movements while they viewed a scene with either one or two apples. When a single apple was present, "on the towel" was often interpreted as the destination, a preference that proponents of syntactic encapsulation attributed to syntactic simplicity and the requirements of the argument structure of the verb "put". When, on the other hand, the scene contained two apples (only one of which was on a towel), "on the towel" was interpreted as modifying the noun "apple". The difference between the processing of the one- vs. two-referent scene, which was captured in early looks to objects, cannot be explained by a syntactic module that is insensitive to information provided by other systems. Later studies demonstrated the early influence of non-syntactic information from a broader range of factors such as referential domains constrained by pragmatic factors (e.g., Brown-Schmidt, & Tanenhaus, 2008), further challenging the view of language comprehension as an informationally-encapsulated system.

It is important to point out that Fodor had a good reason to insist on modular input analyzers: top-down knowledge cannot easily override perceptual illusions such as the Müller-Lyer illusion or the phoneme restoration effect. But this would only be a problem if one assumes that top-down feedback is stronger than bottom-up input. If, on the other hand, one views input analysis as primarily bottom-up, with some constraints induced by top-down processing, it is possible to accommodate the seemingly contradictory findings. There is some data to support this: we played sentences such as "She will write/see with the sharp pencil." while participants viewed black and white drawings of a pencil (target), a pen (verb competitor), a syringe (adjective competitor), and an unrelated item that was neither sharp nor used for writing. Results showed that, upon hearing the word “sharp”, there were significantly reduced looks to the syringe when the verb was “write” compared to “see”. This finding implies that top down processing (in this case the semantic constraints introduced by the verb) had affected the processing of the bottom-up information carried by the word “sharp”. Syringes are not used for writing, thus a syringe is not a plausible referent for a sharp object that one writes with. Importantly, however, hearing the verb “write” did not eliminate fixations to the syringe upon hearing “sharp”; there were still significantly more fixations to the syringe than the unrelated distractor, showing that the influence of top down constraint on processing the bottom-up input was only partial (Nozari, Trueswell, & Thompson-Schill, 2016).

In summary, the empirical evidence suggests that language comprehension is not informationally encapsulated, although top-down feedback does not completely override the influence of bottom-up information.

1.2.4. Domain-specificity

While the ’80s and ’90s were the years of intense research on informational encapsulation (i.e., interactivity), the emergence of sophisticated neuroimaging methods shifted the focus to tests of domain-specificity in the late ’90s and 2000s. Given its strong ties to cognitive neuroscience, the main question of this line of research is often how specialized certain parts of the brain are for certain operations. Fodor’s view of domain-specificity, however, was not primarily neural. For reasons I will discuss in later sections, I believe that it is useful to preserve this distinction. So let us focus for now on Fodor’s line of argument for the domain-specificity of language.

The best place to start is to recall what domain-specificity is *not*. Nobody disputes the fact that you cannot hear with your eye or see with your ear. It thus follows trivially that the mechanisms of auditory and visual perception operate on special classes of input. But it does not tell us anything about the nature of such mechanisms or whether they are similar or different. In other words, differences in the physical nature of the input, per se, are not sufficient to argue for the domain-specificity of the operations that process that input. This, however, does not imply that the nature of input is immaterial to the evaluation of domain specificity. In fact, Fodor’s argument regarding domain-specificity hinges directly on the nature of the stimulus, but his emphasis is on the complexity of the information required for analyzing the stimuli (as opposed to its physical nature). If such analyses require highly specialized information within a domain, then that domain is “eccentric”. Since there are strong arguments for the existence of language universals (e.g., Chomsky, 1975; Pinker & Jackendoff, 2005; cf. Evans & Levinson, 2009) that prevail across human languages and are absent from non-linguistic domains, Fodor concludes that the language domain is highly eccentric.

The argument for domain-specificity is that eccentric domains require specialized operations, and the degree of this specialization is commensurate with the degree of the eccentricity of the domain. But Fodor admits that this inference is loose. While it makes sense to conclude that the existence of specialized operations implies the existence of domain-eccentricity, it does *not* follow that the existence of domain-eccentric information implies the existence of specialized domain-specific operations. The latter must be proven independently. Fodor's solution to this problem is to independently assess the operations of the language faculty in light of the other criteria for modularity. If it is confirmed that language processing fulfills those criteria, this, combined with the eccentricity of the language domain, would enable one to conclude that language processing is domain-specific. As reviewed above, however, language comprehension only meets some of the criteria for modularity.

1.3. Language production in Fodor's taxonomy

As mentioned above, the bulk of Fodor's arguments concern language comprehension. He is conspicuously silent on language production. In so far as language concerns motor acts of speech, Fodor states that "It would please me if the kinds of arguments that I shall give for the modularity of input systems proved to have application to motor systems as well. But I don't propose to investigate that possibility here." (Fodor, 1983, p. 42). However, there is good reason to believe that Fodor did not view language production as reducible to a motor module, or believed that the same kind of simple processes that carry out comprehension are successful in production. For example, while defending the utility of automatic associative processing in input models (e.g., the lock-key example discussed earlier), Fodor resists the extension of associative processing to "belief" and "sentence production", arguing that they are products of "judgment" and "planning", respectively, rather than association (Fodor 1983, p. 81). This and a few other brief allusions to the nature of language production as communicating thoughts and beliefs imply that Fodor viewed language production as better aligned with "thought" (i.e., central processes) than "perception". Similarly, Levelt (1989), who views many of the production operations past the level of message generation as automatic, maintains that the first step of language production, i.e., the conceptualization of a message, is a highly controlled process, far from what can be expected from an automatic module.

While it is understandable why language production might be an ill fit for the "perceptual" category, grouping it with "thought" implies drawing a sharp dichotomy between language production and comprehension. Fodor embraces this dichotomy: in one of his few mentions of language production, he reasons that since communicating one's views through speech requires access to memory and thought, language production cannot be as domain-specific as language comprehension. Differences between language comprehension and production notwithstanding, I will argue in this section that drawing such a sharp distinction between the two, in terms of domain specificity or modularity, is contradictory even to Fodor's own views. The core of the argument is that, when examined carefully, language production and comprehension meet (and fail to meet) the same criteria for modularity.

First, in terms of domain-eccentricity, the exact same universals that apply to the perception of language apply to the production of language: native speakers in their native environment do not apply a different set of syntactic rules to what they produce and what they hear. Moreover, what is perceived during comprehension directly affects what is produced: speakers are prone to producing the same syntactic structure that they have just perceived (Bock, 1986; Pickering & Branigan, 1998). The same is true for the lower levels, acoustic features for perception and articulatory-motor features for production. Even though comprehension and production use different representations at this level, they are still governed by the same rules. For example, in a language like Japanese which does not distinguish between /r/ and /l/, both perception and production of these phonemes are affected. Perception-production loops also work at this stage: production of lexical items is altered and shaped by the speaker's perception of those items on prior occasions (Pierrehumbert, 2002). Finally, neuropsychological evidence points to a common semantic system for language comprehension and production. Individuals with semantic dementia do not just lose the ability to name a horse; they can no longer understand what the word "horse" means (Hodges, Patterson, Oxbury, & Funnell, 1992). Given this overlap between the representations in language comprehension and production, their influence on one another, and the identical rules that govern them, it would be strange to consider one an eccentric domain but not the other.

Next, let us turn to Fodor's other criteria for modularity and compare comprehension and production on those. I begin by reviewing the three criteria for automaticity: fast, mandatory and mostly subconscious processing. If we consider comprehension fast, then we must acknowledge that production is also fast, because the two are intimately tied (the normal speaking rate of a neurotypical adult speaker is 2-3 words/second, despite the large number of processes to be completed during sentence production). To argue for the mandatory nature of language comprehension, Fodor appeals to the common phenomenon of involuntary overhearing of others' speech. One could make the same argument for production by appealing to the many cases where speakers blurt out something they should not have. Documentation of (in)famous embarrassing speech errors is evidence of the ballistic nature of the language production system. But one might argue that overhearing is much more common than producing speech errors (about 1 in every 1000 words; Meyer, 1992). Note, however, that the goal is not to claim that comprehension and production are *equally* automatic; rather, that if mandatory processing without control is taken as evidence for automaticity, then such evidence exists in both comprehension and production. Finally, in terms of limited central access to mental representations, production is strikingly similar to comprehension. Speakers are aware of their thoughts and their utterances and little in between. In fact, this lack of awareness of intermediate representations and processes has been taken as evidence to argue for the automaticity of language production past the stage of conceptualization (Levelt, 1989). Language production, then, meets the same criteria for automaticity as language comprehension.

In terms of informational encapsulation, a large body of work has investigated the presence of feedback between the layers of the production system. In contrast to a strictly feed-forward model of language production (e.g., Levelt, Meyer, & Roelofs, 1999), interactive models of word production (e.g., Dell, 1986; Rapp & Goldrick, 2000) posit feedback from lower levels (e.g., phonemes) to higher levels (e.g., words) in the production system. Evidence for such feedback comes from certain patterns of speech errors: for example, phonological errors (e.g., "cap" instead of "cat") are more likely to produce a word than a nonword (e.g., Baars, Motley, & MacKay, 1975; Nozari & Dell, 2009). Similarly, phonological exchanges between the consonants of two words are more likely when the words share a vowel (e.g., "mad back" → "bad mack") compared to when they do not ("mad bike" → "bad mike") both in speaking (Dell, 1986) and in typing (Pinet & Nozari, 2018). In both cases, the effect is observed because of the feedback from segments (phonemes, graphemes, etc.) to lexical items. The overlapping segments (e.g., /k/, /æ/) help activate the competing lexical item (e.g., "cap", "bad"). These items in turn activate their segments, giving them an advantage over other segments that do not enjoy the support of lexical items which have been activated through feedback.

More recently, we have shown further evidence of feedback between segments and lexical items through incremental learning. Producing words in the context of segmentally-related words leads to interference in both spoken and written production (Breining, Nozari, & Rapp, 2016; Nozari, Freund, Breining, Rapp, & Gordon, 2016). In a nutshell, this is why: when "cat" is the target, feedback from its segments activates the competitor "cap", which in turn activates its segment /p/. Once the production of "cat" is completed, incremental learning (e.g., Oppenheim, Dell, & Schwartz, 2010) strengthens the connections between the target segments and their supporting lexical items, while weakening the connections between the non-target segments and their corresponding lexical items. Thus the connection between /p/ and "cap" is weakened. This weakening puts "cap" at a disadvantage in the next trial when it becomes the target. This mechanism depends critically on the feedback from segments to words and is also observed during learning of new vocabulary (Breining, Nozari, & Rapp, in press).

As noted when discussing informational encapsulation in the context of comprehension, Fodor does not accept interactivity within the language system as evidence against encapsulation. But denying that interaction between layers of the production system provides evidence against informational encapsulation implies that those layers are not specific enough to be considered special domains by themselves. There is, however, converging evidence pointing to separate systems for lexical and segmental processing in language production. For example, neuropsychological studies of individuals with aphasia have shown the possibility of damage to lexical retrieval processes without impaired segmental encoding and vice versa (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Dell, Schwartz, Nozari, Faseyitan, & Coslett, 2013). More importantly, viewing lexical and segmental processing as belonging to the same domain violates Fodor's own criterion for eccentric domains, as these two levels of processing are governed by very different rules. Evidence from speech errors show that lexical (but not segmental) errors tend to preserve their syntactic categories (Garrett, 1975). For example, lexical exchanges on the target "I put the napkin on the table" are much more likely to sound like "I put the *table* on the *napkin*", than "I *napkin* the *put* on the table.". However, segmental exchanges are not constrained by syntax. "I nut the papkin on the table" is not far less likely than "I put the tapkin on the nable". Instead, segmental exchanges are more likely to happen over short (e.g., within-phrase) than long distances, compared to lexical exchanges (Garrett, 1975).

These rules are not specific to English (e.g., Del Viso, Igoa, & García-Albea, 1991), and are precisely the kinds of "universals" that Fodor claims make a domain eccentric. If we accept this premise, then lexical and segmental processing are different domains, and the evidence of feedback between them should be taken as evidence against informational encapsulation, at least as far as various parts of the production system are concerned. To take matters one step further, language production is also influenced by extra-linguistic factors. For example, speakers' syntactic choices can be manipulated by cueing objects in the visual scene (Nozari & Omaki, in press), even when those cues are presented for only 60-80 ms, and none of the participants are aware of them (Gleitman, January, Nappa, & Trueswell, 2007). Moreover, speakers formulate their utterances at least partly based on their addressees' perspectives (e.g., Yoon, Koh, & Brown-Schmidt, 2012). These findings are equivalent of those reported by Tanenhaus et al. (1995) in comprehension: information from the non-linguistic (visual) domain directly influences how a sentence is formulated (in production) or parsed (in comprehension) even though the linguistic information has not changed. It thus seems reasonable to conclude that language production is not informationally encapsulated either with regard to its own sub-systems or with regard to other cognitive systems.

1.4. Summary and a way forward

To summarize, if we evaluate the modularity of language production by the same Fodorian criteria as have been applied to language comprehension, the two have striking similarities. Both are fast and show some degree of involuntary processing. Neither allows central access to intermediate mental representations. Both are highly eccentric domains (perhaps with eccentric subdomains) and the evidence suggests that both are permeable to information from other systems. From this, I draw two main conclusions: (1) given the similarities between language comprehension and production, placing them on opposite sides of Fodor's cognitive taxonomy is likely to hinder our understanding of the relationship between language and other cognitive processes. (2) The evidence I reviewed suggests that language processing meets some, but not all, of Fodor's criteria for modularity. The question thus remains: How special is language processing really?

In the remainder of this chapter, I will try to answer this question for language production. The choice of production—and not comprehension—is mainly motivated by the fact that a considerably smaller body of theoretical and empirical work in psycholinguistics has focused on how we produce—as opposed to comprehend—speech. This is, in large part, due to the much greater methodological challenges involved in studying language production: one can have perfect control over the input in order to achieve good experimental control in studies of comprehension, but it is substantially more difficult to control the thoughts and ideas behind speech planning. Methodological issues aside, the above review showed many commonalities between comprehension and production, so the main approach that I will take here can be (and in some cases has been) extended to comprehension. At the same time, Fodor is absolutely correct that production does differ from comprehension in one major respect: it comes from one's own thoughts. Therefore the empirical findings from language comprehension research cannot be readily extended to production. Production must be studied on its own, with the acknowledgment that production and comprehension are intimately connected processes.

Since Fodor's placement of language production in his cognitive taxonomy proved problematic, the first step is to redefine a framework within which the "specialness" of production can be assessed. We can do so while staying faithful to Fodor's general idea: while we agree that language is an eccentric domain, we can ask: Are its operations also specialized to this domain, or are they shared with other domains? If the cognitive processes that mediate language production are specific to the production system, then language production can be considered a special system separate from other cognitive systems. If, on the other hand, these processes are shared with other cognitive systems, then language production is simply part of domain-general cognition, albeit a highly sophisticated part.

The question now is which processes are reasonable targets for such a study. If we choose to focus on operations that are closely tied to language-specific representations, we will soon come to an inevitable dead end. For example, the sensitivity of lexical errors to grammatical class shows that knowledge of syntactic categories limits lexical selection, e.g., “I put the ----” only accepts a lexical item from the “noun” category. We can now argue that, since the concept of a “noun” is undefined outside of the linguistic domain, then, by definition, a noun-activating mechanism must be specific to language. This is unhelpful. But we can still target the same process from a slightly different angle: when, for example, “table” and “napkin” are both viable candidates for selection, what kinds of processes determine which one is produced? Do these processes have anything in common with processes that adjudicate between response alternatives in other cognitive systems, regardless of the nature of those responses? Are the principles behind detecting competition, resolving competition, and learning to better regulate the system based on this competition shared between language production and other systems?

These questions fall under the general framework of what I call “monitoring and control” in language production, and I argue that they provide the soundest test of the “specialness” of the language production system. An obvious objection might be that such processes are “external” to language production, and as such cannot provide insights into the operations of the language production system per se. For example, one could imagine an external monitor that only peeks at the final output of the production system and issues a pass/fail decision. Presumably such a monitor would not reveal much about the internal dynamics of the production system, and would hardly be useful in evaluating how special its internal operations are—or are not—compared to other systems. However, in what follows, I will argue that such a view of monitoring and control in language production is wrong. Monitoring and control are intimately tied to production processes and are guided, in large part, by information generated within the production system itself. The critical question is whether they also have components shared with other systems or not.

2. Language production and the need for control

Earlier I argued that language production is fast, to some degree ballistic, and that speakers are deprived of access to its intermediate representations. In addition to these properties, language production is extremely well practiced, at least by the time one reaches adulthood. These characteristics make production a good candidate for an automatic process (Levelt, 1983), i.e., a process that does not require cognitive control past the level of conceptualization. If this is in fact true, the whole framework I have proposed is invalid. But, before accepting this, it is important to note that the tasks in which automaticity has been claimed based on the above-mentioned criteria often consist of a repetitive procedure with only small deviations from the task routine. Even a complex task like driving from one’s home to work involves following the same route over the same roads and turns that one drives every day in the same car, more or less in the same traffic and with the same speed. On the other hand, we almost never repeat the exact same set of sentences in the same order to the same person on two occasions. The highly generative nature of language production invites caution in evaluating its automaticity according to the same criteria that have been applied to tasks that are not nearly as creative. But if production is not automatic, how is it so fast, fluent, and relatively error-free? I will argue below that the key is an efficient monitoring-control loop that constantly evaluates the need for control and deploys the control necessary to keep performance optimal.

2.1. Control at the level of word production

To illustrate why producing a word may require cognitive control, I will present a brief overview of the word production system. While details differ, the backbone of the architecture that I review here is shared between several major computational models of word production (Dell, 1986; Levelt, Meyer, & Roelofs, 1999; Rapp & Goldrick, 2000). It is now accepted that producing a word (i.e., getting from meaning to sound) requires at least two stages of processing: in the first stage, semantic features (e.g., animal, four-legged, pet, meows, etc.) must be mapped onto lexical items (e.g., "cat"; Fig. 1). Because semantic features are shared between several items, this process often activates not only the target, but also semantically similar lexical items (e.g., "dog"). In systems with cascading and feedback (e.g., Dell, 1986; Rapp & Goldrick, 2000), the activation from lexical items "trickles down" to segments (phonemes in spoken and graphemes in written production), and those segments, in turn, send activation back to the corresponding lexical items. Whether the system has feedback or not, at some point one of the activated representations at the lexical level is selected for further processing. This marks the end of the first stage of word production.

In the second stage, the selected lexical item is mapped onto its segments. In systems that allow for feedback, activation of segments leads to the activation of segmentally-related words (e.g., "cap"), which in turn activate their own segments to compete with the target's segments. This step also ends with a selection process, which selects segments for their proper positions (segmental encoding). The outcome of this process is passed down to the articulators for articulatory-phonetic encoding and ultimately overt production.

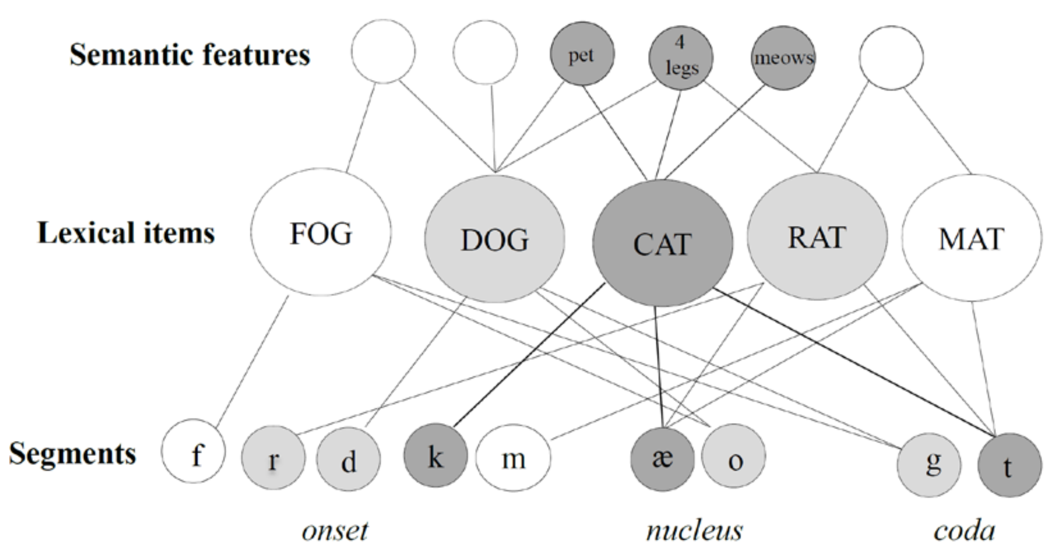

Figure 1. A schematic of the two-step architecture of word production before articulatory-phonetic encoding (e.g., Dell, 1986). The first stage ends by selecting one of the activated lexical items. The second stage ends by selecting one of the segments in each position.

The need for control arises because, in each stage, the spreading activation processes activate multiple lexical items (cat, dog, etc.) and multiple segments (/t/,/p/, etc.) in each position, but only one representation can be selected at each level (see Fig. 1). Empirical evidence shows that greater activation of competitors makes production more difficult by increasing error rates and/or response latencies: naming pictures is harder in the presence of competing words (e.g., Schriefers, Meyer, & Levelt, 1990) and in the context of other semantically-related (e.g., Belke, Meyer, & Damian, 2005; Schnur, Schwartz, Kimberg, Hirshorn, Coslett, & Thompson-Schill, 2009) or segmentally-related (Breining et al., 2016; 2018; Nozari et al., 2016) pictures, at least during tasks that elicit repeated naming of a small set of pictures. It is thus reasonable to hypothesize that a control operation which suppresses competing representations in favor of a single response, i.e., inhibitory control, might be at work in word production.

Consistent with this prediction, Shao, Roelofs and Meyer (2013) reported a correlation between response latencies in naming pictures in the presence of distractors and participants' inhibitory control abilities as measured by the stop-signal task. Importantly, Shao, Roelofs, and Meyer (2012) also reported a correlation between measures of inhibitory control and naming pictures of objects and actions in the absence of overt distractors, demonstrating that the need for control was not specific to cases where competition was enhanced. The correlational nature of this evidence from neurotypical speakers might cast doubt on whether inhibitory control is essential to word production; perhaps it simply helps speed up production through strategic processes that are not really necessary for word production. I recently reported a case study of an individual with aphasia who produced many semantically-related errors, in quick succession, while naming a picture (e.g., "peach, orange, no peach, pineapple!" in response to a picture of a watermelon). Using a battery of linguistic/cognitive tasks I showed that the individual had intact semantic, lexical, and phonological representations and was able to activate lexical items from semantic features. He was, however, selectively impaired on tasks of inhibitory control (e.g., the Simon task), and his performance quickly deteriorated under conditions that increased the activation of competitors (Nozari, under review). This case provides evidence that selective impairment of inhibitory control creates problems during lexical selection, which in turn implies that selection processes are dependent on inhibitory control.

2.2. Control beyond single word production

Since conversations are made of sentences and sentences of words, it would not be surprising to find that the need for inhibitory control at the level of word production extends to sentence production and ultimately conversation. Therefore, instead of focusing on what has to be logically true, I will discuss some evidence for the need for control in operations that are at work during sentence production and conversation but are not part of single-word production. Producing sentences involves not only retrieving words, but also keeping track of dependencies between different parts. For example, in English—and many other languages—the verb must agree in number with its subject. If the sentence has more than one noun phrase, as in "The lion next to the green snakes…”, sometimes speakers mistakenly produce a verb that agrees with the local noun (snakes) instead of the subject (lion). This phenomenon is called “agreement attraction” (Bock & Miller, 1991), and is most common when a singular subject noun is competing against a plural local noun. We recently showed that increasing the demand for control during the assignment of thematic roles increases the chance of attraction errors (Nozari & Omaki, in press). Moreover, analysis of individual differences using path modeling showed that individuals with better inhibitory control abilities were less likely to produce attraction errors. At the level of conversations, beyond producing words and sentences, speakers must also tailor their utterances to the referential context and to their listeners’ perspectives. Analysis of individual differences has shown that both of these are predicted by speakers’ inhibitory control abilities (Trude & Nozari, 2017; Wardlow, 2013).

3. Assessing the need for control: The primary job of a monitoring system

The findings I reviewed above suggest that inhibitory control plays a role at all levels of language production, from producing single words to conversations. This evidence argues against a strong notion of automaticity in language production and calls for mechanisms to assess the need for control. Recall that it is the nature of such mechanisms that can shed light on the “specialness” of language production. Production would not be so special if the regulation of its primary processes such as lexical selection depended on the interaction between the production system and other (domain-general) systems which regulate similar selection processes in the non-linguistic domain. This regulation is the job of the monitoring system. Therefore, to understand how special the production system is, we must adjudicate between different accounts of monitoring in language production.

3.1. Production-external monitoring

The classic view of a speech monitor is a device that captures errors (Levelt, 1983; 1989). Levelt (1983) tasks the comprehension system with this job: a “perceptual loop” monitors the output of the production system (e.g. “cap”), compares it to an intended target (“cat”), and concludes that the speaker has made an error. This proposal is quite elegant, as it proposes that error detection for both self- and other-produced speech is carried out using the comprehension system. If true, such a monitoring mechanism is perfectly compatible with a modular production system, as it operates outside of the production domain and only examines production output once the primary production processes have been completed. However, neuropsychological data suggest a dissociation between monitoring one’s own speech and that of others, as well as a double dissociation between comprehension abilities and detecting one’s errors (see Nozari et al., 2011 for a review of this evidence). This is not to imply that comprehension plays no role in monitoring. Altered auditory feedback indeed changes speakers’ perceptions of what they have produced (Lind, Hall, Breidegard, Balkenius, & Johansson, 2014). The critical question is whether the comprehension system, by itself, is an efficient enough monitor.

Defining the job of a monitor is a key point here. If one narrows monitoring to the detection of errors, especially after they emerge, the comprehension system is not a bad candidate. But I set up this chapter to show the reader that production monitoring goes far beyond error detection. The design and functioning of the production system, i.e., the interconnected nature of the representations along with the activation of multiple alternatives at each stage of processing, might be expected to make production very error-prone. However, production errors are rare in neurotypical adult speakers because production is efficiently controlled at various stages. Such control can only be deployed efficiently if there is a monitoring mechanism that constantly assesses the need for control. Viewed this way, the primary job of a monitoring system is to *prevent* errors from happening. Catching errors which have already happened is a much smaller part of the job.

In order to achieve this goal, the monitor must work closely in parallel with primary production processes. It thus has to be fast and not require much conscious effort, or it would make production slow, effortful, and disfluent. Importantly, it must have access to various stages of production because, as discussed above, the need for control arises at different levels. A monitor that only has access to the output of the production process is simply too slow to do much good. If we want to accept comprehension as the monitor for the production system, we must define a comprehension system that has access to intermediate representations in production. But as Fodor pointed out, speakers do not have access to the intermediate linguistic processes, so this access must be subconscious. Moreover, the comprehension system should be able to perform a comparison between the representation that is about to be selected and the target, while this comparison remains subconscious. One could try to redefine comprehension to accommodate this definition, but it would be a different process than assumed by the perceptual loop, which is "conscious and attentional" and "more or less…deliberate" (Postma, 2000, p. 115).

3.2. Semi-production-external monitoring

A fully production-external monitor that evaluates the intermediate representations of the production system on a constant basis is not very plausible on both theoretical and empirical grounds that I discussed in the previous section. There is, however, one kind of monitoring that can be achieved through an interaction between the production and perception systems (hence I refer to it as semi-production-external) in a fast unconscious manner. Hickok (2012) defines such a monitor as follows: activation of semantic representations during conceptualization of a message activates both production and perception processes. At lower levels, the production and perception systems' representations diverge: we have articulatory-motor representations in production and acoustic representations in perception. Hickok argues that, under normal circumstances, as the motor plan for "cat" gains activation, the perceptual representation of "cat" is gradually suppressed via inhibitory connections. If, however, an erroneous word (e.g., "dog") is sent to the articulator, its motor representation cannot suppress the perceptual representation of "cat", simply because the two are not connected in the system. This leads to persistent activation of the perceptual representation of "cat" which is read as an error alarm by the system.

This account is plausible for several reasons: (a) articulatory-phonetic and acoustic representations are distinct and most likely connected (we can learn to produce /k/ by listening to someone else's production of /k/), (b) perceptual representations in the auditory cortex are activated quickly during production (Tian & Poeppel, 2010), and (c) similar mechanisms are at work in other domains where motor actions have perceptual consequences. But precisely the same reasons that make this kind of monitor a viable model for monitoring motor aspects of speech make it much less of a viable candidate for monitoring at higher levels of language production. There has been much interest in recent years in extending this model (sometimes referred to as the "forward model" or "internal model") to monitoring at lexical and syntactic levels of production (e.g., Pickering & Garrod, 2013). Such a model, however, must assume the existence of *two* sets of such representations (i.e., different lexical nodes for production and one for comprehension, equivalent to motor and perceptual representations) and their simultaneous activation during production. I am not aware of any theoretical arguments that motivate double sets of representations at the higher cognitive levels, or of empirical evidence showing that such double sets exist[1]. This is, of course, not proof that they do not exist, but for such a monitoring model to be proposed as a viable alternative at that level, verification of its basic assumptions is a necessary first step. Thus, to the degree that control is required at pre-articulatory levels, it is unclear whether this kind of monitoring mechanism provides the right answer.

3.3. Production-internal monitoring

Theoretically speaking, the system with the fastest and most complete access to information in the production system is the production system. One should then at least consider the possibility that the production system has a strong hand in monitoring itself. This idea has been floating around since the early '80s (e.g., De Smedt & Kempen, 1987; Laver, 1980; MacKay, 1992; Schlenck, Huber, & Willmes, 1987). The original proposals (e.g., Laver, 1980) still viewed monitoring as a comparative process, in which the produced response was compared to the target response. When situated within a production-internal model, such a comparative process would require the target to have been available in the production system, in which case one might wonder why the correct target was not produced in the first place. A major theoretical development was the proposal of non-comparative monitoring. For example, MacKay (1992) proposed that monitoring can still be successful without the monitor having access to the right response. Instead, it would detect a new pattern (e.g., of activation in the network), one not previously experienced by the speaker, as an error signal. This perspective had a groundbreaking implication: it viewed conscious awareness of an error as a *consequence* (and not a precursor) of detecting the error. When the system falls into an unusual state, it triggers an error message which signals to the speaker that something is not right.

The most recent proposal for a production-based monitor is the *conflict-based monitor* (Nozari et al., 2011; Hanley, Cortis, Budd, & Nozari, 2016), which was motivated by a number of properties I have discussed in earlier sections: (1) It is parsimonious in the sense that it uses the information already generated by primary production processes, so no assumptions about a new set of representations are required. Using production-internal information also allows for fast monitoring at all levels of production without privileged access to intermediate representations

[1] Note that by "representation" here, I am referring to the nodes (and not the connections) in a network. While the body of knowledge is not stored only in nodes, but also in the connections between nodes that comprise a network, the concept of activation —and inhibition of activation— applies to nodes. Thus a model such as a forward model would require two sets of nodes, and not simply separate connections. It is this duplication of nodes as representations that I question.

by an external system. (2) It in non-comparative, i.e., it is blind to the "right response", thus avoiding the problem of producing an error despite having the correct response already at hand. Instead, it uses a certain pattern of information about the state of the production system to deduce the need for additional control. (3) The information it employs is precisely the information needed to deploy control to the production system. When reviewing the need for control at multiple levels of the production system, I pointed out that one of the major reasons—perhaps *the* major reason—for the necessity of control in production is competition between alternative representations at each level (e.g., lexical, segmental, syntactic, etc.). It is thus reasonable to posit that information about such competition is what is being monitored. Specifically, the conflict-based monitor is sensitive to the difference between the activation of two (or more) representations in the same part of the production system (e.g., the lexical layer). If this difference is small, the representations are in high competition with one another. This is exactly the kind of situation in which the deployment of cognitive control is necessary to resolve the competition. Importantly, the same kind of mechanism can determine the need for control in other (non-linguistic) systems as well.

3.4. Monitoring as decision making

In the above sections, I argued that the best source of information for a monitor whose goal is to quickly determine the need for control in the production system is information about the level of competition between alternative representations in each part of the system. This information can be quantified as *conflict*, i.e., the inverse of the difference in the activation between two (or more) representations (Nozari et al., 2011). Conflict is high when several items have similar levels of activation, i.e., a little bit of noise can create a high chance for any of them to be selected as the response. If only one response is correct, then the chance of committing an error is proportional to the chance of the competing representations being selected. Conflict is thus a direct index of the *probability* of an error: high conflict equals a high probability of making an error. Note that the system does not need to know what the right response is for this to be informative. When "cat" and "dog" have comparable levels of activation, there is a 50% chance that one of them gets selected. This chance is the same whether "cat" is the actual target or "dog". Moreover, this index allows for *prospective* assessment of error commission; the monitor does not have to wait for selection to be over before it can determine the need for control. It can thus act as a preventive measure to keep the system from making errors, a function I defined earlier as the primary function of the monitoring system.

But here is the hard question: How much conflict is too much conflict and should signal the need for control? This is the kind of question that all non-comparative theories of monitoring—which do not require comparison of a response to the "right" response—must address, as all of them work with some form of information that serves as a proxy for error probability. I believe the answer is that monitoring is a form of decision making. This is illustrated in Figure 2, using a framework adopted from the signal detection theory. The figure shows two hypothetical distributions of conflict, corresponding to correct and error responses. As explained above, error trials are associated with higher levels of conflict (see Nozari et al., 2011 for a computational demonstration of this claim), so the distribution of errors falls to the right of the distribution of correct trials. However, not all high-conflict trials will necessarily lead to errors, because there is always some chance of the correct response being selected from among the activated alternatives, so the two distributions have some overlap.

The monitor determines which responses to flag as potential errors by setting a criterion somewhere between the two distributions. The location of this criterion is determined by the speaker's goals: if the goal is to produce speech which is free of any errors, the criterion can be placed far to the left so that all errors are caught. Note, however, that this will come at the cost of increased "false alarms". These false alarms can be observed as overt rejection of the correct response in individuals with aphasia (Nozari et al., 2011), but their more common manifestation in neurotypical adult speakers is disfluency (e.g., Postma, Kolk, & Povel, 1990). If, on the other hand, speed and fluency are to be emphasized over accuracy, the criterion would be placed to the far right. When neither speed nor accuracy is to be prioritized over the other, one way to determine the position of the criterion is by allowing it to settle at a point of equilibrium; for example, a point at which the probability that the criterion would have to be moved to the left to avoid a miss would match the probability of it having to be moved to the right to avoid a false alarm (Kac, 1962). Our computational simulations showed that this strategy predicts a detection rate of around 50% with a false alarm rate of <1% (Nozari et al., 2011). This detection rate matches what has been reported in conversational speech (e.g., Nooteboom, 2005) and what we have observed in my lab during typing of single words.

A useful feature of this framework is the possibility of extending it from detecting an error once selection is completed to preventing an error before a response has been selected. We have recently shown that differences in the pattern of responses (longer response latencies and more omission errors vs. shorter response latencies and more commission errors) within the same paradigm can be explained by shifting the position of a criterion that determines whether lexical selection should be allowed to proceed or halted until competition can be resolved, most likely by the intervention of control processes (Nozari & Hepner, 2018).

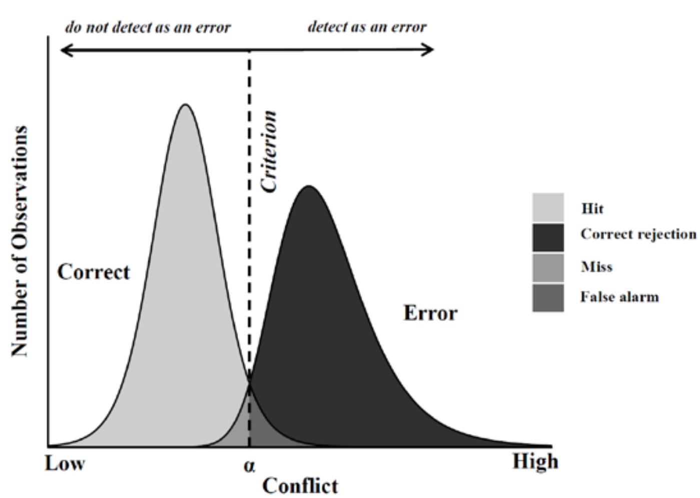

Figure 2. Conflict-based monitoring in the signal detection framework.

A complete picture is beginning to emerge now: monitoring is a process of decision making, with a flexible criterion, using information generated during production processes to determine the probability of an error. In neurotypical systems, in most cases, this information is used before selection to recruit the control necessary to prevent errors from occurring. Occasionally, however, the relevant information may only be available after selection has been completed. In such cases, the monitor acts as an error detector, but the main function of monitoring remains surveying the system in order to determine the need for control. Such a monitor plays a critical role in regulating and optimizing production processes. For example, when faced with a high-conflict picture naming trial, it can tune the production system to more efficiently resolve conflict in the next high-conflict picture naming trial (e.g., Freund, Gordon, & Nozari, 2016). Importantly, this adaptation from one high-conflict trial to the next seems to persist despite a temporal gap between the two trials and regardless of the nature of the intervening tasks, suggesting the involvement of learning mechanisms in the allocation of control (Freund & Nozari, 2018).

4. Domain-generality of monitoring and control in language production

What I have explained so far can be summarized in a few sentences: language production involves competition at all levels. If unresolved, this competition leads to frequent errors, as is the case with aphasic individuals. However, competition is often successfully resolved because the production system is constantly monitored and control is appropriately applied. Since competition and its resolution are integral parts of the production processes, the operations that affect competition resolution are also part of successfully producing speech. We can now evaluate whether such processes are specialized for language production, or are shared with other systems, and by answering this question determine how "special" language production is relative to the rest of cognition.

In the early parts of this chapter, I reviewed what Fodor considered to be the criteria for computational domain specificity, and alluded to the surge of interest in probing neural domain-specificity in the late '90s and 2000s. Most recently, the investigation of domain-generality has turned to paradigms that assess whether an increased need for control in one task/domain leads to the recruitment of control in another task/domain (see Egner, 2014 for a review). These three approaches have arrived at different answers to the question of the domain-generality of cognitive functions—partly because of methodological variations, but more importantly because the question turns out to have different meanings when viewed from different angles (Nozari & Novick, 2017). I will thus briefly review the evidence for domain-generality from these three angles.

4.1. Domain-generality of computations

Competition is everywhere in cognitive systems, and must often be settled in favor of a single winner. It is thus reasonable to assume that the mechanisms that mediate conflict resolution apply across domains. The evidence I reviewed on the role of inhibitory control in various levels of language production lend support to this claim. Analysis of individual differences suggest that individuals' ability to inhibit a response in tasks like stop-signal or go-nogo tasks predicts how quickly and accurately they can name pictures, produce the correct subject-verb agreement, and tailor their utterance to the referential context. While some part of this process may be carried out locally within each system through lateral inhibition between representations, neural evidence suggests that at least some part of the process is mediated by central processes (see below). By the same logic, it is plausible that the need for control is assessed in similar ways across systems that require competition resolution. The conflict-based monitor in speech production was, in fact, inspired by a similar model applied to forced-choice manual responses in a variety of tasks which induce conflict between response alternatives (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Similarly, semi-production-based monitors such as Hickok's (2012) model use the same principle of suppression of a perceptual response by its motor equivalent that has been reported in vision.

In short, similar computational problems exist in language production as in other systems. As such, similar computational solutions are applicable to both, as long as the nature of the representations are compatible with those solutions. A general mechanism of motor control through perceptual suppression is applicable where motor and perceptual representations are concerned in the language system. A general conflict detecting mechanism is applicable where several representations of the same kind are simultaneously competing for selection.

4.2. Domain generality of neural underpinning

A thorough review of the neural underpinnings of language production is beyond the scope of this chapter, but I will review a few key pieces of evidence related to the issue of the domain-generality of monitoring and control processes in speech production. Generally speaking, three techniques have been used to probe this issue: event-related potentials (ERPs), neuroimaging (mostly fMRI) and brain stimulation. The first of these is concerned more with uncovering common neural signatures (if they exist) than with pinning down that signature to a brain region with millimeter precision. There are now several ERP studies showing a common signature (called the ERN, or error-related negativity) for monitoring in speech and non-speech domains (e.g., Ganushchak & Schiller, 2008; Riès, Janssen, Dufau, Alario, & Burle, 2011; see Nozari et al., 2011 for a comprehensive review of this evidence). The ERN is a modality-independent, response-locked ERP component with greater negativity on error than correct trials. A similar component, N2, is observed on high-conflict correct trials, whose origin, similar to the ERN, can be traced back to the anterior cingulate cortex (ACC; e.g., Van Veen & Carter, 2002). These results strongly suggest the presence of a common monitoring mechanism across domains, with a central component, housed in the ACC, which is sensitive to the need for control.

What happens when the need for control is detected? Decades of research on the prefrontal cortex have implicated the VLPFC as being involved in deploying control based on task demands (e.g., Thompson-Schill et al., 1997; see Nozari & Thompson-Schill for a review). Neuroimaging studies of monitoring and control in language production are sparse, but the results implicate the involvement of the same neural regions observed during monitoring and control of other tasks. For example, Gauvin, De Baene, Brass, and Hartsuiker (2016) found activation of the supplementary motor area (pre-SMA), dorsal ACC, and VLPFC during monitoring of a tongue-twister task. Moreover, increased competition by repeated naming of semantically-related pictures has been shown to activate the VLPFC (e.g., de Zubicaray, Fraser, Ramajoo, & McMahon, 2017; Schnur et al., 2009). Similarly, stimulation of the left PFC through transcranial direct current stimulation (tDCS) has been shown to decrease the rate of semantic errors (Nozari, Arnold, & Thompson-Schill, 2014) and phonological errors on words the speakers attend to (Nozari & Thompson-Schill, 2013).

While the neuroimaging and brain stimulation studies provide some evidence for the involvement of the medial and lateral PFC in implementing monitoring and control in language production, this evidence does not imply that the *same population of neurons* in these regions is involved in both linguistic and non-linguistic tasks. For example, it is possible that, at the individual level, slightly different parts of the PFC mediate control in linguistic and non-linguistic tasks (e.g., Fedorenko, Behr, & Kanwisher, 2011), although the significance of this finding must be interpreted in light of the physical constraints of a prefrontal cortex that must accommodate connections from various specialized systems such as language, vision, etc.

In summary, the investigation of common neural correlates for monitoring and control between language production and other domains is relatively new. The evidence so far points to some level of domain-generality, although the exact degree of neuronal overlap within these domain-general regions (e.g., ACC, LPFC) between monitoring and control of language and other domains is not yet known.

4.3. Domain generality of functional adjustment in control

The evidence for domain-generality in terms of computations and neural substrates has led researchers to push the boundaries on the issue by asking whether the increased need for control in one domain recruits control that can then be applied to another domain. The hypothesis here is that signaling the need for control, regardless of the source of conflict, mobilizes domain-general control resources that can be readily applied to any task in any domain. A small number of studies have reported results that support this prediction. For example, Kan, Teubner-Rhodes, Drummey, Nutile, Krupa, and Novick (2013) reported that reading a high-conflict (syntactically-ambiguous) sentence, which signals the need for control to the monitor, led to the recruitment of control that improved performance on high-conflict trials from a word Stroop or a visuospatial task (see also Hsu & Novick, 2016). In contrast to these results, a large body of studies have found no evidence of cross-task transfer of control (e.g., Akçay & Hazeltine, 2011; Egner, Delano & Hirshc, 2007; Kiesel, Kunde, & Hoffmann, 2006; see Egner, 2014 for a review).

We recently tested this prediction directly in language production. Trials from a picture-word interference (PWI) paradigm were interleaved with trials from a visuospatial task (Exp. 1) and a sentence comprehension task (Exp. 2; Freund & Nozari, 2018). All three tasks were designed to contain high-conflict trials (e.g., semantically-related words superimposed on target pictures, arrows with opposing directions, or lexical or syntactic ambiguity, respectively) and low-conflict trials (e.g., target word superimposed on the target picture, same direction arrows, no ambiguity). Since PWI trials alternated with trials from the other tasks, it was possible to assess two effects: *within-task adjustment of control* was assessed by measuring performance on the current trial as a function of a 2-back trial from the *same* task (for each of the PWI, visuospatial, and sentence comprehension tasks), and *cross-task adjustment of control* was assessed by measuring performance on a current trial as a function of 1-back trial from a *different* task (e.g., PWI as a function of visuospatial conflict, etc.).

The results were clear: all three tasks showed robust evidence of within-task adjustment of control, despite being separated by a lag of 4s (Exp. 1) and 8s (Exp. 2) and a trial from a different task to perform in between. On the other hand, there was no sign of performance improvement in any of the three tasks as a function of increased demand for control on the previous trial from a different task. These results speak against a domain-general account in which control is deployed indiscriminately to all systems, and instead indicate that deployment of control is specific to the source of conflict: if conflict is detected in the language production system, control is deployed to that system. In fact, quite possibly, if the need for control is detected in a specific part of the production system (e.g., to resolve lexical competition), then control is deployed to exactly that part of the system (Nozari et al., 2011).

Interestingly, there were suggestions of a *reversed* effect: in some cases, the increased need for control in one task led to the *deterioration* of performance on a high-conflict trial on another task (see Notebaert & Verguts, 2008, for a similar finding). This reversed effect speaks against completely independent control systems. How can we reconcile all these findings? This pattern of results is compatible with a central controller (e.g., part of the LPFC), with at least some level of domain-generality in terms of neurons that implement control, but with specific connections to domain-specific representations. Control is implemented in each task by a special configuration of connections between the central controller and task-specific representations. Critically, this configuration can be adapted in order to facilitate performance the next time the system experiences high conflict in the same task (Freund & Nozari, 2018; Notebaert & Verguts, 2008). Similar to incremental learning processes during lexical selection (Oppenheim et al., 2010) and segmental encoding (Breining, Nozari, & Rapp, 2016; in press), the connections between the part of the central controller and the domain-specific representations involved in this configuration are strengthened. At the same time, the connections between the same population of neurons in the central controller and other systems are weakened, in order to allocate the limited resources efficiently to the immediate needs. To the extent that the neuronal populations in the central controller overlap in implementing control across tasks, this mechanism of incremental learning facilitates performance on the next trial of the same kind of task, and interferes with the implementation of control on a trial of another task whose connections to the controller have been weakened. Note, however, that in many cases control simply does not transfer from one task to another, and no reliable reversed effect is observed. This finding implies that the overlap between the neuronal populations in the central controller for mediating control in different tasks is limited (see Jiang & Egner, 2014, for neuroimaging evidence supporting this claim).

In summary, increasing the need for control in the language production system does not lead to enhanced implementation of control in other systems and vice versa. This is, however, completely expected from the operation of a limited-capacity central system which is trying to optimize task performance: distribution of limited resources is most effective when control is deployed exactly where it is needed. At the same time, there is some indication that the neuronal populations in the central controller which implement control in different tasks/domains may indeed have a certain degree of overlap.

5. Conclusion: How special is language production?

We are finally in a position to answer this question: language production is special in so far as its representations are special. I have claimed that even subdomains within production, e.g., lexical items vs. segments, are each special in that they meet Fodor's criteria for eccentricity. What was in question was the assumption that a system with special representations requires specialized operations. Part of the argument for a specialized system was efficiency. Language production is fast, fluent, and relatively error-free, and without the speaker's conscious awareness of the processes that map thoughts onto sounds. So perhaps it is a highly specialized automatic module. The evidence I reviewed in this chapter suggest otherwise. The nature of production processes engenders competition at all levels of production, and empirical evidence suggests that such competition indeed requires control in order to be resolved properly. I thus proposed an alternative: language production is so fast and efficient because of a highly efficient monitoring-control loop which constantly monitors the need for control and consequently deploys control to the appropriate part of the production system. I then reviewed various accounts of monitoring and argued that a specialized monitoring system such as the perceptual loop system is, in fact, neither the fastest nor the most efficient option.

While the contribution of the perceptual loop to detecting errors in overt speech is undeniable, I also discussed two domain-general mechanisms for speech monitoring, both of which are good candidates for a fast and unconscious monitoring process. The first is a forward model which suppresses the perceptual consequences of self-produced actions. This is an excellent candidate for monitoring speech at the articulatory-motor level. Fodor's argument about eccentricity has a critical implication here: while eccentric representations do not necessarily require specialized operations, certain operations do require specific kinds of representations. A monitor that guides motor actions via perceptual consequences of those actions requires separate sets of corresponding motor and perceptual representations. If it is to be applied to subsystems which lack such representations, the process must be redefined.

The second mechanism is a conflict detection mechanism which monitors the difference between the activation of multiple representations in the same part of the system. This is an excellent candidate for monitoring language production, the dynamics of which naturally lead to the activation of multiple representations at various levels. The plausibility of these two mechanisms for speech monitoring (see Hickok, 2012; Nozari et al., 2011; Hanley et al., 2016 for implementations and detailed discussion of the empirical evidence in various populations), together with the empirical evidence for the domain-generality of monitoring and control in language production (e.g., ERP, brain stimulation, and neuroimaging evidence) strongly suggests the involvement of a domain-general process in the regulation of speech production.

In the last section, I cautioned against a sweeping interpretation of this domain-generality. The key is in remembering that these domain-general operations still operate on domain-specific representations. Functionally, this means that a monitoring-control loop for lexical selection consists of domain-specific lexical representations, a domain-general central control system (the same general region(s) in the brain, with some overlap between the subregions that mediate control in specialized domains), and the connections between the two. Signals generated within this loop during monitoring recruit control, which, in neurotypical individuals, is commensurate to the need for control. This control, however, does not benefit other systems, as they are simply not part of this particular loop. In fact, in so far as the same neuronal populations in the central controller engage in mediating control across different loops, implementation of control in other loops suffers, just like in the case of overt selective attention. We have shown that in language production, as in other domains, attending to a word increases accuracy on that word at the expense of accuracy on the other words in the sequence (Nozari & Dell, 2012). Our later work showed that stimulation of the central controller (left PFC) exaggerates this effect by widening the gap between the accuracy of attended and unattended items as control increases (Nozari & Thompson-Schill, 2013). Both the conscious attentional effects manipulated in these two studies and the subconscious control effects discussed throughout this chapter follow the same principle: functional specificity is an inevitable consequence of systems that share a resource-limited domain-general component. This combination of shared computations through (partially shared) machinery, along with specific loops to monitor and regulate domain-specific operations is an optimal solution to guarantee fast, efficient and relatively error-free processing.

To conclude, production processes are monitored and regulated through operations that are very similar to those in other systems, contrary to the notion of special mechanisms for special domains. In that sense, language production is not that special. However, the specificity of linguistic representations to the language domain creates special monitoring and control loops in production that can be selectively trained and selectively damaged. This specificity has a critical clinical implication: the most effective way to train or rehabilitate control in the production system is by using tasks that engage the monitoring and control loop in the context of language production. In this sense, language production is as special as any other cognitive system.

Acknowledgments

This work was supported in part by the Therapeutic Cognitive Neuroscience Fund at Johns Hopkins University, and in part by the NSF grant NSF-1631993 to Nazbanou Nozari and Akira Omaki. I would like to thank Christopher Hepner, Duane Watson and Kara Federmeier for their feedback on this chapter.

References

Akçay, Ç. & Hazeltine, E. (2011). Domain-specific conflict adaptation without feature repetitions. *Psychonomic Bulletin & Review*, *18*(3), 505-511.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior*, *14*(4), 382-391.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*(4), 667-692.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, *18*(3), 355-387.

Bock, K., & Miller, C. A. (1991). Broken agreement. Cognitive Psychology, 23(1), 45–93.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624.

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic bulletin & review*, *23*(2), 500-506.

Breining, B., Nozari, N. & Rapp, B. (in press). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive science*, *32*(4), 643-684.

Chomsky, N. (1975). Reflections on language. New York: Pantheon

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

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, *104*(4), 801.

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*(3), 380-396.

De Smedt, K., & Kempen, G. (1987). Incremental sentence production, self-correction and coordination. In *Natural language generation* (pp. 365-376). Springer, Dordrecht.

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

de Zubicaray, G., Fraser, D., Ramajoo, K., & McMahon, K. (2017). Interference from related actions in spoken word production: Behavioural and fMRI evidence. *Neuropsychologia*, *96*, 78-88.

Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247.

Egner, T., Delano, M., & Hirsch, J. (2007). Separate conflict-specific cognitive control mechanisms in the human brain. *neuroimage*, *35*(2), 940-948.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, *32*(5), 429-448.

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428-16433.

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language*, *25*(3), 348-368.

Fodor, J.A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Freund, M., Gordon B., & Nozari, N. (2016). Conflict-based regulation of control in language production. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.),*Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1625–1630). Austin, TX: Cognitive Science Society.

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

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. *Neuroimage*, *39*(1), 395-405.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 133-175). San Diego, CA: Academic Press.

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

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of memory and language*, *57*(4), 544-569.

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

Hickok, G. (2012). Computational neuroanatomy of speech production. Nature Reviews Neuroscience, 13(2), 135.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*(6), 1783-1806.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological science*, *27*(4), 572-582.

Jiang, J., & Egner, T. (2014). Using neural pattern classifiers to quantify the modularity of conflict–control mechanisms in the human brain. *Cerebral Cortex*, *24*(7), 1793-1805.

Kac, M. (1962). A note on learning signal detection. IRE Transactions on Information Theory, IT-8, 126–128.

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, *129*(3), 637-651.

Kiesel, A., Kunde, W., & Hoffmann, J. (2006). Evidence for task-specific resolution of response conflict. *Psychonomic Bulletin & Review*, *13*(5), 800-806.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), Errors in linguistic performance: Slips of the tongue, ear, pen, and hand (pp. 287–305). New York: Academic Press.

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

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press. Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, *25*(6), 1198-1205.

MacKay, D. G. (1992). Awareness and error detection: New theories and research paradigms. Consciousness and Cognition: An International Journal, 1(3), 199–225

Marslen-Wilson, W. D., & Tyler, L. K. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society B, 295*(1077), 317-332.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. Cognition, 42(1), 181-211.

Nooteboom, S. G. (2005). Listening to oneself: Monitoring speech production. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 167–186). Hove: Psychology Press.

Notebaert, W., & Verguts, T. (2008). Cognitive control acts locally. *Cognition*, *106*(2), 1071-1080.

Nozari, N. (under review). The dual origin of semantic errors in access deficit: activation vs. inhibition deficit.

Nozari, N., Arnold, J. E., & Thompson-Schill, S. L. (2014). The effects of anodal stimulation of the left prefrontal cortex on sentence production. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *7*(6), 784-792.

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a "lexical editor" be?. Journal of Memory and language, 60(2), 291-307.

Nozari, N., & Dell, G. S. (2012). Feature migration in time: Reflection of selective attention on speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1084.

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

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, cognition and neuroscience*, *31*(7), 886-903.

Nozari, N. & Hepner, C. (in press). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. Cognitive Neuropsychology.

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403-410.
Nozari, N. & Omaki, A. (2018). Syntactic production is not independent of inhibitory control: Evidence from agreement attraction errors. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), Proceedings of the 40th Annual Conference of the Cognitive Science Society (pp. 824-829). Austin, TX: Cognitive Science Society.
Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt?. *Neuropsychologia*, *51*(13), 2770-2780.
Nozari, N., & Thompson-Schill, S.L. (2015). Left ventrolateral prefrontal cortex in processing of words and sentences. In G. Hickok, & S.L. Small (Eds.), *Neurobiology of Language* (pp. 569–584). San Diego, CA: Academic Press.
Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227-252.
Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and language*, *39*(4), 633-651.
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.
Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), Laboratory phonology (pp. 101139). Berlin: Mouton de Gruyter.
Pinet, S. & Nozari, N. (2018). "Twisting fingers": the case for interactivity in typed language production. Psychonomic Bulletin & Review, 25, 1449–1457.
Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it?. *Cognition*, *95*(2), 201-236.
Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*(2), 97-132.
Postma, A., Kolk, H., & Povel, D.-J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, *33*(1), 19–29.
Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review*, *107*(3), 460.
Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior*, *22*(3), 358-374.
Riès, S., Janssen, N., Dufau, S., Alario, F. X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, *23*(6), 1419-1436.
Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142.
Schlenck, K. J., Huber, W., & Willmes, K. (1987). ''Prepairs'' and repairs: Different monitoring functions in aphasic language production. Brain and Language, 30(2), 226–244

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

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, *29*(1), 86-102.

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & cognition*, *41*(8), 1200-1211.

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology*, *65*(10), 1927-1944.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

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

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*, 166.

Trude, A., & Nozari, N. (2017). Inhibitory Control Supports Referential Context Use in Language Production and Comprehension. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society Society (pp. 1218-1223).* Austin, TX: Cognitive Science Society.

Van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & behavior*, *77*(4-5), 477-482.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. Psychonomic Bulletin & Review, 20, 766-772.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. Science, 167(3917), 392-393.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1.

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