The dual origin of semantic errors in access deficit:

Activation vs. inhibition deficit

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

Abstract

According to most models of language production, to name a picture one must first map semantic features onto lexical items. Even if both sets of representations are intact, problems in mapping semantic to lexical representations can impair production. Individuals with this problem, sometimes referred to as "access deficit", often demonstrate evidence of preserved semantic knowledge (e.g., good comprehension), increased rate of lexical (usually semantic) errors in production, and inconsistent accuracy on naming the same picture on different occasions. In this paper, I argue that access deficit can have two distinct etiologies. I will present a case of double dissociation between two individuals with chronic post-stroke aphasia, one of whom shows a profile compatible with impaired activation of the target lexical item from semantic features (*activation deficit*), while the other shows a profile compatible with impaired inhibition of competing lexical items (*inhibition deficit*). These results have three key implications: (a) they provide support for the theoretical separation between activation and selection processes in computational models of word production, (b) they point to the critical role of inhibitory control in lexical selection, and (c) they invite a closer inspection of the origin of semantic errors in individuals with access deficit in order to choose the best treatment option.

Keywords: lexical selection, access deficit, inhibitory control, aphasia

Introduction

According to most models of word production, once the semantic features of an object are activated, the first step of producing a word like CAT is mapping those features (e.g., 4 legs, pet, furry, meows, etc.) onto the lexical representation of cat (Dell, 1986; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000). This step itself is considered to entail two phases: *activation* and *selection*. During the activation phase, semantic features activate not only the target word (cat) but also other words that share semantic features with cat (e.g., cow, dog, etc.; Figure 1). During the selection phase, only one of the several activated lexical representations will be selected for further processing. While activation and selection are theoretically separable operations, it is unclear how separable they are in an interactive system such as language production (see Dell, Nozari, & Oppenheim, 2014 for a review of production models). In this paper, I present neuropsychological data from two individuals with a similar profile of lexical retrieval deficit, one of whom shows a selective deficit in activating lexical items, and the other a selective deficit in inhibiting competitors during selection.
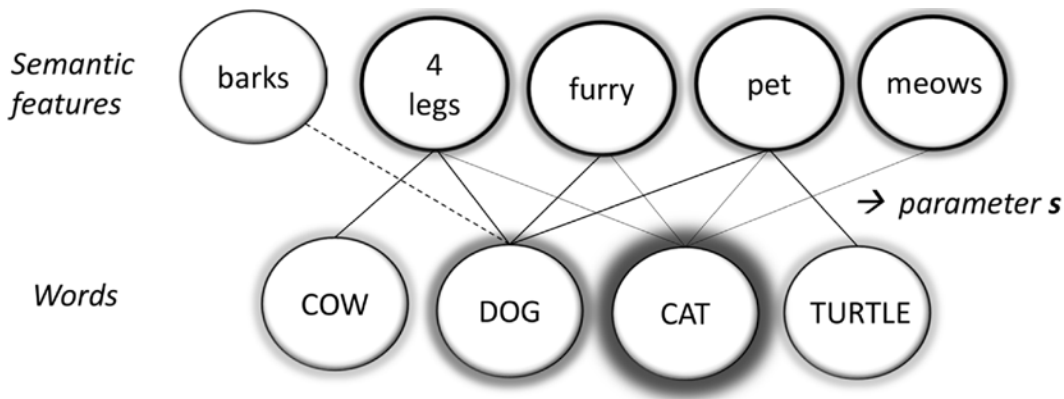

Figure 1. A schematic of the mapping of semantic features to lexical representations (words) for the target CAT. Darker circles represent higher activation, which is proportional to the number of features which send activation to the lexical representation. The dotted line represents an inactive connection. Parameter s (semantics) marks the strength of the connections between semantic features and lexical items.

Lexical retrieval deficit in aphasia

Aphasia, or the impairment of language processing as the result of brain damage, can manifest as a number of distinct syndromes. For the purpose of this study, I will focus on a specific type of aphasia marked by the predominance of semantic errors during picture naming. Two deficits have been linked to this profile in the literature. The first deficit is the loss of semantic concepts as in *semantic dementia*, which is marked by a pervasive semantic deficit across both comprehension and production tasks, regardless of specific task demands (e.g., Hodges & Patterson, 2007; Jefferies, & Lambon Ralph, 2006; Rogers, Patterson, Jefferies, et al., 2015). For example, once the individual has lost the concept of CAT, they would be unable to recognize or name cat. Loss of semantic knowledge can be probed using tasks that tap semantic associations without requiring verbal knowledge. For example, the individual may be asked to determine which of the two pictures (a cat or a dog) would go better with the picture of a bone. Loss of labels can be probed with word-to-picture matching tasks which probe lexical knowledge, without requiring labels to be produced by the individual. A typical test includes presenting an auditory word (e.g., "cat") together with a picture of the target (cat) and some distractor pictures.

Individuals who perform well on both semantic and lexical knowledge tasks but still have trouble producing the correct labels present the second type of deficit, sometimes called *access deficit* (Warrington, & McCarthy, 1983) or *semantic aphasia* (e.g., Corbett, Jefferies, Ehsan, et al., 2009; Jefferies, & Lambon Ralph, 2006; Noonan, Jefferies)*,* which manifests as an increased rate of semantic errors, especially under conditions of high competition, without the loss of semantic concepts. Unlike individuals with semantic dementia, those with semantic aphasia

show more variable performance and less sensitivity to the frequency or familiarity of stimuli, as well as a marked deterioration in performance when competitors are highly activated, such as when words have multiple meanings or a picture name is miscued (e.g., presenting a picture of a cat along with the onset /d/ for the competitor dog; see Lambon Ralph, Jefferies, Patterson, et al., 2017 for a review, and Hoffman, McClelland, & Lambon Ralph, 2018 for a computational model).

In an excellent review of the empirical evidence and theoretical accounts proposed for access deficits, Mirman and Britt (2014) showed that no single theoretical account was able to explain all the empirical findings surrounding access deficits. These authors rightfully concluded that the problem was, at least in part, due to the vague definition of access deficit. In this paper, I focus on individuals who show a selective increase in the rate of semantically-related errors in picture naming, intact semantic and label knowledge, intact auditory word repetition (i.e., preserved lexical-phonological mapping; Nozari, Kittredge, Dell, et al., 2010; Nozari & Dell, 2013), and no evidence of apraxia or articulatory-motor problems. I thus define "access deficit" literally, as a deficit of accessing (intact) lexical representations from (intact) semantic features (see Figure 1), for any reason. Using a psycholinguistic framework, I will focus the investigation on disentangling the two processes involved in mapping semantic to lexical representations, and will argue that access deficit, as described above, can itself take two forms: a deficit in activating lexical items from semantic features (*activation deficit*), or a deficit in inhibiting competitors (*inhibition deficit*).

Activating lexical items from semantic features is an essential part of the process of lexical retrieval. The importance of this process is highlighted in a class of computational models of aphasia which have been particularly successful in explaining the error patterns in the affected

individuals (e.g., Dell, Schwartz, Martin et al., 1997a; Schwartz, Dell, Martin, et al., 2006). In such models, the strength of the connections that carry out this mapping is represented by one of the critical parameters in the model, the parameter *s* (semantic; Figure 1). The lower the *s* parameter, the weaker the transmission of information from the semantic to the lexical layer and its convergence on a single target, and thus the larger the number of semantic (cat → dog), mixed (cat → rat), and, in extreme forms, unrelated lexical (cat → bed) errors (Schwartz et al., 2006). It is thus compatible with the framework of these models that generally low levels of activation (i.e., an *activation deficit*) could be a primary cause of semantic errors.

At the same time, a lower *s* parameter also leads to smaller differences in the activation levels of lexical items (Nozari, Dell, & Schwartz, 2011). If the general level of activation for all items is very low (i.e., so low that an absolute selection threshold cannot be reached), the problem will be indistinguishable from *activation deficit* discussed above. It is, however, possible that despite the close activation levels of multiple lexical items, the activation is high enough for individual items to pass the absolute selection threshold and potentially be selected. In such a case, different theories of selection make different predictions: Competitive accounts of selection (e.g., Roelofs, 1992) propose that the ensuing competition should delay production until competition is clearly resolved in favor of one representation. Resolving such competition may depend on inhibitory control. Non-competitive selection accounts (e.g., Mahon, Costa, Peterson, et al., 2007; Navarrete, Del Prato, Peressotti, & Mahon, 2012; 2014), on the other hand, do not view close levels of activation among competitors as an obstacle to selection; the first word that reaches an absolute threshold is selected. It is thus an empirical question whether inhibition of competing responses is a critical part of lexical selection, and whether a failure of applying such control (i.e., an *inhibition deficit*) could also cause semantic errors, although

reports of poor performance by individuals with semantic aphasia in the presence of strong competitors strongly suggests this possibility (e.g., Corbett et al., 2009; Jefferies, & Lambon Ralph, 2006; Noonan et al., 2010).

The current study

The current study compares the performance of two individuals with a similar profile of good semantic comprehension and a predominance of semantically-related errors in picture naming. Despite this similarity, however, the naming pattern differed between the two: one individual showed long response latencies and mostly single responses on the majority of trials, raising the possibility that she had difficulty activating any lexical representations. The other individual produced several responses in quick succession, potentially hinting at a problem in inhibiting the already-activated competitors. This paper tests three predictions to dissociate these two deficits.

Prediction 1. The individual with *activation deficit* should have difficulty in maintaining the activation of lexical items and their connection to their semantic features, while the individual with *inhibition deficit* should not demonstrate such a problem.

Prediction 2. The individual with *inhibition deficit* should be selectively impaired in conditions that increase the activation of the competitor. Specifically, he should be prone to the selection of the competitor as the response in these conditions. Since the individual with *activation deficit* is still capable of suppressing the competitor, her performance should not suffer under such circumstances. If she fails to produce the correct word, she should be more likely to produce no response (an omission) than to produce the competitor.

Prediction 3. Related to both of the previous predictions, opposite effects of increased semantic similarity are expected in the two individuals. The individual with *activation deficit* should benefit from the overall increase in the activation of items in the lexical semantic space provided by semantic similarity[1]. The performance of the individual with *inhibition deficit*, on the other hand, should suffer from increased activation of semantic competitors.

Lexical perseverations. In addition to testing the three predictions described above, I will examine lexical perseverations in the two participants. A perseveration is the inappropriate intrusion of a previously produced response on the current trial, and is a common feature of immature or damaged production systems (Ackerman & Ellis, 2007; Fischer-Baum, McCloskey, & Rapp, 2010; Dell, Burger, & Svec, 1997b; Martin & Dell, 2007; McCloskey, Macaruso, & Rapp, 2006; Moses, Sheard, & Nickels, 2007). Two general mechanisms have been proposed for perseveratory errors: the *failure-to-activate account* (Cohen & Dehaene, 1998; Dell et al., 1997b) posits that when the current target fails to gain enough activation, the residual activation of a previously produced response may be enough for its selection in lieu of the current target. The *failure-to-inhibit account* hinges on the malfunction of a hypothesized inhibitory process that suppresses the activation of an item once it has been produced (e.g. Houghton, Glasspool, & Shallice; 1994; MacKay, 1986).

One study has directly assessed the contribution of these two types of deficit to perseverations, albeit for letter (as opposed to lexical) perseverations. In a sample of 12 individuals with dysgraphia, Fischer-Baum and Rapp (2012) found clear support for the failure-to-activate account, with a likely contribution of a failure-to-inhibit deficit in one participant. For

---

[1] Note that while semantic similarity increases the activation of the competitors, in systems with feedback from lexical to semantic representations, the activation has the greatest convergence on the target.

example, all participants in their sample who had both lexical and sublexical impairment—which prevents the activation of the target through either route, resulting in an *activation deficit*— showed above-chance rates of perseveration errors (see Fischer-Baum & Rapp, 2012 for additional details on the relationship between error types and route impairment). But an *activation deficit* alone does not predict perseverations on a task such as direct copy transcoding, in which participants simply copy a written word that is in front of them, since in this task graphemes can be activated via other routes that bypass the damaged lexical and sublexical routes. The authors found one individual who perseverated at above chance rates on this task, pointing to a potential *inhibition deficit.*

The study sample, however, was selected specifically to contain individuals with a definite activation deficit (i.e., damage to both lexical and sublexical routes). It thus remains an open question whether the failure-to-inhibit deficit alone can result in perseverations, and whether both mechanisms can also lead to lexical (as opposed to letter) perseverations. Since the failure-to-activate and failure-to-inhibit mechanisms closely mirror what is described in this paper as *activation deficit* and *inhibition deficit*, these questions can be answered by studying the pattern of lexical perseverations in the two participants.

Methods

Participants

XR is a 56 year old female with a high school education. She was premorbidly left-handed and had a right hemisphere stroke 22 years before the study. QD is a 65 year old male, also with a high school education. He was right-handed before he had a left-hemisphere stroke 12 years prior to the study. Appendix A contains multi-view images from structural MRI scans

that show the extent of damage to the left and right hemisphere in QD and XR, respectively. QD had pervasive damage to the frontal and parietal lobes with some extension to the superior temporal gyrus. In comparison, XR had relatively preserved anterior and ventrolateral (but significantly damaged dorsolateral) prefrontal cortex. Parietal damage was present but was less severe than QD. Her lesion also showed some extension to the superior temporal gyrus together with severe undercutting of the white matter connecting frontal and temporal lobes. Both participants presented with word-finding difficulty and production of short nonfluent utterances that were often missing function words. Both had been in group therapy for enhancing speech and life skills for over five years at the Snyder Center for Aphasia Life Enhancement (SCALE; https://www.leagueforpeople.org/scale) from where they were recruited.

In addition, 12 neurotypical native speakers of English (six females; $M_{age} = 61.67$, $SD = 7.01$ years) participated as controls. All participants were consented under an IRB protocol approved by Johns Hopkins University and received monetary compensation for their participation.

Background language tests. The two participants were chosen because of the similarity in their language profiles. Table 1 summarizes the results of their background language tests. Both had near-intact semantic and auditory-lexical comprehension. The first task probed semantic knowledge (without words): they had to pick whichever one of the three pictures best matched a target picture. Half of the trials probed taxonomic (e.g., banana-pineapple) and the other half thematic relationships (e.g., hair-comb). In both conditions, the foils were neither thematically nor taxonomically related to the target. Out of 28 trials, 39 control subjects had a mean score of 27.74 ($SD = 0.88$), and both participants scored 27. The next task was a word-to-picture matching task, probing auditory word comprehension. Participants matched each of the 175

labels for pictures in the Philadelphia Naming Task (PNT; Roach, Schwartz, Martin et al., 1996) to the target picture in the presence of two distractors that were unrelated to the target. Both participants' scores were near perfect on this task. In addition to good comprehension, they also showed high accuracy on auditory repetition of PNT words, showing intact lexical-to-phonological mapping (e.g., Nozari et al., 2010; Nozari & Dell, 2013).

Both, however, were visibly impaired in picture naming on the PNT, with QD showing an overall lower accuracy than XR (Table 1). Importantly, both had a predominantly semantic error profile. Despite this similarity, however, analysis of response latencies for the first 100 correctly-named PNT trials showed that XR was significantly slower than QD (see Table 1, PNT RTs; Mann Whitney U test's $Z = 4.32$, $p < .001$).

Large-set picture naming. To confirm the dominance of semantic errors, and to show the variability in performance typical of access deficits, we tested XR and QD on naming 444 pictures (colored photographs from public image repositories) administered on two separate occasions (a total of 888 trials/participant). These lists were designed such that items from the same semantic category (e.g., vehicles, fruits, clothing items, birds, insects, etc.) were at least 12 items apart to reduce the chance of semantic blocking (Schnur, 2014). Participants were given 20 seconds to respond unless they indicated that they were done with the trial. No feedback was provided on the responses, except encouraging them to produce a single word in cases where they were producing descriptions of items, or providing general encouragement when they expressed frustration regarding their performance, which is not unusual in this population. Because of the much greater diversity in the pictures in terms of lexical frequency, age of acquisition, name agreement, and length, the overall accuracy was lower compared to the PNT, but, as seen in Table 1, the relative accuracy, as well as the dominance of semantically-related

errors, was preserved in the two participants. Both participants showed some awareness of their errors by expressing dissatisfaction with some of their incorrect responses, although this awareness was only partial, as some errors remained undetected. Item-level consistency (whether the same item was named accurately or inaccurately on both naming occasions) was measured by Cronbach's alpha, which was 0.69 for XR and 0.66 for QD, showing that neither participant demonstrated highly consistent performance at the item level. Despite these similarities, one difference was striking in the pattern of responses when a semantically-related error was made: QD produced multiple semantically-related responses (e.g., "orange", "peach", "no, apple", for the target "watermelon") on many more trials than XR (57 vs. 28; Fisher's exact test; $p = .006$).

In summary, both participants had preserved comprehension and auditory word repetition abilities and a predominantly semantic error profile in picture naming. However, XR had a tendency to respond slowly with single words, while QD responded more quickly and more often with multiple semantically-related words.

Table 1. Scores of the background language tests for the two participants. The numbers in the parentheses next to the tests show the total number of items. The first five columns show the similarities in XR and QD's profiles. The last two columns show the differences. Percentages of semantically-related errors were calculated as the proportion of semantic (e.g., cat → dog) + mixed (cat → rat) errors over all commission errors. See Appendix B for the coding scheme, as well as the breakdown of error types for the 1st and 2nd encounter with items in the large naming test. Comp = Comprehension; PNT = Philadelphia Naming Test; PRT = Philadelphia Repetition Test; RT = Response times (latencies); Sem = semantic. XR has a hypothesized activation deficit, while QD has a hypothesized inhibition deficit.

	Similarities in language profiles					Differences in language profiles	
	Sem comp (28)	PNT comp (175)	PRT (175)	PNT (175)	Large naming set (444 x 2)	PNT RTs (ms) (100 correct)	Multiple-response trials (out of sem-related errors in the large naming set)
XR	27	173	172	154 correct; 79% sem-related	63% correct; 85% sem-related	2064 (SE=142)	28 (14%) out of 198
QD	27	174	171	134 correct; 82% sem-related	50% correct; 84% sem-related	1658 (SE =128)	57 (28%) out of 201

Experimental tasks testing the three predictions

Four tasks—a modified version of the Category Probe task (Freedman & Martin, 2001), the Simon task (Hommel, Proctor, & Vu, 2004), the Miscueing task (Noonan et al., 2010) and the Word-pair Stroop task (Nozari, Freund, Breining et al., 2016)—were used to test the predictions of the *activation* vs. *inhibition* deficit accounts. The relevance of Simon, Miscueing, and Stroop tasks for testing inhibitory control abilities is clear from the literature. The original Category Probe task, however, has been used as a test of working memory, and it may not be immediately clear why it is appropriate for testing the activation deficit hypothesis. The logic is as follows: Successful performance in this task requires that participants keep the list items and their connections to their semantic categories activated in working memory. Importantly, it is not critical whether the exact item is remembered or not; what is critical is that the activation of a given item reinforces the relevant semantic category, and vice versa. In Baddeley and Hitch's

(1974) terms, this requires "verbal rehearsal", i.e., continuous mapping of semantic features to lexical items in inner speech. This mapping is expected to be deficient in *activation deficit*, but preserved in *inhibition deficit*.

Theoretically speaking, it is also possible to complete the Category Probe task using an alternative strategy: upon hearing each item, participants could convert that item to a label for a semantic category and then rehearse that label. This is unlikely for two reasons: (a) because the items are presented in 1 s intervals which leaves too little time for inferential processing, especially in individuals with brain damage, and (b) there are many possible categories in the experiments, and participants are not told in advance what they are. Thus, upon hearing an item such as "pigeon", it may not be immediately obvious which category label is most relevant here, birds or animals? But when the probe "sparrow" is heard, it becomes obvious that two birds have been mentioned. It is thus much less likely that what is being memorized is inferred category labels detached from the lexical items. This, in turn, strengthens the assumption that the task relies on keeping the semantic-lexical mapping loop activated.

For two reasons, this task also does not require much inhibitory control. First, remembering the specific lexical item does not matter for generating the correct response. For example, if the item "cat" is presented, it activates the general category of animals, with higher activation of features that correspond to the categories of mammals and pets. If these features then also activate "dog", the competition between "dog" and "cat" is irrelevant to the performance in this task, as both items still contribute to the activation of the same relevant semantic categories. Second, opportunities for interference have been intentionally minimized by avoiding the repetition of any items across trials in the adapted version used in this experiment. Although the original Category Probe task did not show a reliable correlation with executive

control tasks (Allen, R. Martin & N. Martin, 2012), other studies have found parallel impairment in performance on inhibitory control and short term memory tasks (e.g., Hamilton & Martin, 2005). The source of impaired performance in such memory tasks is often interference from a previous trial, i.e., an item from trial t-1 that is relevant to the answer on trial t, could be mistakenly remembered as having been presented on trial t. Such interference is induced by repeating combinations of items from a small set (e.g., triplets of letter from a total set of 16 letters), which keeps each item highly activated and thus requires good inhibitory control for teasing apart relevant and irrelevant presentations of the same item. By modifying the Category Probe task not to include any item repetition, the problem of interference is significantly reduced. Therefore, *activation deficit*, but not *inhibition deficit*, should lead to significantly lower than normal scores on the Modified Category Probe task (prediction 1).

Since the details of the tasks are necessary for interpreting the results, I will present each task before discussing the results from that task in the Results section.

Evaluating lexical perseveration

Lexical perseverations were examined in the large-set naming task. First, all lexical substitutions (semantic errors, mixed errors, and unrelated lexical errors) were identified for each participant. Next, the number of lexical perseveration errors (i.e., a repetition of a response that had been produced previously in the same session) was counted. Note, however, that a lexical perseveration may arise due to the influence of a prior production (which I call true perseveration), or simply by chance, when the same response happens to be produced on multiple trials independently of one another. If the latter is the case, then the order of the responses should be arbitrary. In other words, the responses prior to a given lexical substitution

should represent a random sample drawn without replacement from the set of responses produced in the session. This probability follows a hypergeometric distribution X with a probability mass function given by $p(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$; if x is the response produced on the current trial, K is the total number of x responses produced in the session excluding the current trial, k is the number of x responses on previous trials, N is the total number of responses produced in the session excluding the current trial, n is the number of previous trials, and $\binom{a}{b}$ is the binomial coefficient, calculated as $\frac{a!}{b!(a-b)!}$. Thus, the probability that preceding responses contain at least one instance of x by chance, i.e., $p(X > 0)$, can be calculated as $1 - p(X = 0)$. Here is an example of how the probability of observing a perseveration by chance is calculated for a given trial, trial 59, with the response "desk" for the target "table", when "desk" has also been produced on two other occasions, once for the target "desk" and once for the target "dresser", in a session containing a list of 123 items. The probability of at least one of these two other responses appearing before trial 59 by chance is calculated as $1 - \frac{\binom{2}{0}\binom{120}{58}}{\binom{122}{58}} = 0.73.$

Note that the above probability depends critically on the values of K and n. The greater the number of times a particular response has been produced in the session, the higher the probability that it would be observed as a lexical perseveration by chance. For instance, if the response set used in the above example contained three—instead of two—instances of "desk" aside from the current trial, the chance of observing a lexical perseveration by chance would have increased from 0.73 to 0.86. Similarly, the greater the number of the preceding trials, the higher the chance that one of them would be randomly repeated and observed as a lexical

perseveration. For instance, the chance of observing a lexical perseveration by chance increases from 0.73 on trial 59 to 0.96 on trial 99.

The probability of perseveration due to chance can thus be calculated, as described above, for each lexical substitution in each session. The result is a vector of probabilities with a Poisson binomial distribution, which is similar to a binomial distribution but with a different chance of success in each trial. Using this distribution, one can then determine whether the observed number of lexical perseverations (i.e., lexical substitutions with at least one identical response preceding them) is statistically different from chance.

Procedures

All but two control participants completed all the tasks in the same order. These two control participants were tested on an older version of the Category Probe task which contained words (e.g., flower names) unfamiliar to the average male participant in the Baltimore area, and also many repetitions of the same word, potentially creating retroactive interference. The task was amended to avoid these problems for the next 10 control participants, as well as XR and QD. Control participants were tested in two sessions. XR and QD were tested in a total of seven and eight 1–1.5 h sessions, respectively, including the large-set naming test (but excluding background language tests which had been collected previously). The difference in the number of sessions was due to the breakdown of the large naming set: QD requested more breaks, hence the additional session. Each picture in the large naming set was presented for 20 seconds or until the participant produced a response. Once all items had been administered once, the whole set was administered a second time in a different order.

The data from two of the four administrations of the Word-pair Stroop were collected along with the naming tests. XR and QD and all control participants completed the rest of the experimental tests over two sessions. The first session included Miscueing (part I), Modified Category Probe, and the third administration of the Word-pair Stroop task. The second session included the fourth administration of the Word-pair Stroop task, followed by the Simon task, and Miscueing (part II). Simon and Word-pair Stroop tasks were presented in the E-Prime 2.0 software (Psychology Software Tools, Pittsburg, PA). Miscueing was presented in PowerPoint. The Modified Category Probe task was administered by a research assistant who was a native speaker of English and was supervised by the author. Button-press responses were collected using two buttons on the keyboard (all participants used the index and middle fingers of their non-dominant hands). Verbal responses were recorded digitally for offline transcription.

Results

Verbal responses were transcribed and coded by a native English speaker blind to the hypotheses of the study and were double-checked by the author. Details of the coding system are available in Appendix B. The first response was used in all codings, even if the participant later changed it. RTs were hand-coded based on visual inspection of the acoustic wave in Praat, version 6.0.43 (Boersma & Weenink, 2018, http://www.praat.org/). Where RT analyses are reported, error trials are excluded from the analyses. Unless stated otherwise, statistical analyses were performed using techniques advocated by Crawford, Garthwaite, and Porter (2010) which correct for small control samples. Table 2 summarizes the results of the experimental tests.

Modified Category Probe task (testing prediction 1). This task was loosely adapted from Freedman and Martin (2001), but with new materials (in order to avoid using words that may be

unfamiliar to individuals of a specific gender, e.g., low-frequency flower names), no repetition, and fully counterbalanced lists such that the answer appeared equally often in every position in each list. The examiner read a list of words (1/second) followed by a probe word. The list grew from 1 to 6 words and participants indicated whether the probe (e.g., "monkey") belonged to the same semantic category as any of the words in the list (e.g., "train, cherry, zebra, shoe, bed" for a 5-item list; correct answer = "yes"). Accuracy of 75% or higher was required to move on to the next level, i.e., a longer list. Scores are reported as the number of items in the longest list in which the criterion was met, followed by the proportion of correct items in the next list as a decimal. For example, a score of 4.6 indicates that the participant completed more than 75% of trials correctly in a 4-item list and correctly responded to 60% of items in the 5-item list. For reasons explained in the earlier sections, *activation deficit*, but not *inhibition deficit*, is expected to lead to significantly lower than normal scores on the Modified Category Probe task (prediction 1).

Results of the Modified Category Probe task. Control participants' performance was above chance on lists containing at least 4 items and above (range = 4.6–6). XR was only able to pass the criterion (75% accurate) on lists with up to 2 items. QD, on the other hand, successfully passed the criterion for lists of up to 4 items, and correctly responded to 60% of the items in the 5-item list. QD's score was not significantly different from that of the control group, but XR's was significantly below the group mean, compatible with an *activation deficit*.

Simon task (testing prediction 2). This was a classic Simon task (Hommel, Proctor, & Vu, 2004) which required participants to press the button (red or blue) that matched the color of a square on the screen. On the congruent trials ($N = 40$), the square appeared on the same side of the screen as the button. On the incongruent trials ($N = 40$), it appeared on the opposite side. The response

deadline was 1 s, with a fixed inter-trial interval of 1550 ms. Markedly impaired performance in the incongruent condition—which requires suppression of the same-side response—is a hallmark of a general disorder of inhibitory control, and could thus be expected in *inhibition deficit,* if such a deficit encompasses language-specific inhibitory control. *Activation deficit*, on the other hand, should not cause impaired performance on the Simon task. **Results of the Simon task.** XR and QD showed a Simon effect (Incongruent − Congruent) in error rates comparable to each other and not significantly different from control participants (Table 2). Figure 2 shows the RTs. The Simon effect in RTs was significantly larger than controls for QD, but not for XR (Table 2). The performances of the two participants were compared directly against each other in a multi-level regression model with the random effect of trials, in which "Subject" was treated as a fixed effect, along with Condition (Congruent vs. Incongruent) and the interaction between the two. In addition to a main effect of congruency ($t = 10.4$, $p < .001$), the Subject by Condition interaction term was statistically significant ($t = -4.61$, $p < .001$). Post-hoc tests showed that XR was significantly slower than QD in the Congruent ($t = 3.17$, $p = .002$), but significantly faster than him in the Incongruent condition ($t = −3.53$, $p < .001$). These results show QD's selective impairment in the Incongruent condition, compatible with an *inhibition deficit*.
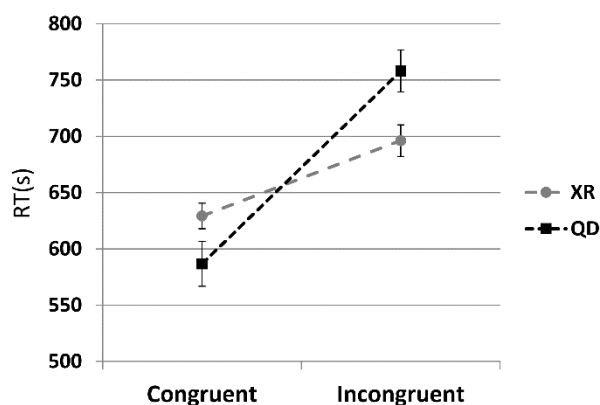

Figure 2. RTs (±SE) for XR and QD in Congruent and Incongruent conditions in the Simon task

Table 2. Scores on the experimental tests for XR, QD, and control participants. Scores in each condition are reported along with the critical score which is most relevant to the prediction tested by each task. Reported *p* values, as well as point estimate and 95% CI for effect sizes, reflect differences between XR or QD's score and the control critical scores, and have been calculated using Crawford et al.'s (2010) proposed statistical methods. Results of direct comparisons between XR and QD, when appropriate, are reported in the text. C = Congruent, I = Incongruent. Asterisk (*) indicates a difference from controls' score that is statistically significant at α = .05.
[a] Percentages—instead of error counts—are reported for Word-pair Stroop, because of the different number of trials completed by different participants (see text).
[b] No statistics are reported for comparisons between XR and QD and control participants because the *SD* of the control sample was 0 (no errors).

	Score per condition		Critical score	p value	Effect size (95% CI)
Modified Category Probe					
XR	NA		2.5	< .001*	-5.60(-8.20 to -2.98)
QD	NA		4.6	.10	-1.91(-2.96 to -0.83)
Controls mean ±*SD*	NA		5.69±0.57		
Simon (Error)	**C**	**I**	**Simon Effect**		
XR	0	5	5	.23	1.33 (0.52 to 2.1)
QD	1	6	5	.23	1.33 (0.52 to 2.1)
Controls mean ±*SD*	1±1.96	2.92±3.43	1.75±2.45		
Simon Effect (RT)	**C**	**I**	**Simon Effect**		
XR	630	697	67	.87	-0.17 (-0.73 to 0.41)
QD	587	758	171	.01*	3.25 (1.78 to 4.68)
Controls mean ±*SD*	488±67	560±70	72±31		
Miscueing (Error)	**Cue**	**Miscue**	**% Miscue error**		
XR	4	8	25% of errors; 10% of miscue trials	NA[b]	NA
QD	2	13	100% of errors; 65% of miscue trials	NA[b]	NA
Controls mean ±*SD*	0±0	0±0	0±0		
Word Pair Stroop (Error)[a]: Unrelated	**Straight naming**	**Reversed naming**	**Word-pair Stroop effect**		
XR	3%	26%	23%	< .001*	5.51 (3.17 to 7.84)
QD	5%	35%	30%	< .001*	7.41 (4.30 to 10.50)
Controls mean ±*SD*	1%±1%	4%±4%	3%±4%		
Word Pair Stroop (Error)[a]: Semantic	**Straight naming**	**Reversed naming**	**Word-pair Stroop effect**		
XR	2%	8%	6%	.17	1.52 (0.66 to 2.34)
QD	2%	55%	53%	< .001*	16.68 (9.80 to 23.55)
Controls mean ±*SD*	0±0	1%±3%	1%±3%		

Miscueing task (also testing prediction 2). Adapted from Noonan et al. (2010), this task comprised the presentation of 20 pictures (each for 10 seconds or until a response was produced), once with a *cue* (the correct onset, presented both visually and auditorily) and once with a *miscue* (the onset of a strong competitor, e.g., a picture of a tiger presented with "L" for "lion"). The test was administered in two sessions, each with 10 cue and 10 miscue trials, such that no picture was repeated within the session. Cues should help in both *activation deficit* and *inhibition deficit*. However, the individual with *inhibition deficit* should be much more susceptible to producing the competitor in the miscue condition, while the individual with *activation deficit* should show more omissions than the production of the competitor (because she can inhibit the competing response even if it is activated by the miscue).

Results of the Miscueing task. Average baseline naming accuracy for the items in the Miscueing task, assessed by the large-set naming task, was 70% and 68% for XR and QD respectively. As expected, good cues helped: in the cue condition their accuracy increased to 80% and 90%. In the Miscue condition, XR made 8 errors (60% accuracy) and QD 13 errors (35% accuracy). The non-parametric (Fischer's exact test) was used to analyze these data. The overall number of errors in the cue vs. miscue conditions did not differ between the two participants ($p = 0.36$). However, the pattern of errors in the Miscue condition was different (Figure 3). Of XR's errors, only 25% were miscue errors (i.e., errors where she named the cued competitor). The rest were omissions, with the exception of one semantically-related error that was unrelated to the cue. In contrast, all of QD's errors in the miscue condition were the miscued competitor, in keeping with an *inhibition deficit*. Fischer's exact test showed that the distribution of error types was significantly different between XR and QD ($p < .001$).
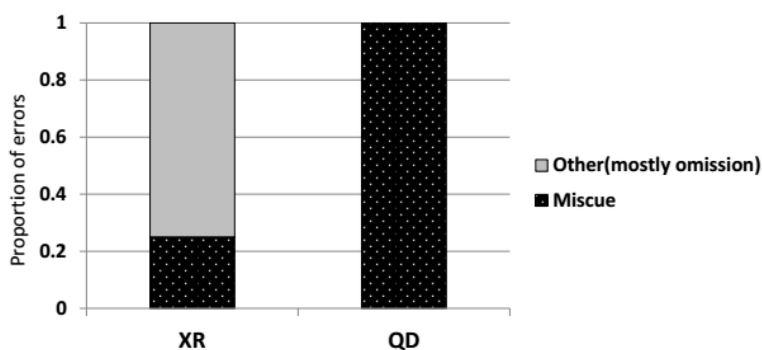
Figure 3. Proportion of miscue errors vs. other error types in all error trials in the miscue condition in XR and QD. Total number of errors in this condition is 8 for XR and13 for QD.

Word-pair Stroop (testing prediction 3). Adapted from Nozari et al. (2016), this task presents

blocks containing two pictures. In the first block (16 trials) participants view one of the two

pictures on each trial in pseudo-random order (no more than 3 repetitions in a row) and name it

with a deadline of 3 s and an inter-trial interval of 2 s. This creates a situation similar to blocked

cyclic naming but with only two items. In the second block (16 trials), they learn to reverse the

names of the two pictures, e.g., say "pig" when they see a picture of a car, and say "car" when

they see a picture of a pig. Since viewing the picture of a car naturally elicits the response "car",

the reversed blocks create a Stroop-like situation, where the natural prepotent response must be

suppressed in order for the instructed response (i.e., the name of the other object in the block) to

be produced.

Two sets of materials were used over different sessions. Since performance on naming

tasks is known to be variable in individuals with access deficits, in order to obtain more stable

estimates, XR and QD completed two sessions with each set for a total of four sessions. On the

other hand, performance is much more stable in neurotypical controls, so they only completed

two sessions, each with a new set of materials. Each session included two conditions: unrelated

(cat/pen, pig/car) and semantically-related (cat/dog, pig/fox), administered in counterbalanced

23

order. Each condition contained the two straight and reversed blocks as described above. Each block was preceded by 4 practice trials, which were repeated if necessary until it was clear that the participant was comfortable with the picture names and was following the task instructions.

Both participants are expected to show poorer performance in the reversed than the straight condition for different reasons: the removal of the to-be-named picture makes activating the correct lexical item more difficult in *activation deficit*; likewise, the replacement of the target picture with the competitor picture overactivates the competitor, which is hard to inhibit in *inhibition deficit.* However, the two deficits make the opposite predictions regarding the effect of similarity. The individual with *activation deficit* may benefit from the extra activation in the lexical semantic space provided by the semantic competitor in the block. This is because in such a deficit lexical items do not have enough activation to reach a minimum threshold for selection. Additional activation provided by a semantically-related item could give the target a better chance of passing that threshold. There is always a chance that the competitor's activation reaches the threshold first by mistake, but in such cases the intact inhibitory control can prevent it from slipping out. Thus the net effect of semantic similarity in *activation deficit* can be expected to be facilitation. On the other hand, the performance of the individual with *inhibition deficit* should deteriorate as a result of increased activation of the semantic competitor[2], as the problem is not reaching a minimum threshold of activation, but keeping the competitors from being produced instead of the target. The more activated these competitors, the higher their chance of passing the threshold before the target.

---

[2] The reason for examining the effect of semantic similarity in the reversed as opposed to straight naming was that the two participants were fairly high-level, and made few errors in the straight conditions (i.e., there was a ceiling effect in accuracy). The reversed condition increased task difficulty and allowed differences to show up.

Results of the Word-pair Stroop. Error rates in the unrelated and semantically-related conditions

for all participants are reported in Table 2. Not surprisingly, both XR and QD made significantly

more errors than controls in both straight (XR: p =.006, effect size = 3.50, 95% CI [1.94, 5.03];

QD: p <.001, effect size = 4.83, 95% CI [2.76, 6.89]) and reversed (XR: p =.002; effect size =

4.23, 95% CI [2.39, 6.05]; QD: p < .001, effect size = 12.09, 95% CI [7.08, 17.08]) conditions.

However, the pattern of errors was different between the two participants (Figure 4).

While both XR and QD had larger Stroop effect sizes than controls in the unrelated

condition, only QD had a significantly larger effect size than controls in the semantic condition

(Table 2). The performance of the two participants were compared directly against each other

using Fisher's exact test, which revealed a significant difference in the Stroop effect in error rates

for semantic and unrelated conditions between XR and QD (p =.003). Finally, the dissociation in

performance within each participant in the unrelated and semantic conditions was investigated

using the Bayesian inferential methods for the difference between a case's scores on two tasks

against a control sample (Crawford et al., 2010). This dissociation was reliable both in XR (p =

.001; effect size = 5.00, 95% Credible Interval [2.41, 8.09]) and QD (p = .001; effect size =

−11.59, 95% Credible Interval [−20.27, −4.19]). Together, these results showed opposite effects

of semantic relatedness on XR's and QD's performance: XR's performance improved, while

QD's performance deteriorated as a result of semantic similarity during reversed naming.
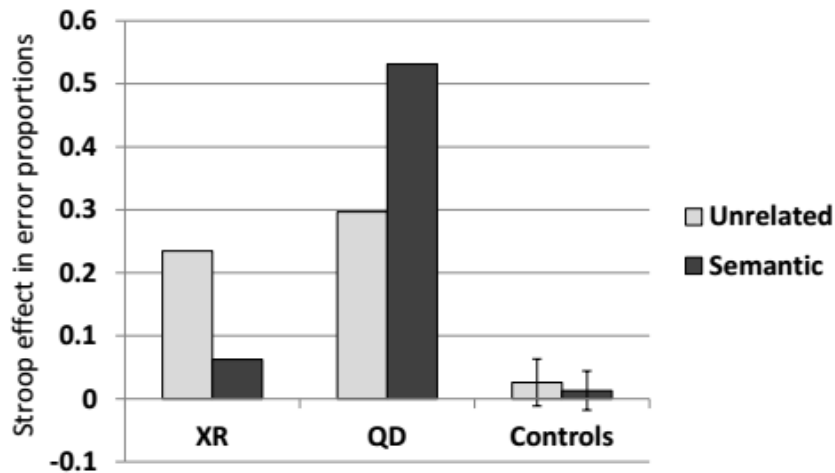
Figure 4. Stroop effects in errors (errors in reversed naming − errors in straight naming) for XR and QD and controls (±SD*) in the unrelated and semantically-related conditions in the word-Stoop task.*

Lexical perseverations. Perseveration errors were examined in the large naming set. XR and QD

completed the 888 trials in five and six sessions, respectively, and produced 212 and 209 lexical

substitutions (semantic, mixed, and unrelated lexical errors), of which 42 (20%) and 59 (29%)

were coded as lexical perseverations. Figure 5 shows the probability density of perseveration

errors over different lags for the two participants, the general shape of which complies with prior

reports of lexical perseveration: a peak at a lag > 1 followed by a monotonic decrease in the

probability of perseveration in later lags (Hsiao et al., 2006). Interestingly, instead of repeating

the most recent response(s), e.g., a peak at lag 2, the peak was close to the average distance

between two semantically-related items in the list ($M = 20.86$), showing an increased rate of

perseverations selectively for semantically-related responses. To determine whether observed

perseveration errors occurred at a rate above chance, the procedure described in the Methods

section was used. The observed perseveration rates were then compared against the Poisson

binomial distribution of chance using the R package poisbinom version 1.0.1 (Olivella &

Shiraito, 2017). This analysis returned a *p* value of *p* = .003 for XR and *p* = .064 for QD.
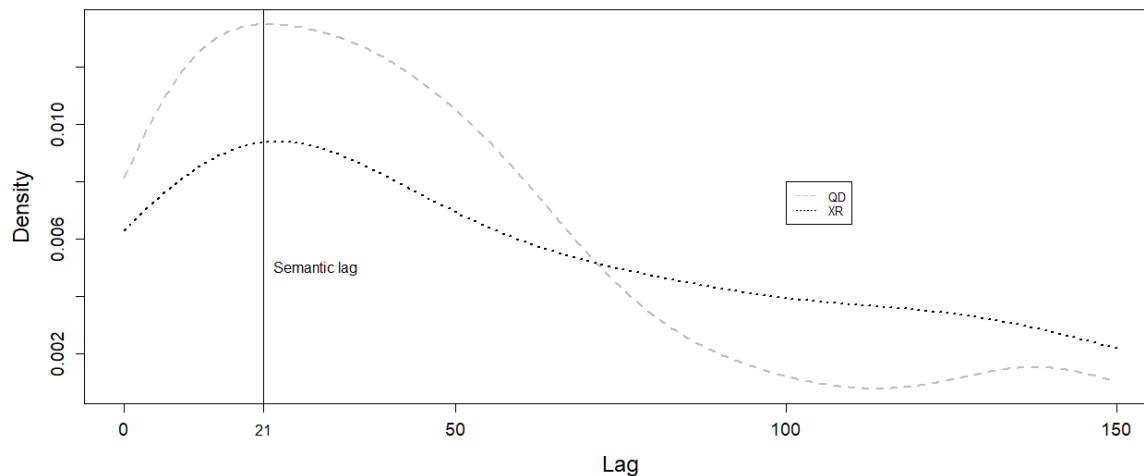

Figure 5. The probability density of lexical perseveration by lag for QD and XR. The vertical line (lag 21) represents the average lag between two semantically-related items, calculated as the distance between each item and the next item from the same semantic category, averaged over all items. This lag coincides with the peak in the probability of producing a lexical perseveration in both participants.

Discussion

The study aimed to determine whether the process of mapping semantic features to lexical items

has two dissociable components, activation and selection, which could be selectively impaired.

Although the concepts of activation and selection appear in virtually all models of language

production (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000; Roelofs,

1992), recent neural evidence has called into question the separability of these two processes

(Riès et al., 2017). To make matters more complicated, while the dynamics of the activation

process have been studied extensively (e.g., Dell et al., 1997a), the mechanisms for selection are

not well understood. Importantly, it is unclear whether the activation of competing lexical

representations slows down target selection (competitive selection; e.g., Roelofs, 1992), has no

effect on the speed of target selection (e.g., Mahon et al., 2007; Navarrete et al., 2014), or depends on a selection criterion that changes flexibly based on task goals and the state of the production system (Nozari & Hepner, 2018).

The current results suggest that the dissociation between activation and selection processes is real (see also Piai, Roelofs, Jensen, et al., 2014 for different electrophysiological signatures of activation and selection). Moreover, these processes can be selectively impaired in aphasia. Table 3 provides a summary of the findings of the study. The language profiles of XR and QD were indistinguishable on paper (see the first five columns of Table 1). However, XR (*activation deficit*) produced single-word responses in picture naming with long delays, compatible with slow spreading of activation and difficulty for any item to reach a selection threshold, while QD's (*inhibition deficit*) production was significantly faster and contained many more multiple-response trials, in keeping with quick spreading of activation pushing multiple items over the threshold without sufficient control over selection. When the task required sustained activation of lexical items and their links to their semantic categories (the Modified Category Probe task), XR's performance was significantly impaired. Since the task required no inhibitory control, however (i.e., *any* response in the same semantic category as the probe was acceptable), QD showed no significant impairment on this task. On the other hand, when the task required the selection of a single response among activated competitors, QD's performance was visibly poorer than XR and neurotypical control participants. He often failed to suppress the production of the miscued competitor, performed more poorly in the incongruent condition of the Simon task, and had great trouble reversing words in the semantically-related condition. While these results do not imply that activation and inhibition deficits cannot coexist, they suggest that the two deficits—and the activation and selection processes that are impaired in these deficits,

respectively—can be distinguished. In other words, an "access deficit" could be an *activation deficit*, an *inhibition deficit*, or a combination of the two. In the rest of the discussion, I will discuss several ramifications of these results.

Table 3. Summary of findings.	Activation deficit (XR)	Inhibition deficit (QD)
Picture naming	Longer RT, often a single semantically-related response	Shorter RT, multiple semantically-related responses
Modified Category Probe task	Impaired	Unimpaired
Simon task	Unimpaired	Impaired
Miscue task	Few miscue errors	Lots of miscue errors
Word-pair Stroop task	Helped by semantic similarity	Hurt by semantic similarity
Lexical perseveration	Significantly above chance	Marginally above chance

How separable are activation and selection processes?

The data I presented in this paper argued for a potential dissociation between activation and selection processes in word production. At first glance, this finding may seem at odds with recent electrocorticography data that have suggested a large spatiotemporal overlap of activation and selection (Riès et al., 2017). But that would only be the case if activation and selection are assumed to be strictly serial processes in a modular system. Production models that allow for cascading and interactivity between the layers of representations (e.g., Dell, 1986; Rapp & Goldrick, 2000) readily predict at least some overlap between activation and selection processes. The degree of such overlap also depends on the nature of the selection process. For example, a selection process that uses information from the activation levels of the lexical representations to dynamically adjust the selection criterion (Nozari & Hepner, 2018) is much more likely to overlap with activation processes compared to a selection process which triggers selection as soon as any of the representations passes a fixed threshold.

It is important to note, however, that interactivity and a certain degree of modularity are not mutually exclusive. For example, the interactive model of Dell (1986) and all its subsequent variations maintain a globally modular two-step production process, despite cascading and feedback between layers: first semantic features are mapped onto lexical items, then the selected lexical item is mapped onto phonemes. This balance between modularity and interactivity is necessary to explain a wide range of empirical findings (see Dell, Nozari, & Oppenheim, 2014 for a review). Importantly, while the computational parameters that index these two steps (derived from behavioral response patterns) are uncorrelated in large samples, the neural correlates of the two steps (identified by voxel-based lesion-symptom mapping) show some overlap (Dell, Schwartz, Nozari, et al., 2013). It is thus possible to observe neural overlap for processes that are computationally dissociable. In the context of the current results, this means that activation and selection are distinct but interacting processes.

Facilitatory and inhibitory effects of semantic similarity on production

Despite being one of the most researched topics in psycholinguistics, the exact consequences of the activation of a semantically-related competitor on target selection under different circumstances has remained difficult to grasp. On the one hand, there is ample evidence that semantically-related competitors can interfere with the production of the target (e.g., Belke, Meyer, & Damian, 2005; Costa, Alario, & Caramazza, 2005; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Nozari et al., 2016; Schnur et al., 2006; 2009; Vitkovitch, Rutter, & Read, 2001; Wheeldon & Monsell, 1994). On the other hand, there are also many reports of semantic facilitation (e.g., Bloem & La Heij, 2003; Bloem, van den Boogaard, & La Heij, 2004; Finkbeiner & Caramazza, 2006; Kuipers, La Heij, & Costa, 2006; Rabovsky, Schad, & Abdel Rahman, 2016; Roelofs, 1992, 2003; Vitkovitch & Tyrrell, 1999; Wheeldon and Monsell, 1994).

Interestingly, sometimes both effects emerge within the same task, in which case semantic facilitation usually precedes semantic interference. A prominent example is the cyclic naming paradigm, where participants repeatedly name a small set of pictures presented in cycles (e.g., Belke et al., 2005; Schnur et al., 2006). The common finding in this paradigm is a transient semantic facilitation, observed only in the first cycle, which then switches to semantic interference in later cycles. This brief semantic facilitation has been attributed to spreading activation (e.g., Damian & Als, 2005) or strategic mechanisms (e.g., Belke, Shao, & Meyer, 2017; cf. Belke & Stielow, 2013), although Navarrete et al. (2012; 2014) have questioned the switch of facilitation to interference by showing that the apparent interference arises because items in the related condition do not show repetition priming past the first cycle, while those in the unrelated conditions do.

Finally, Abdel Rahman and Melinger (2009) proposed that semantically-related contexts tend to interfere with production when a cohort of competitors are activated, generally through mutual reinforcement between the target (e.g., cat) and the context (e.g., dog) which helps activate other category cohorts (e.g., other four-legged animals such as sheep, goat, etc.). Such interference is absent (and may even turn into facilitation) if the item-context combination does not mutually activate a large cohort (e.g., cat, milk; see also Mahon & Caramazza, 2009 for a criticism).

The data I reported here add to these findings by showing that the effects of semantic similarity depend on the state of the production system. When the target can use additional input to become activated, activation of semantic competitors can be helpful, because this activation propagates to the target through semantic features shared with the competitors. This is often the situation when neurotypical individuals name a picture for the first time, hence the semantic

facilitation observed in the first cycle of the cyclic naming task. Once the picture has been named, however, repetition priming is often strong enough to overshadow the little extra activation the target may receive from semantic competitors, so the facilitation effect disappears and the balance shifts towards mechanisms that create interference, such as incremental learning (e.g., Oppenheim, Dell, & Schwartz, 2010).

In an individual with *activation deficit*, the normal mechanisms of target activation are by definition impaired. Therefore any help in activating the target is beneficial. Consistent with this idea, XR's performance on reversing the names of the two pictures in the Word-pair Stroop task actually improved when the words were semantically related, compared to when they were unrelated. Note that there is no reason to assume that individuals with *activation deficit* are fundamentally immune to the mechanisms that produce similarity-induced interference. The claim, instead, is that the balance between facilitation and interference effects might be different in these individuals. The situation created by the Word-pair Stroop provides an ideal testbed for this balance: the current picture is not the target, so the target word receives little direct input on each trial. Yet it must be produced. This is the situation with the highest chance for any semantic information to help activate the target. Even if the competitor is activated along with the target, the inhibitory control operations can suppress its interfering effect. Thus the facilitation provided by the semantically-related competitor outweighs the interference it may cause, and the net effect is facilitatory. Compare this to the miscue task, where the actual target is presented on every trial. The semantic activation afforded by the actual stimulus is already much stronger than what can be activated through the onset of a semantically-related competitor. Thus, even though the semantically-related miscue could provide a slight facilitatory effect, it is too weak compared to

the actual stimulus to trump the interference effects caused by the activation of the competitor. Thus the net effect here is not facilitation.

In short, in an individual with *activation deficit*, the semantic facilitation is expected to outweigh semantic interference under conditions where direct input to the target lexical item is weak. The story is more straight-forward for *inhibition deficit*: since QD did not have trouble activating lexical items, he did not benefit from the additional activation of the target by the semantically-related competitors and only showed the interference effect, which was exaggerated because he could not suppress the semantic competitors when necessary.

Inhibitory control in word production

Several studies have demonstrated the importance of inhibitory control or the regions involved in implementing such control in word production (see Roelofs & Piai, 2011 for a related review). For example, using a Picture Word Interference paradigm, Shao, Meyer, and Roelofs (2013) showed that participants' mean response latencies were correlated with their inhibitory control ability as measured by the stop-signal task (participants made a binary button-press response on simple stimuli except when they heard a tone which prompted them to withhold their response). They also used delta plots to demonstrate that participants with small mean semantic interference effects employed selective inhibition more effectively than did participants with larger semantic interference effects (see also Sikora, Roelofs, Hermans, & Knoors, 2016). On the neural side, the link between the lateral prefrontal cortex (LPFC) and conditions with increased activation of semantic competitors has been demonstrated in both fMRI and lesion studies (e.g., de Zubicaray, Fraser, Ramajoo, & McMahon, 2017; de Zubicaray, McMahon, & Howard, 2015; Schnur et al., 2009). Furthermore, studies of verb generation have consistently found a role for

LPFC in conditions of high lexical competition (Thompson-Schill et al., 1997; Snyder, Banich, & Munakata, 2011; see Nozari & Thompson-Schill, 2015 for a review). High competition in these studies is defined as conditions where several verbs are equally plausible for a given noun (e.g., ball → play, kick, throw, etc.) as opposed to conditions where one verb is more strongly associated with the noun than other verbs (e.g., scissors → cut). Finally, Piai, Ries, and Swick (2016) found larger interference effects from lexical distractors (compared to neutral distractors such as a string of Xs) in the picture naming of six individuals with left LPFC damage than age-matched controls, showing a causal role for LPFC in naming. However, they failed to find a consistent effect of semantic relatedness (semantic vs. unrelated lexical distractors) in their patient group.

The majority of the findings discussed above point to a role for inhibitory control (or its neural correlates) in production, but all of those studies used paradigms that exaggerate the activation of lexical competitors. To my knowledge, two studies have examined the role of inhibitory control in picture naming in the absence of over-activated competitors. Shao, Roelofs, and Meyer (2012) showed that response latencies in naming both objects and actions were correlated with individuals' inhibitory control ability as measured by the stop-signal task. In another study comparing performance on pictures with high and low name agreement, Shao, Roelofs, Acheson, and Meyer (2014) found longer latencies, larger N2 amplitudes for naming pictures with low name agreement, and a negative correlation between the slowest delta segment and the N2 amplitude difference in naming high- and low-name-agreement pictures. The bulk of these arguments, however, hinges on correlational evidence.

The results of the current study thus add an important piece of evidence to the literature on the role of inhibitory control in word production by showing that a problem of inhibitory

control can impair word production even under circumstances when competitors are not artificially exaggerated, i.e., a simple picture naming task with long lags between semantically-related items. This implies that inhibitory control has a causal role in simple word production. While the greater damage to the ventrolateral prefrontal cortex and parietal cortex in QD compared to XR is compatible with an important role for these regions in implementing inhibitory control in language production, the current study was not designed to test neural hypotheses. For one thing, the individuals had premorbidly different dominant hemispheres for language processing, which makes comparisons tricky. Moreover, the extent of the damage is not exactly the same in the two individuals. Finally, both are chronic stroke survivors with years of speech therapy and very good language production abilities for the extent of their lesions, which most likely reflects substantial neural reorganization. Future studies in individuals with more comparable lesions are required to achieve a better understanding of the neural correlates of *activation deficit* vs. *inhibition deficit.*

Is inhibitory control in language production domain-general or domain-specific?

A common question in the cognitive control literature is whether different domains (e.g., language processing, visual processing, motor processing, etc.) are regulated by the same (domain-general) or different (domain-specific) control processes. As discussed above, QD was impaired on a range of tasks that required inhibitory control, including the Simon task (a spatial task) and the Word-pair Stroop task (a language production task). Can this be taken as evidence for domain-general inhibitory control?

Evidence for the domain-generality of inhibitory control often comes from neuroimaging studies that explore the cortical regions preferentially activated under conditions with high

inhibitory control demands. Some studies have found overlapping control regions across diverse tasks (e.g., January, Trueswell, & Thompson-Schill, 2009; Thompson-Schill et al., 1997). Other studies, however, have found distinct (e.g., Fedorenko, Behr, & Kanwisher, 2011) or partially distinct (Bahlmann, Blumenfeld, and D'Esposito, 2014) control regions for different tasks. The sum of this evidence suggests that while the same general regions are involved in implementing control across different tasks, there are most likely subpopulations of neurons within these regions that are specialized for regulating performance with specific types of representations. Significant damage to a large brain area, as in large strokes, can therefore manifest as impaired inhibitory control across multiple domains, without necessarily suggesting that the exact same population of neurons implements inhibitory control across different tasks.

Another piece of evidence that suggests some degree of separation between the implementation of control in language production and other domains comes from tests of "functional domain-generality". Nozari and Novick (2017) defined functional domain-generality as increased implementation of control in one task as a function of increased demand for control in another task. For example, Hsu and Novick (2016) found that responding to an incongruent (control-demanding) button-press Stroop trial facilitated the implementation of control in processing an ambiguous (control-demanding) sentence on the next trial. Freund and Nozari (2018) tested whether increasing inhibitory control demands in one task (a spatial task, Exp 1; a sentence comprehension task; Exp 2) led to better implementation of control in a subsequent word production trial in which participant had to name a picture while ignoring a distractor. No evidence of such between-task transfer was found. This absence of cross-task transfer of control was in sharp contrast to a robust within-task transfer of control: participants were better able to ignore the distractor word during picture naming after another naming trial that required them to

ignore a distractor, compared to one that did not. This effect was robust when the first naming trial immediately preceded the second (Freund, Gordon, & Nozari, 2016; see also Shitova et al., 2017 for the localization of the effect to lexical selection), as well as when the two naming trials were on average 4 s and 8 s apart with an intervening spatial or sentence comprehension trial in between them, respectively (Freund & Nozari, 2018; Exps 1 and 2). This robust within-task adaptation of control in the absence of cross-task adaptation points to functional specificity of inhibitory control in language production (see Freund & Nozari, 2018 for a proposed learning mechanism). The clinical implication of this functional specificity is that the most effective way of improving inhibitory control in language production is training inhibitory control on a language production task, as opposed to a task from a different domain like the Simon task.

Perseverations

The specific interest in lexical perseverations in this study was to test whether *activation deficit* or *inhibition deficit* alone is sufficient to produce perseveratory responses. Both participants produced tens of repeated responses in the large naming test. In keeping with past reports of linguistic perseverations, the peak of perseveration was not at lag 1, but later (Hsiao et al., 2006; Vitkovitch, Kirby, & Tyrrell, 1996), and it coincided with the average distance between two semantically-related items. To ensure that the repeated production of items indicated true perseveration, i.e., repeated productions that were influenced by a prior production as opposed to multiple independent productions of the same item, the same logic applied by McCloskey et al. (2006) was used here. Note that while perseverations are thought of as repeating a response from a previous trial (i.e., what has happened *after* a response has been produced), the statistical assessment of whether perseverations occur more often than chance or not probes the origin of a perseveratory error (i.e., what has happened *before* this particular response has been produced).

This backward focus should help with understanding the logic of the statistical procedure: it is first assessed whether an intruded response has its origin in a window within which previous responses are likely to influence the current trial. McCloskey et al. (2006) defined this window as the five preceding trials for letter perseverations. Since lexical perseverations are known to persist over longer lags (e.g., Vitkovitch et al., 1996), the window in the current study was defined as all preceding trials. In the next step, one asks how likely it is for that window to contain the origin of the intruded response by chance. If the responses in the defined window have no special influence on the production of the intrusion, then that window should have no greater likelihood of containing another instance of the same response than the rest of the list, a probability that is determined only by the number of times an item has been produced and the position of the trial in the list. This probability is then compared to the actual frequency of the perseveration origins in the defined window to determine whether that frequency exceeded chance. If so, it is reasonable to conclude that items were perseverated in the list at a rate above chance, and are thus true perseverations.

Using this method, XR had significantly above chance lexical perseverations, suggesting that an activation deficit was sufficient to produce perseverations. This finding is in agreement with the report of Fischer-Baum and Rapp (2012), who identified a failure-to-activate the current response as a definite origin of letter perseverations. We can thus conclude that failing to activate the current representation, lexical or segmental, is sufficient to generate a perseveration. QD also showed a marginal probability of producing perseverations at a rate above chance. This suggests the possibility of true perseverations in an individual with inhibition deficit, although a definite claim in this regard requires more solid evidence to be collected in future studies.

In a recent paper in this journal, we have argued that lexical selection depends on the state of the production system, as well as a selection criterion (Nozari & Hepner, 2018). The gist of this proposal is that various situations generate different levels of conflict (defined as the inverse of the difference in the activation of two items) between lexical representations. Based on where the selection criterion is placed relative to the conflict distribution, the response profile changes: a less conservative criterion allows the most highly-activated response to be selected even when conflict is high, i.e., the selected response could be very close in activation to another response. A more conservative criterion, on the other hand, only allows the selection of a given response if its activation is sufficiently higher than the competing alternative(s). In practice, the former manifests as relatively quick but error-prone responses, whereas the latter leads to slow but potentially more accurate responses. Criterion setting is an integral part of any cognitive task that entails explicit or implicit decision making, and is the key to flexible task performance in correspondence to task goals. As such, questions of criterion placement are always relevant to production performance. The critical question is whether criterion placement provides a better account of the findings discussed here than the proposed account.

Given the above descriptions, it seems that QD has a less conservative criterion for lexical selection than XR. The question is, why? One possibility is that QD has a criterion placement deficit (Anders et al., 2017). There are two reasons to be skeptical of this. (1) Analysis of individual differences in criterion setting in metacognitive judgments—which provide the ideal testbed for studying criterion placement and shifts—has revealed that when emphasis on the task goals, such as speed and accuracy, remains constant across tasks, criterion placement is consistent across tasks within the same individual. This means that individuals who tend to have a conservative criterion for judgments on a linguistic task also tend to have a conservative

criterion for judgment on non-linguistic tasks (Song et al., 2011). If we attribute the difference between QD and XR to a criterion placement deficit, then this deficit should be consistent across domains. The data do not support this prediction. While QD's performance on language tasks can be attributed to his liberal criterion, his performance on the Simon task shows the opposite pattern. His problem with the incongruent condition manifests as significantly longer RTs compared to XR and the controls, while he keeps his error rate low, as expected from a conservative criterion. Put differently, if criterion setting is the problem, then it seems to go in opposite directions in linguistic and non-linguistic tasks. One may try to justify this by arguing that the problem is simply the random placement of the criterion, but that claim is also countered by the observation that his performance across many sessions of picture naming is remarkably consistent and in keeping with a liberal selection criterion. The sum total of these findings point to a stable liberal criterion specifically in the language production system. But is this a deficit?

(2) To determine that a certain criterion placement is deficient, one must be able to show that placing the criterion in a different place would have led to better performance with respect to the task goal. We have shown that in damaged production systems, performance is likely to be suboptimal regardless of where the criterion is placed (Nozari & Hepner, 2018). The speaker with aphasia has the choice of producing either faster and more error-prone speech or many omissions, neither of which are aligned with the task goal. Thus there is no "good criterion" per se in such speakers against which a "bad criterion" can be measured. The argument for a criterion setting problem can thus only be made if the same individual cannot adjust their criterion based on changing task goals. Otherwise, where an individual with aphasia chooses to place their selection criterion may reflect their personal strategy for coping with the challenges of lexical selection in a damaged production system, as opposed to a deficit per se.

In summary, XR and QD are both adopting perfectly reasonable strategies as far as criterion placement goes. This strategy, however, does not explain other differences between them including the difference in their performance on the Simon task. More generally, appealing to the placement of the criterion as a mechanistic explanation is useful in cases where performance is expected to change based on changing task goals; otherwise, pointing out differences in criterion placement is simply a redescription of the observed data patterns, with little explanatory value. In the same vein, deficits of criterion setting can only be established by demonstrating that speakers fail to make adjustments to the criterion in order to better meet the new goals.

Conclusions

The data reported in this paper show that activation and selection are separable (albeit interacting) operations and can be selectively impaired. While in both cases the behavioral outcome is more lexical errors, the underlying etiologies are different: impaired activation of lexical items (*activation deficit*) stems from weak connections between sematic features and lexical representations. Impaired selection (*inhibition deficit*) stems from an impaired inhibitory control mechanism which cannot successfully suppress the activated competitors. Theoretically, these findings point to a critical role for inhibitory control in lexical selection. Clinically, the results call for more careful evaluation of the etiology of "access deficit" on a case by case basis, in order to devise the optimal treatment plan for each individual.

Acknowledgments

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*(5), 713-734. https://doi-org.proxy1.library.jhu.edu/10.1080/01690960802597250

Ackerman, T., & Ellis, A. W. (2007). Case study: Where do aphasic perseverations come from?. *Aphasiology*, *21*(10-11), 1018-1038. https://doi.org/10.1080/02687030701198361

Allen, C. M., Martin, R. C., & Martin, N. (2012). Relations between short-term memory deficits, semantic processing, and executive function. *Aphasiology*, *26*(3-4), 428-461. http://dx.doi.org/10.1080/02687038.2011.617436

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47-89. https://doi.org/10.1016/s0079-7421(08)60452-1

Bahlmann, J., Blumenfeld, R. S., & D'esposito, M. (2014). The rostro-caudal axis of frontal cortex is sensitive to the domain of stimulus information. *Cerebral cortex*, *25*(7), 1815-1826. https://doi.org/10.1093/cercor/bht419

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

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

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *Quarterly Journal of Experimental Psychology*, *66*(11), 2135-2160. https://doi.org/10.1080/17470218.2013.775318

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and language*, *48*(3), 468-488. https://doi.org/10.1037/e501882009-509

Bloem, I., van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, *51*(2), 307-323. https://doi.org/10.1016/j.jml.2004.05.001

Cohen, L., & Dehaene, S. (1998). Competition between past and present. Assessment and interpretation of verbal perseverations. *Brain: a journal of neurology*, *121*(9), 1641-1659. https://doi.org/10.1093/brain/121.9.1641

Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. A. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. *Brain*, *132*(9), 2593-2608. https://doi.org/10.1093/brain/awp146

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*(1), 125-131. https://doi.org/10.3758/bf03196357

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cognitive neuropsychology*, *27*(3), 245-260. https://doi.org/10.1080/02643294.2010.513967

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1372. https://doi.org/10.1037/0278-7393.31.6.1372

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

Dell, G. S., Burger, L. K., & Svec, W. R. (1997b). Language production and serial order: A functional analysis and a model. *Psychological review*, *104*(1), 123. https://doi.org/10.1037//0033-295x.104.1.123

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

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997a). Lexical access in aphasic and nonaphasic speakers. *Psychological review, 104*(4), 801. https://doi.org/10.1037//0033-295x.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. https://doi.org/10.1016/j.cognition.2013.05.007

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. https://doi.org/10.1016/j.neuropsychologia.2017.01.010

de Zubicaray, G., McMahon, K., & Howard, D. (2015). Perfusion fMRI evidence for priming of shared feature-to-lexical connections during cumulative semantic interference in spoken word production. *Language, Cognition and Neuroscience, 30*(3), 261-272. https://doi.org/10.1080/01690965.2013.848990

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. https://doi.org/10.1073/pnas.1112937108

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*(6), 790-796. https://doi.org/10.1016/s0010-9452(08)70419-2

Fischer-Baum, S., McCloskey, M., & Rapp, B. (2010). Representation of letter position in spelling: Evidence from acquired dysgraphia. *Cognition*, *115*(3), 466-490. https://doi.org/10.1016/j.cognition.2010.03.013

Fischer-Baum, S., & Rapp, B. (2012). Underlying cause (s) of letter perseveration errors. *Neuropsychologia*, *50*(2), 305-318. https://doi.org/10.1016/j.neuropsychologia.2011.12.001

Freedman, M. L., & Martin, R. C. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology*, *18*(3), 193-226. https://doi.org/10.1080/02643290042000080

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. https://doi.org/10.1016/j.cognition.2018.03.009

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 1-13.

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *The Lancet Neurology*, *6*(11), 1004-1014. https://doi.org/10.1016/s1474-4422(07)70266-1

Hoffman, P., McClelland, J. L., Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological review*, *125*(3), 293. http://dx.doi.org/10.1037/rev0000094

Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological research*, *68*(1), 1-17. https://doi.org/10.1007/s00426-003-0132-y

Houghton, G., Glasspool, D., & Shallice, T. (1994). Spelling and serial recall: Insights from a competitive queueing model. In G. D. A. Brown, & N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention.* John Wiley & Sons: Chichester. https://doi.org/10.1002/(sici)1099-0909(199806)4:2%3C112::aid-dys76%3E3.0.co;2-a

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

Hsiao EY, Schwartz MF, Schnur TT, Dell GS. Temporal characteristics of semantic perseverations induced by blocked-cyclic picture naming. Brain and Language 2009;108:133–144. https://doi.org/10.1016/j.bandl.2008.11.003

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence

processing. *Journal of Cognitive Neuroscience*, *21*(12), 2434-2444.

https://doi.org/10.1162/jocn.2008.21179

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus

semantic dementia: a case-series comparison. *Brain*, *129*(8), 2132-2147.

https://doi.org/10.1093/brain/awl153

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011).

Relating inter-individual differences in metacognitive performance on different perceptual

tasks. *Consciousness and cognition*, *20*(4), 1787-1792.

http://dx.doi.org/10.1016/j.concog.2010.12.011

Kuipers, J. R., La Heij, W., & Costa, A. (2006). A further look at semantic context effects in

language production: The role of response congruency. *Language and Cognitive

Processes*, *21*(7-8), 892-919. https://doi.org/10.1080/016909600824211

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and

computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42-55.

https://doi.org/10.1038/nrn.2016.150

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.

https://doi.org/10.1017/s0140525x99001776

MacKay, D. (1986). Self-inhibition and the disruptive effects of internal and external feedback in

skilled behavior. In H. Heuer, & C. Fromm (Eds.), *Generation and modulation of action

patterns.* London: Springer. https://doi.org/10.1007/978-3-642-71476-4

Mahon, B. Z., & Caramazza, A. (2009). Why does lexical selection have to be so hard? Comment on Abdel Rahman and Melinger's swinging lexical network proposal. *Language and Cognitive Processes*, *24*(5), 735-748. https://doi-org.proxy1.library.jhu.edu/10.1080/01690960802597276

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

Martin, N., & Dell, G. S. (2007). Common mechanisms underlying perseverative and non-perseverative sound and word substitutions. *Aphasiology*, *21*(10-11), 1002-1017. https://doi.org/10.1080/02687030701198346

McCloskey, M., Macaruso, P., & Rapp, B. (2006). Grapheme-to-lexeme feedback in the spelling system: Evidence from a dysgraphic patient. *Cognitive Neuropsychology*, *23*(2), 278–307. https://doi.org/10.1080/02643290442000518

Mirman, D., & Britt, A. E. (2014). What we talk about when we talk about access deficits. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634), 20120388. https://doi.org/10.1098/rstb.2012.0388

Moses, M. S., Sheard, C., & Nickels, L. A. (2007). Insights into recurrent perseverative errors in aphasia: A case series approach. *Aphasiology*, *21*(10-11), 975-1001. https://doi.org/10.1080/02687030701198312

Navarrete, E., Del Prato, P., & Mahon, B. Z. (2012). Factors determining semantic facilitation and interference in the cyclic naming paradigm. *Frontiers in psychology*, *3*, 38. http://dx.doi.org/10.3389/fpsyg.2012.00038

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

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

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

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

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

Nozari, N., Hepner, C. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology.*

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in

aphasia: Steps, routes, and frequency effects. Journal of Memory and Language, 63(4), 541–559.

https://doi.org/10.1016/j.jml.2010.08.001

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current*

Directions in Psychological Science, *26*(5), 403-410. https://doi.org/10.1177/0963721417702419

Nozari, N. & Thompson-Schill, S. L. (2015). Left Ventrolateral Prefrontal Cortex in Processing

of Words and Sentences. In G. Hickok & S.L. Small (Eds.), *The Neurobiology of*

Language. Academic Press. https://doi.org/10.1016/b978-0-12-407794-2.00001-8

Olivella, S., & Shiraito, Y. (2017). Poisbinom: A Faster Implementation of the Poisson-Binomial

Distribution. R package version 1.0.1.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning:

A model of cumulative semantic interference during lexical access in speech

production. *Cognition*, *114*(2), 227-252. https://doi.org/10.1016/j.cognition.2009.09.007

Piai, V., Riès, S. K., & Swick, D. (2016). Lesions to lateral prefrontal cortex impair lexical

interference control in word production. *Frontiers in human neuroscience*, *9*, 721.

https://doi.org/10.3389/fnhum.2015.00721

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct Patterns of

Brain Activity Characterise Lexical Activation and Competition in Spoken Word Production.

PLoS ONE, 9(2), e88674. https://doi.org/10.1371/journal.pone.0088674

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240-244. https://doi.org/10.1016/j.cognition.2015.09.016

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

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

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia naming test: scoring and rationale. *Clinical aphasiology*, *24*, 121-133. https://doi.org/10.1037/t56477-000

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

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: a review. *Frontiers in psychology*, *2*, 307. https://doi.org/10.3389/fpsyg.2011.00307

Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220-239. https://doi.org/10.1016/j.neuropsychologia.2015.04.015

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

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

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, *92*, 289-303. https://doi.org/10.1016/j.cortex.2017.04.017

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and language*, *54*(2), 228-264. https://doi.org/10.1016/j.jml.2005.10.001

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & cognition*, *41*(8), 1200-1211. https://doi.org/10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *brain research*, *1586*, 130-142. https://doi.org/10.1016/j.brainres.2014.07.009

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. https://doi.org/10.1080/17470218.2012.670252

Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *The Quarterly Journal of Experimental Psychology*, *69*(9), 1719-1740. https://doi.org/10.1016/j.neuropsychologia.2016.01.037

Snyder, H. R., Banich, M. T., & Munakata, Y. (2011). Choosing our words: retrieval and selection processes recruit shared neural substrates in left ventrolateral prefrontal cortex. *Journal of cognitive neuroscience*, *23*(11), 3470-3482. https://doi.org/10.1162/jocn_a_00023

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

Vitkovitch, M., Kirby, A., & Tyrrell, L. (1996). Patterns of excitation and inhibition in picture naming. Visual Cognition, 3, 61–80. https://doi.org/10.1080/713756730

Vitkovitch, M., Rutter, C., & Read, A. (2001). Inhibitory effects during object name retrieval: The effect of interval between prime and target on picture naming responses. *British journal of psychology*, *92*(3), 483-506. https://doi.org/10.1348/000712601162301

Vitkovitch, M., & Tyrrell, L. (1999). The effects of distractor words on naming pictures at the subordinate level. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(4), 905-926. https://doi.org/10.1080/027249899390864

Warrington, E. K., & McCarthy, R. (1983). Category specific access dysphasia. *Brain*, *106*(4), 859-878. https://doi.org/10.1093/neucas/8.3.192

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language*, *33*(3), 332-356. https://doi.org/10.1006/jmla.1994.1016

Figure captions

Figure 1. A schematic of the mapping of semantic features to lexical representations (words) for the target CAT. Darker circles represent higher activation, which is proportional to the number of features which send activation to the lexical representation. The dotted line represents an inactive connection.

Figure 2. RTs (±*SE*) for XR and QD in Congruent and Incongruent conditions in the Simon task.

Figure 3. Proportion of miscue errors vs. other error types in all error trials in the miscue condition in XR and QD. The total number of errors in this condition is 8 for XR and 13 for QD.

Figure 4. Stroop effects in errors (errors in reversed naming − errors in straight naming) for XR, QD, and controls (±*SD*) in the unrelated and semantically-related conditions in the word-Stoop task.

Figure 5. The probability density of lexical perseveration by lag for QD and XR. The vertical line (lag 21) represents the average lag between two semantically-related items, calculated as the distance between each item and the next item from the same semantic category, averaged over all

items. This lag coincides with the peak in the probability of producing a lexical perseveration in both participants.

Footnotes

1. Note that while semantic similarity increases the activation of the competitors, in systems with feedback from lexical to semantic representations, the activation has the greatest convergence on the target.

2. The reason for examining the effect of semantic similarity in the reversed as opposed to straight naming was that the two participants were fairly high-level, and made few errors in the straight conditions (i.e., there was a ceiling effect in accuracy). The reversed condition increased task difficulty and allowed differences to show up.
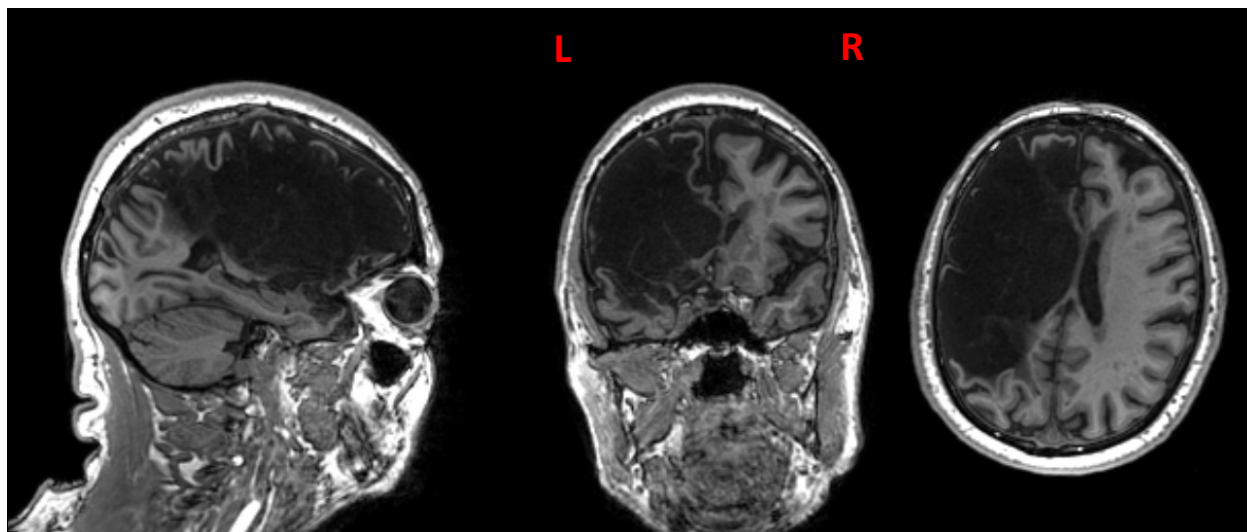
APPENDIX A

Multiview (sagittal, coronal and axial) structural MRI scans for XR and QD. Lesions are shown in native space (the images have not been normalized). L = left, R = right.

XR (X = 116, Y = 114, Z = 157)
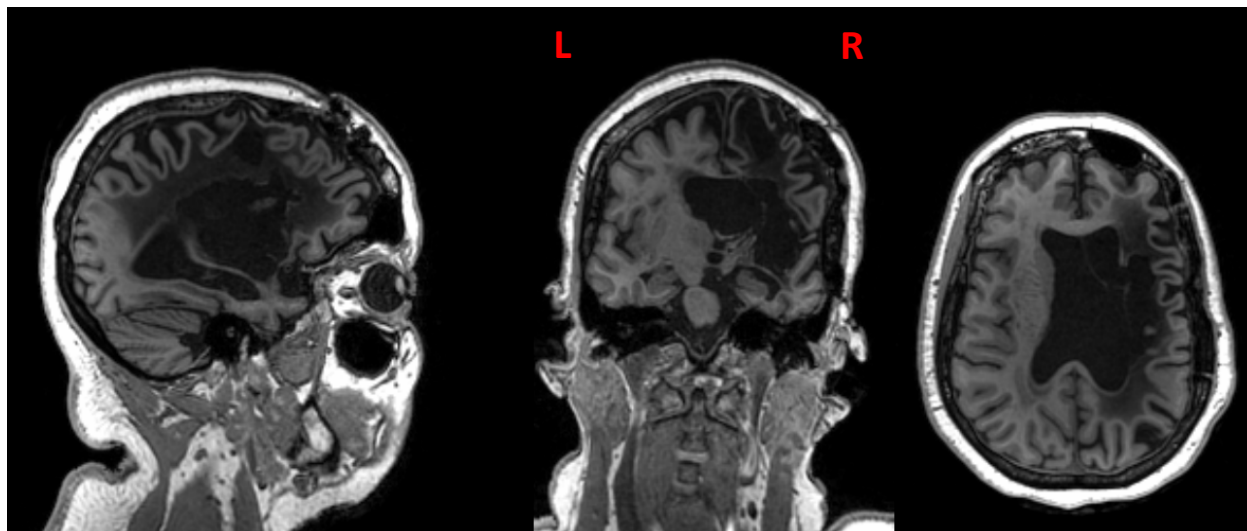
QD (X = 61, Y = 148, Z = 157)

Breakdown of response types in the large naming test for XR and QD. The 444 pictures were named twice on different occasions, and responses are presented separately by the 1st and 2nd encounters. Each response type is described underneath the table. The classification system is based on Dell et al. (1997a).

Response type	XR		QD	
	1st encounter	2nd encounter	1st encounter	2nd encounter
correct	281	280	218	223
semantic	68	76	63	95
phonological	4	10	8	6
mixed	26	23	17	23
compound	3	5	3	9
unrelated	6	2	3	1
nonword	2	2	5	4
non-commission	54	46	127	83
Semantically-related	94	99	80	118
All commissions	109	118	99	138
Semantic/Commission	0.86	0.84	0.81	0.86
Sum of all responses	444	444	444	444

1. Correct: response is the same lexical item as the target and is pronounced correctly. Plural form were accepted as correct responses, e.g., cat → cats. Elaborations that contained the correct label were also accepted as correct, e.g., cat → a lovely cat.

2. Semantic: response bears a taxonomic or thematic relation to the target, e.g., cat → dog, or saddle → horse.

3. Phonological: target shares at least one phoneme in the same syllabic position as in the target word, or two or more phonemes in any position, e.g., cat → cap.

4. Mixed: response bears both a semantic and a phonological relation to the target, e.g., cat → rat.

5. Compound: response replaces one of the morphemes in a polymorphemic words, e.g., waterfall → watermelon. These responses are coded as a separate category, since they sometimes result in semantically-related and sometimes semantically-unrelated words, but share the critical feature of having one morpheme (often the first one) produced correctly.

6. Unrelated: response is a word with no clear semantic or phonological relation to the target, e.g., cat → pencil.

7. Nonword: response is a nonlexical item that may or may not bear phonological resemblance to the target, e.g., cat → lat.

8. Non-commission: this error type encompasses several categories that have the common feature of never containing a clear single-word response that can be taken as the response to the picture. These include: omissions (no response was provided), descriptions and circumlocutions (participant makes general or specific comments about the item without ever providing a label, e.g., cat → it walks around in my yard, or cat → I don't like them), and non-item-related comments (e.g., "why can't I remember this? It's so simple…").

* Since all pictures were of real-life objects, we did not encounter "visual errors" which are often a separate category of responses in black and white line-drawings. We also avoided using pictures that contained multiple objects in order to minimize the chance of producing picture-part errors, such as a garage with a car in it.