Listening to Yourself and Watching Your Tongue: Distinct Abilities and Brain Regions for Monitoring Semantic and Phonological Speech Errors

Joshua D. McCall[1], Andrew T. DeMarco[1,2], Ayan S. Mandal[1,3], Mackenzie E. Fama[1,4], Candace M. van der Stelt[1,5], Elizabeth H. Lacey[1,5], Alycia B. Laks[1], Sarah F. Snider[6], Rhonda B. Friedman[6], and Peter E. Turkeltaub[1,2,5,6]

Abstract

■ Despite the many mistakes we make while speaking, people can effectively communicate because we monitor our speech errors. However, the cognitive abilities and brain structures that support speech error monitoring are unclear. There may be different abilities and brain regions that support monitoring phonological speech errors versus monitoring semantic speech errors. We investigated speech, language, and cognitive control abilities that relate to detecting phonological and semantic speech errors in 41 individuals with aphasia who underwent detailed cognitive testing. Then, we used support vector regression lesion symptom mapping to identify brain regions supporting detection of phonological versus semantic errors in a group of 76 individuals with aphasia. The results revealed that motor speech deficits as well as lesions to the ventral motor cortex were related to reduced detection of phonological errors relative to semantic errors. Detection of semantic errors selectively related to auditory word comprehension deficits. Across all error types, poor cognitive control related to reduced detection. We conclude that monitoring of phonological and semantic errors relies on distinct cognitive abilities and brain regions. Furthermore, we identified cognitive control as a shared cognitive basis for monitoring all types of speech errors. These findings refine and expand our understanding of the neurocognitive basis of speech error monitoring. ■

INTRODUCTION

Speech Error Monitoring in Aphasia

Effective spoken communication relies on detection and correction of speech errors, an ability referred to as speech error monitoring (SEM). While there are many proposals concerning the cognitive and neural basis of SEM, there is little consensus about how these proposals fit together and their relative accuracy in depicting SEM (Gauvin & Hartsuiker, 2020; Nozari, 2020). Our understanding of SEM might be clarified by identifying deficits and lesions that are associated with reduced SEM. Such associations can be identified in aphasia, a language impairment that commonly results from stroke. Reduced SEM in aphasia depends on the location and size of the stroke lesion and is associated with certain behavioral deficits (e.g., Mandal et al., 2020). Examining the causes of reduced SEM in aphasia could broadly inform our understanding of the cognitive and neural basis of SEM. Moreover, understanding more about SEM in aphasia could inform future therapies to improve SEM ability or inform treatment decisions for individuals with SEM deficits that might preclude specific therapeutic approaches that strongly rely on SEM.

Critical Speech and Language Processes for SEM in Aphasia

To understand the mental processes that are critical for SEM in aphasia, we must first consider the various proposed models of how speakers monitor their own speech. Speech processing encompasses various stages, including lexical access, motor aspects of speech production, speech perception, and speech comprehension (Levelt, Roelofs, & Meyer, 1999; Dell & O'Seaghdha, 1992). Correspondingly, SEM models vary by which aspects of speech processing they emphasize. Notably, however, the different stages described in SEM models are not necessarily mutually exclusive (Nozari, 2020; Postma, 2000). In addition, these models differ in the extent to which SEM relies

[1]Center for Brain Plasticity and Recovery, Neurology Department, Georgetown University Medical Center, Washington, DC, [2]Rehabilitation Medicine Department, Georgetown University Medical Center, Washington, DC, [3]Brain-Gene Development Lab, Psychiatry Department, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, [4]Department of Speech, Language, and Hearing Sciences, The George Washington University, Washington, DC, [5]Research Division, MedStar National Rehabilitation Hospital, Washington, DC, [6]Center for Aphasia Research and Rehabilitation, Georgetown University Medical Center, Washington, DC

on cognitive control (Gauvin & Hartsuiker, 2020; Roelofs, 2020; Nozari, Dell, & Schwartz, 2011). Here, the term cognitive control is used broadly to refer to cognitive abilities that support dynamic adaptation of mental processing. Because contemporary models of SEM differ on the aspects of speech-language processing and cognitive control that contribute to SEM, they make distinct predictions about which deficits should reduce SEM in aphasia. Broadly, three groups of models exist, which are discussed below.

One group of SEM models emphasizes the role of motor speech and perceptual processing in SEM. This group of models, as well as the speech processing models that contain them, is referred to by many names, including forward models, feedforward models, efference models, production-perception models, and sensorimotor models (Gauvin & Hartsuiker, 2020; Nozari, 2020; Hickok, 2012; Tourville & Guenther, 2011; Postma, 2000). Here, we refer to such SEM models as sensorimotor models (Figure 1, blue and magenta). Two prominent sensorimotor models include Hickok's Hierarchical State Feedback Control (HSFC) model (Hickok, 2012), and Guenther and colleagues' Directions into Velocities of Articulators (DIVA) model (Tourville & Guenther, 2011; Bohland, Bullock, & Guenther, 2010; Guenther, 2006). In HSFC as well as DIVA, expectations of the perceptual consequences of speech (i.e., the somatosensory and auditory consequences) are generated during preparation for motor speech execution. These expectations are then compared with the actual perception of one's own speech. If the

actual perception does not match the expectation, then a speech error is detected and corrected. As the comparison between expectation and perception is performed in perceptual modalities, these models suggest that perceptual deficits might impair SEM. Although these models describe motor speech execution, they do not clearly predict that motor speech deficits would impair SEM. On the contrary, motor speech deficits such as apraxia of speech could potentially distort perceptual expectations (Miller & Guenther, 2021). We consequently suggest that such distortions might paradoxically increase SEM, because these distortions would increase the mismatch between expected and actual perceptions.

Neither DIVA nor HSFC incorporates a role for cognitive control, so they do not predict that cognitive control deficits in aphasia should impair SEM. Importantly, as noted by Gauvin and Hartsuiker (2020), although these models present clear mechanisms for SEM of phonological errors (i.e., errors related to the sound of the intended words), they do not present clear mechanisms for SEM of semantic errors (i.e., errors related to the meaning of the intended words; Gauvin & Hartsuiker, 2020; Nozari, 2020). Thus, predictions by DIVA and HSFC may be more relevant in the context of SEM of phonological errors than semantic errors.

Notably, HSFC can be elaborated to account for SEM of lexical errors, because it includes architecture at the level of the lemma as well as a conceptual level (Hickok, 2012). Furthermore, models with similar principles to HSFC describe syntactic processing (e.g., Matchin & Hickok,
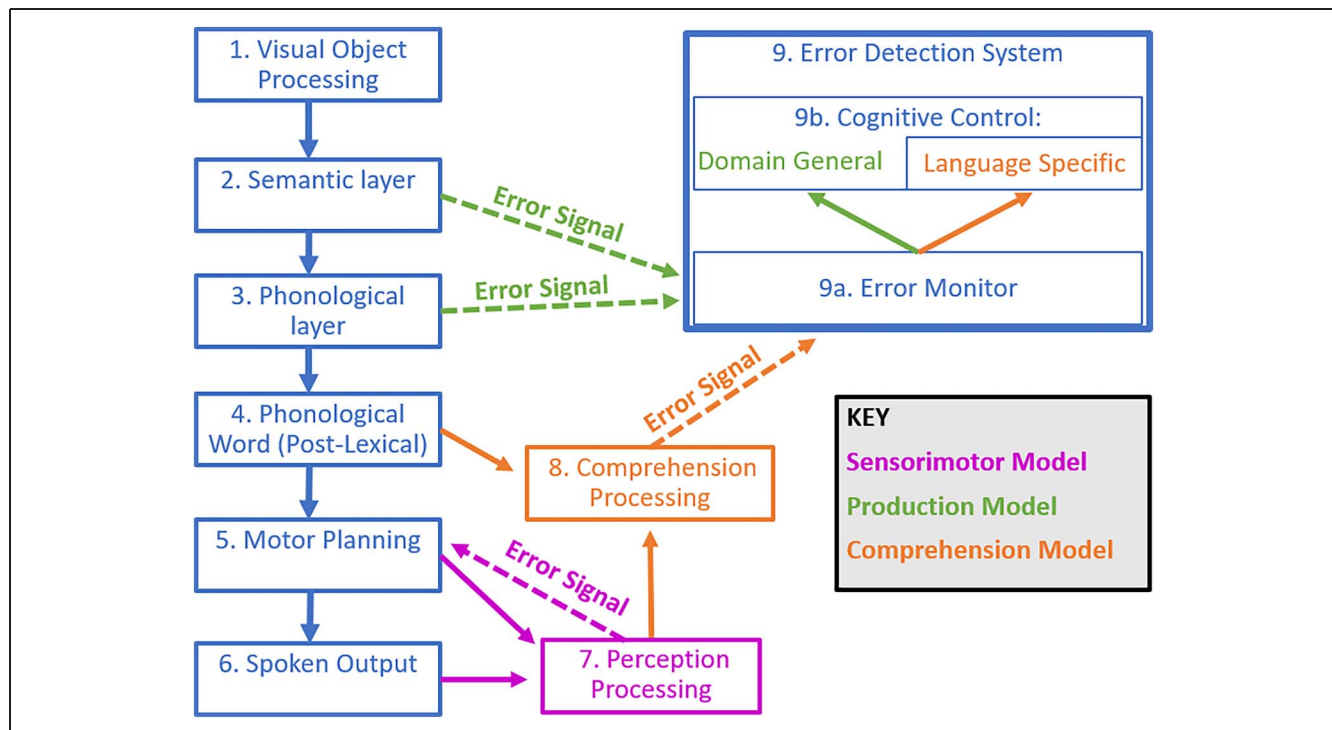

Figure 1. Models of SEM during picture naming. Adapted from Goldrick and Rapp (2007) incorporating aspects of Fama and Turkeltaub (2020), Roelofs (2020), Hickok (2012), Nozari et al. (2011), Tourville and Guenther (2011), and Levelt (1983). Theoretical components shared across families of SEM models are outlined in blue. Components specific to production models are shown in green; comprehension models, in orange; sensorimotor models, in magenta.

2020) and could thus account for SEM of syntactic errors, an error type not readily addressed by most models of SEM. Although HSFC and related models can potentially account for SEM of lexical and syntactic errors, it could be useful for them to address SEM of semantic errors as well.

Consistent with sensorimotor models, it appears that many people with aphasia are impaired at monitoring the sensory consequences of their speech. Such a sensorimotor monitoring impairment can be measured in altered auditory feedback studies, where participants produce speech while hearing an electronically altered version of their speech (Sangtian, Wang, Fridriksson, & Behroozmand, 2021; Johnson, Sangtian, Johari, Behroozmand, & Fridriksson, 2020). Often in these paradigms, participants adjust their speech to feedback, despite not being explicitly instructed to do so (e.g., Burnett, Freedland, Larson, & Hain, 1998). For example, if the altered auditory feedback alterations include a raised pitch, then a participant might compensate by lowering their pitch. These altered auditory feedback studies reveal that people with aphasia are impaired at compensating their speech in response to altered feedback; specifically, they are slower to compensate (Johnson et al., 2020) and less accurate at detecting feedback alterations (Sangtian et al., 2021). However, researchers have pointed out limitations of altered auditory feedback paradigms including that the altered auditory feedback does not approximate naturally occurring speech errors (Meekings & Scott, 2021), and that the paradigm is inherently limited to external detections of speech errors (i.e., because the error is externally applied after production, compensation cannot reflect SEM during preparation for speech production; Gauvin, De Baene, Brass, & Hartsuiker, 2016). It is also worth noting that altered auditory feedback paradigms may have limited relevance for SEM of semantic errors, as sensorimotor processing does not clearly explain SEM of semantic errors. Therefore, the current study investigated the role of motor speech and perceptual processing in SEM of spontaneous semantic and phonological speech errors during picture naming to more closely approximate naturally occurring speech errors.

A second group of SEM models emphasizes the role of lexical access in SEM. This group has been referred to as production-based models (Figure 1, blue and green). One notable example of a production-based model is the conflict-based account proposed by Nozari and colleagues (2011), which emphasizes the role of SEM during lexical access in speech production. In the conflict-based account, semantic and phonological errors are detected when there is increased conflict during the semantic or phonological encoding stages of lexical access (Figure 1, Boxes 2 and 3). As such, the conflict-based account predicts that deficits in the semantic encoding stage of lexical access would impair SEM of semantic errors, and that deficits in the phonological encoding stage of lexical access would impair SEM of phonological errors. Another notable production-based model is Gauvin and

Hartsuiker's Hierarchical Conflict Model for Self and Other Monitoring, here referred to as the Gauvin and Hartsuiker model (Gauvin & Hartsuiker, 2020). Similar to the conflict-based account, the Gauvin and Hartsuiker model describes the importance of lexical access in SEM (Gauvin & Hartsuiker, 2020). The model builds on the conflict-based account by adding a role for speech perception in SEM. Thus, the Gauvin and Hartsuiker model makes the same predictions as the conflict-based account with regard to how lexical access deficits would impair SEM in aphasia, with an additional prediction that deficits in speech perception might impair SEM in aphasia.

Consistent with production-based models, a study of spontaneous speech errors in a group of people with aphasia found that SEM of semantic errors relates to *s weights*, a measure of semantic aspects of lexical access derived from a computational model of picture naming performance (Nozari et al., 2011). The same study also found that SEM of phonological errors relates to *p weights*, a measure of phonological aspects of lexical access derived from the same model mentioned above (Nozari et al., 2011). In addition, another study in a group of people with aphasia found that SEM of spontaneous errors was related to a measure of fluency, specifically the mean length of utterance produced while describing a picture (Mandal et al., 2020). The relationship between fluency and SEM in aphasia is broadly consistent with a role of lexical access in SEM as predicted by production-based models. Alternatively, the relationship between fluency and SEM might be explained by other aspects of speech processing such as motor speech abilities.

A third group of SEM models emphasizes the role of comprehension in SEM. These models are referred to here as comprehension-based models of SEM (Figure 1, blue and orange). One of the most well-known comprehension-based models is Levelt's Perceptual Loop Theory (Levelt, 1983), and one of the most recent comprehension-based models is Roelofs' WEAVER++ (Roelofs, 2014). In comprehension-based models, one detects errors in their own speech using the same comprehension abilities used for understanding someone else's speech. One may comprehend either their externally produced speech (Figure 1, Box 6) or their inner speech (Figure 1, Box 4). In comprehension-based models, SEM of semantic errors occurs when a person comprehends their own speech, and notices that it is different from what they intended to say. The specific mechanisms for SEM of phonological errors are not described in depth in comprehension-based models. One mechanism discussed for comprehension-based models is that comprehended speech is evaluated using a lexicality criterion (Roelofs, 2004), which could conceivably allow for SEM of nonword phonological errors. On the other hand, SEM of real-word phonological errors could conceivably occur through the same mechanism of SEM of semantic errors. In their simplest interpretation, comprehension-based models would predict that comprehension deficits impair SEM in aphasia.

The evidence for comprehension deficits impairing SEM in aphasia is mixed. On one hand, several researchers have identified a double dissociation between SEM and comprehension in case studies of people with aphasia (for a review, see Nozari et al., 2011; but also see Roelofs, 2020, for an explanation of how double dissociations can still be consistent with a comprehension-based model). On the other hand, two studies have found that SEM of spontaneous errors in groups of people with aphasia relates to performance on an auditory word to picture matching task, suggesting some role for comprehension abilities in SEM (Mandal et al., 2020; Dean, Della Sala, Beschin, & Cocchini, 2017).

Critical Cognitive Control Abilities for SEM in Aphasia

Several domain-general error monitoring theories, and some of the SEM models discussed above, propose that error monitoring interacts with cognitive control. For example, Conflict Monitoring Theory, as described by Botvinick and colleagues, suggests that upon detecting an error, the error monitoring system recruits cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Recruitment of cognitive control can then aid in correction of the error, as well as other posterror behavioral adjustments (Danielmeier & Ullsperger, 2011). In addition, Hierarchical Representation Theory suggests a bidirectional relationship in which the error monitoring recruits cognitive control, and cognitive control also can boost error monitoring (Alexander & Brown, 2015).

In line with domain-general theories of error monitoring, both production-based models reviewed above (i.e., the conflict-based account as well as the Gauvin and Hartsuiker model) as well as WEAVER++ emphasize a role of cognitive control in SEM. In particular, the production-based models both describe a role of *domain-general* cognitive control (Figure 1, Box 9B) in error monitoring. For instance, the conflict-based account proposes that dynamics in lexical access are monitored for conflict via a domain-general monitor that recruits domain-general cognitive control (Figure 1, green arrows). On the other hand, WEAVER++ discusses a role of *language-specific* cognitive control (Figure 1, Box 9B; Roelofs, 2020). As mentioned above, in WEAVER++, errors are detected as mismatches in the comparison between one's intended message and the comprehension of their spoken output. WEAVER++ proposes that language-specific cognitive control is needed to perform this comparison (Figure 1, orange arrows). Thus, the conflict-based account and the Gauvin and Hartsuiker models predict that domain-general cognitive control deficits should impair SEM in aphasia, whereas Roelofs' WEAVER++ predicts that language-specific cognitive control deficits should impair SEM in aphasia. Again, in contrast, sensorimotor models of SEM do not describe any role of cognitive control in SEM.

There is little evidence concerning whether cognitive control deficits underlie reduced SEM ability in aphasia. Mandal and colleagues (2020) found that SEM in aphasia relates to an indirect measure of cognitive control, the difference between forward minus backward span performance (Mandal et al., 2020). This measure is a broad assessment of cognitive ability in that it may reflect multiple cognitive control processes. Moreover, there are multiple interpretations of poor scores on span tasks that do not reflect cognitive control deficits, such as impaired articulatory rehearsal (Ghaleh et al., 2019). In contrast, Dean and colleagues (2017) examined another broad measure of cognitive ability, the Brixton spatial anticipation test, and found it had no relationship with SEM in aphasia (Dean et al., 2017). Overall, available evidence relating cognitive control deficits to SEM in aphasia does not differentiate between specific cognitive control abilities, nor specify whether the cognitive control abilities are domain-general or language-specific.

Critical Neural Substrates for Monitoring Phonological and Semantic Errors in Aphasia

Our neurocognitive understanding of SEM requires identifying the critical brain regions for SEM. Critical brain regions have been investigated using functional neuroimaging in neurotypical speakers and lesion studies in aphasia. In neurotypical speakers, errors in verbal and nonverbal tasks activate regions in the posterior medial frontal cortex (pMFC; Ridderinkhof, 2004). Correspondingly, functional neuroimaging of SEM of spontaneous speech errors consistently demonstrates activation of the pMFC, as well as left hemisphere (LH) regions thought to be involved in speech processing such as the inferior frontal gyrus (IFG) and insula (Runnqvist et al., 2021; Gauvin et al., 2016; Abel et al., 2009). Thus, functional neuroimaging studies suggest that the pMFC, IFG, and insula may critically support SEM.

To date, few studies have investigated the neural structures that, when lesioned, reduce SEM in aphasia. Altered auditory feedback studies in aphasia indicate that feedback compensation is impaired by lesions in the left middle and superior temporal gyri, IFG, and supramarginal gyrus (Behroozmand et al., 2018). However, as noted above, altered-auditory feedback paradigms might not generalize to SEM of natural speech errors and may have limited relevance for detecting semantic errors. One prior lesion study evaluated SEM of spontaneous speech errors during picture naming, which might better approximate natural speech errors. This study found that LH frontal white matter damage reduces SEM in aphasia (Mandal et al., 2020). This study further implicated frontal white matter damage in phonological SEM but did not identify lesions that reduce semantic SEM, possibly because the sample size was too small. A follow-up analysis of the participants from Mandal and colleagues (2020) directly examined white matter fiber tracts that, when lesioned, reduce SEM. Using a connectome-based approach, McCall,

Dickens, and colleagues (2022) found that reduced SEM corresponded to lesioned tracts connecting lateral frontal brain regions implicated in speech production with medial frontal brain regions implicated in cognitive control. These lesioned connections corresponded to SEM of phonological errors; however, the authors did not identify lesioned connections that correspond to SEM of semantic errors.

Importantly, the SEM models described in the Introduction section predict that different neural substrates may underlie SEM for phonological versus semantic errors. For example, production-based models predict that SEM of phonological errors should rely in part on brain regions involved in phonological encoding, which are thought to include dorsal stream regions spanning from the left posterior superior temporal cortex to the left precentral gyrus and the left IFG (Roelofs, 2014; Schwartz, Faseyitan, Kim, & Coslett, 2012; Indefrey, 2011). In contrast, production-based models predict that SEM of semantic errors should rely on brain regions involved in the semantic encoding stage of lexical access. Although semantic knowledge is thought to be widely distributed throughout the cortex (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017; Binder, Desai, Graves, & Conant, 2009), some models suggest that semantic aspects of lexical access relies on the left middle temporal gyrus (Indefrey, 2011; Hickok & Poeppel, 2007), as well as the left anterior temporal lobe (Schwartz et al., 2009; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004). In contrast to production models, comprehension models do not clearly predict that different neural substrates underlie SEM for phonological versus semantic errors. In these models, SEM of all error types should rely on brain regions involved in speech comprehension, which include the superior temporal cortex (Specht, 2014).

Language comprehension more broadly has been proposed to rely on a set of brain regions organized sequentially in a posterior–anterior direction, starting posteriorly in the temporal cortex and ending anteriorly in anterior temporal cortex and/or anterior inferior frontal cortex (Matchin & Hickok, 2020; Pylkkänen, 2019; Friederici, 2017; Hickok & Poeppel, 2007). This organization, or pathway, of brain areas is often referred to as the *ventral stream* (Rauschecker & Scott, 2009; Hickok & Poeppel, 2007). Comprehension models accordingly predict that SEM should rely on ventral stream brain regions.

As noted above, sensorimotor models do not address how semantic errors are monitored. Because these models propose that comparison between expected and actual auditory perception occurs in the posterior superior temporal cortex, the clearest neuroanatomical prediction of these models is that SEM of phonological errors should rely on this region (Hickok, 2012; Tourville & Guenther, 2011). In HSFC, this prediction is tempered by the fact that the posterior superior temporal cortex also contains area Spt, which is proposed to generate perceptual expectations. Damage to area Spt therefore could distort perceptual expectations, which as explained above would increase SEM.

Although several models suggest that SEM of phonological errors relies on brain structures that are distinct from those of SEM of semantic errors, lesion–behavior studies have not investigated distinct lesions that reduce SEM of phonological errors versus SEM of semantic errors. It is important to identify such distinctions because they may aid in clinical prognosis and assessment, as well as inform our understanding of the mechanisms underlying SEM.

Present Study

Here, we investigated the behavioral and neural basis of SEM in aphasia. We specifically focused the basis of monitoring phonological versus semantic errors because the models outlined above make specific predictions for these different error types, but only with modest empirical support. Critically, all of the SEM models predict that speech and/or language deficits should impair SEM in aphasia, but only some predict that cognitive control deficits impair SEM in aphasia. Therefore, our first goal was to investigate speech and language processes that predict SEM performance in LH stroke survivors (Experiment 1).

Next, we investigated whether cognitive control deficits reduce SEM, while accounting for the speech and language abilities identified as important for SEM (Experiment 1). To that end, we employed several cognitive control tasks to disentangle specific cognitive control abilities important for SEM, including language-specific cognitive control measures and nonlinguistic cognitive control measures. We further examined whether cognitive control of phonological representations (i.e., phonological control) relates to phonological SEM, and whether cognitive control of semantic representations (i.e., semantic control) relates to semantic SEM, as predicted by some models.

Finally, we used support vector regression lesion symptom mapping (SVR-LSM) to determine whether there are distinct neural substrates for SEM of phonological errors versus SEM of semantic errors (Experiment 2). We expected these neural substrates would correspond to the behavioral predictors of SEM in aphasia. This Experiment 2 investigation included data from participants in Experiment 1, as well as prior studies in the laboratory, namely, Mandal and colleagues (2020) and McCall, Dickens, and colleagues (2022; see Experiment 2 Methods).

EXPERIMENT 1

Methods

Experiment 1 analyzed behavioral data from an ongoing cross-sectional study of individual language outcomes after stroke (R01DC014960).

Participants

Participants were native English speakers with no history of psychiatric or neurological disorders aside from stroke. In addition, participants were screened for hearing and

vision loss. This study was approved by Georgetown University's institutional review board, and all participants provided written informed consent.

Participants included 41 LH stroke survivors at least 3 months post stroke (Table 1). Prior work in this laboratory has found that sample sizes of approximately 50 participants are sufficiently powered for regression-based behavioral analyses (Fama, Henderson, et al., 2019). LH stroke survivors were only included if they committed at least five errors during confrontation picture naming. A cutoff of five errors has been adopted in prior studies of SEM (e.g., van der Stelt, Fama, McCall, Snider, & Turkeltaub, 2021; Mandal et al., 2020) to promote precision in the measurement of SEM abilities. For example, if a participant only commits one error, their SEM ability can only be measured as 0% success (did not detect the error) or 100% success (detected the error). Furthermore, if a participant commits two errors, their SEM abilities can be measured as either 0% success (detected neither error), 50% success (detected one of the two errors), or 100% success (detected both errors). Consequently, precision improves with more errors, and at five errors, SEM abilities can be measured as 0%, 20%, 40%, 60%, 80%, or 100%.

Measures

We collected scores of SEM, five speech and language measures, five cognitive control measures, and two personality and motivation measures. These measures are described in detail below.

SEM measurement. SEM was scored as detections of errors committed during picture naming. Participants named 120 black-and-white images presented one at a time on a computer. Images included two 30-item short forms of the Philadelphia Naming Test (Walker & Schwartz, 2012), as well as 60 items from an internally normed publicly available corpus (www.cognitiverecoverylab.com /researchers; Fama, Snider, et al., 2019).

Participants were instructed to give the best one-word name for each picture and were told that "what really counts is the first thing you say, but if you make a mistake you can try to fix it." Participants had up to 20 sec to name each picture and could choose to advance to the next item sooner. No feedback was provided. Item-wise performance measures included accuracy, the type of errors committed, and whether each error was detected.

Error scoring. Trained researchers coded picture naming errors on the basis of semantic and/or phonological relationships to the target. Errors were coded as having a semantic relationship with the target if the error has a relevant meaning to the target. Errors were coded as having a phonological relationship with the target if both the error and target share any of the following: an initial phoneme, a final phoneme, a phoneme at the same syllable and word position, a stressed vowel, or two phonemes (not counting unstressed vowels) at any word position. Semantic and phonological relationships to the target were considered orthogonally. That is, errors could have both semantic and phonological relationships to the target, have just one relationship (i.e., semantic but not phonological and vice, versa), or have neither a semantic nor phonological relationship to the target. Overall, 58.0% of errors had a phonological relationship to the target, 28.4% of errors had a semantic relationship to the target, and 24.6% had neither a semantic nor phonological relationship to the target. In addition, 11.0% of errors had both a semantic and phonological relationship to the target.

Error detection scoring. Participants had the opportunity to detect and correct their errors, but they were not explicitly directed to judge the accuracy of their response on each trial. Much of the prior research on error monitoring, particularly research on the relationship between error monitoring and clinical outcomes, uses this approach (McCall, Dickens, et al., 2022; van der Stelt et al., 2021; Mandal et al., 2020; Schwartz, Middleton,

Table 1. Demographics for Experiment 1

	LH Stroke Survivors
Total	41
Sex	27 M, 14 F
Age	$M = 59.2$, $SD = 11.4$ [40–84]
Education (years)	$M = 16.5$, $SD = 2.8$ [12–21]
Western Aphasia Battery (WAB) Aphasia Quotient	$M = 73.9$, $SD = 22.7$ [20.1–97.3]
Time since stroke (months)	$M = 48.6$, $SD = 48.3$ [3.3–184.1]
WAB aphasia diagnosis	21 Anomic, 10 Broca, 4 Conduction, 6 Transcortical Motor
Picture naming errors (out of total 120 items)	$M = 30.1$, $SD = 23.8$ [5–106]
Picture naming error detection rate	$M = 0.46$, $SD = 0.29$ [0–0.98]

M = mean; SD = standard deviation; [min–max] = range.

Brecher, Gagliardi, & Garvey, 2016; Nozari et al., 2011; Marshall, Neuburger, & Phillips, 1994). In Experiment 1, detections were scored by trained researchers, and afterward, detection scores were verified by a separate researcher to ensure consistency with the scoring protocol.

Any verbal statement indicating awareness of an error on a trial was recorded as a detection. For example, detection of an error would be recorded if a participant either verbally declared that their response is incorrect, for example, "garlic…no that's not it!" [Target: Onion] or produced a second attempt to repair their initial utterance, for example, "thermometer um uh raincoat" [Target: Umbrella]. In the absence of verbal statements indicating error awareness, nonverbal behaviors (grunts, facial expressions, etc.) were not recorded as detections.

Speech and language measures.   In total, five speech and language scores were measured in LH stroke survivors (s weights, p weights, Word–Picture Matching accuracy, Pseudoword Discrimination accuracy, and Apraxia of Speech Rating Scale 3; Table 2). Each measure was collected in all 41 participants, with the exception of Word–Picture Matching, which was collected in 40 participants. In addition, Word–Picture Matching accuracy was measured in 42 age- and education-matched neurotypical controls to establish a normal performance range. These speech and language scores were selected as prospective predictors of SEM because they met at least one of the following two criteria:

(1)   The score has been previously found to relate to SEM in aphasia. This criterion was met by s weights, p weights, and Word–Picture Matching accuracy (Mandal et al., 2020; Dean et al., 2017; Nozari et al., 2011).

(2)   The score has strong theoretical potential to relate to SEM in aphasia. This criterion was met by Pseudoword Discrimination accuracy, as well as Apraxia of Speech Rating Scale 3.

S weights and *p* weights. Performance on the picture naming task was used to evaluate phonological and semantic aspects of word retrieval in individual participants based on Dell's interactive two-step model of lexical access (Dell & O'Seaghdha, 1992). Here, Foygel and Dell's semantic-phonological computational model was used to estimate s weights representing the weight of semantic–lexical connections, and p weights representing the weight of lexical–phonological connections (Foygel & Dell, 2000). These weights were estimated via the web-based interface (available at https://langprod.cogsci.illinois.edu/cgi-bin /webfit.cgi; Dell, Martin, & Schwartz, 2007).

Word–Picture Matching (48 items). Participants were presented with an auditory word, and selected a target picture representing that word. The target picture was presented simultaneously in an array of five semantic foils. This task is a computerized version of the word–picture matching task described and normed in Rogers and Friedman (2008). Each target and foil were selected from a standardized corpus of black-and-white drawings (Rogers & Friedman, 2008; Snodgrass & Vanderwart, 1980). Accuracy scores on this measure correlated with participant's scores of Western Aphasia Battery–Revised (WAB; from Kertesz, 2007) subtests of speech comprehension (WAB Auditory Word Recognition $r = .62$, $p = .000018$; WAB Yes No Questions $r = .56$, $p = .00015$; WAB Sequential Commands $r = .64$, $p = .000011$).

Performance was scored as percent accuracy. As there was an apparent ceiling effect and a highly skewed distribution in scores, LH stroke survivors' accuracy scores were categorized on the binary basis of whether they were impaired relative to neurotypical performance. Impairment was defined as performance worse than the entire range of neurotypical scores. Thus, the cutoff for impairment was the minimum neurotypical score, which was 94% accuracy.

Pseudoword discrimination task (46 items). Participants indicated via button-press if two pseudowords are the same or different in a task adapted from the Temple Assessment of Language and Short-term Memory in Aphasia (TALSA; Martin, Minkina, Kohen, & Kalinyak-Fliszar, 2018). Performance was scored as percent accuracy. Accuracy on this task has theoretical potential to relate to SEM because the task may require abilities critical for evaluating phonological errors against their intended target. Although pseudoword discrimination was not measured in neurotypical control participants in the present study, in-house norms were available via a prior study's scores in 20 neurotypical participants (Fama, Snider, et al., 2019). In the present study, this measure correlated with participant scores of WAB subtests

Table 2. Performance on Speech and Language Measures

Measure	LH Stroke Survivors	Percent Impaired
Word to Picture Matching	$M = 85\%$, $SD = 20\%$ [13–100%]	45.0%
S Weights	$M = 0.01$, $SD = 0.03$ [0–0.10]	NA
P Weights	$M = 0.02$, $SD = 0.02$ [0–0.10]	NA
Pseudoword Discrimination	$M = 89\%$, $SD = 11\%$ [48–98%]	36.6%
Apraxia of Speech Rating Scale 3	$M = 9.66$, $SD = 8.81$ [0–30]	NA

Percent impaired column indicates the proportion of participants categorized as impaired at either Word to Picture Matching, or Pseudoword Discrimination. M = mean; SD = standard deviation; [min–max] = range.

of speech comprehension (WAB Auditory Word Recognition $r = .61$, $p = .000027$; WAB Yes No Questions $r = .50$, $p = .00094$; WAB Sequential Commands $r = .46$, $p = .0023$).

Similar to scores on the word–picture matching task, the scores on the pseudoword discrimination task had an apparent ceiling effect and skewed distribution, so they were binarized using the same approach described above. The cutoff for impairment was the minimum neurotypical control score, which was 89% accuracy.

Apraxia of Speech Rating Scale 3. Apraxia of speech scores such as the Apraxia of Speech Rating Scale 3 (ASRS3) have theoretical potential to relate to SEM because deficits in motor speech planning and execution may contribute to the incidence of phonological speech errors. Lesion–symptom mapping evidence supports the contribution of motor speech brain areas to the incidence of phonological errors during picture naming in people with aphasia (Schwartz et al., 2012). Furthermore, motor-speech somatosensory (e.g., proprioceptive) feedback is thought to support online correction of the articulation of phonemes (Hickok, 2012; Guenther, 1994). To account for the potential role of motor speech deficits in SEM, we included the ASRS3 (Wambaugh, Bailey, Mauszycki, & Bunker, 2019). Performance was scored as the sum total across all 16 ASRS3 item scores. Each item-level ASRS3 score is based on a 5-point scale describing the presence and severity of motor speech characteristics as identified by the examining clinician, a trained speech-language pathologist. For each item, a higher score indicates greater severity. Therefore, a higher sum total of all scores suggests a more severe motor speech deficit. Table 2 includes the raw ASRS3 scores. Before statistical modelling, these scores were rescaled by mean-centering and dividing by the standard deviation.

Cognitive control measures. Five cognitive control scores were measured in LH stroke survivors (i.e., Simon cost, no-go accuracy, task switching cost, phonological control, semantic control; Table 3). Three of these scores were selected because they measure cognitive control of processes that do not critically rely on language (Simon cost, no-go accuracy, task switching cost). The final two scores were selected because they measure cognitive control of language-based processes (semantic control, phonological control).

Table 3. Performance on Cognitive Control Measures

Measure	LH Stroke Survivors
Phonological Cost	$M = 0.25$, $SD = 0.16$ [0.04–0.98]
Semantic Cost	$M = 0.39$, $SD = 0.23$ [0.03–1.21]
Switching Cost	$M = 0.34$, $SD = 0.12$ [0.11–0.64]
Simon Cost	$M = 0.10$, $SD = 0.10$ [−0.03–0.40]
No Go Accuracy	$M = 0.85$, $SD = 0.22$ [0–1.00]

M = mean; SD = standard deviation; [min–max] = range.

Simon task (100 items: 50 congruent, 50 incongruent). Participants must select between two alternative colored (red vs. blue) blocks in trials that are either congruent or incongruent on dimensions of color and location (left vs. right). Performance in each condition is measured as the inverse efficiency (mean reaction time divided by accuracy; Townsend & Ashby, 1983). The behavioral index, cost, is measured as the difference in inverse efficiency between the incongruent and congruent conditions. Cost on the Simon task is thought to reflect abilities in conflict monitoring and resolution (Cespón, Hommel, Korsch, & Galashan, 2020).

Go/no-go task (100 items: 82 go, 18 no-go). Participants are presented with shapes one at a time and press a button when they see a circle (the go trials), but they do not press the button when they see any other shape (the no-go trials). The task consists of 100 total trials: 82 go trials and 18 no-go trials. Performance is scored as percent accuracy on the no-go trials. Accuracy on no-go trials reflects inhibitory control ability, as those trials require the participant to refrain from pressing the button.

Task switching. Participants perform the nonverbal version of the Antelopes and Cantaloupes task paradigm, which is briefly described below (see McCall, van der Stelt, et al., 2022, for an in-depth description of task and summary measures). This paradigm was completed by 40 participants. Participants select the picture of a target object in a 2 × 2 array of four pictures on a touch screen as many times as possible within seven 20-sec blocks. To minimize the use of inner speech-based self-cueing strategies in this nonverbal version of the task, the four pictures in the 2 × 2 array are abstract shapes that have no canonical name. These blocks require selecting a single target repeatedly (S1), two separate targets successively (S2), or three separate targets successively (S3). Performance is calculated as the block duration divided by number of successful selections and is referred to as time per target selection (TTS). Switching cost is calculated as average TTS across S2 and S3 blocks, minus TTS in the S1 blocks. A lower switching cost reflects greater ability in task switching.

Phonological control task. Participants perform two versions of the Antelopes and Cantaloupes paradigm. In the phonological version, the four pictures in the array are phonologically related to one another. In the baseline version, the four pictures in the array are neither phonologically nor semantically related to one another. Phonological cost is measured as the difference between the TTS in the phonological version and the TTS in the baseline version.

Semantic control task. Participants perform two versions of the Antelopes and Cantaloupes paradigm. In the semantic version, the four pictures in the array are semantically related to one another. The baseline version is the same as in the phonological control task. Semantic cost is

measured as the difference between the TTS of the semantic version and the TTS in the baseline version (McCall, van der Stelt, et al., 2022).

Personality and motivation measures. To measure the extent to which participants are generally motivated to monitor their errors across task domains, we scored personality and motivation in two ways: (1) We utilized a self-report personality measure, and (2) we examined spontaneous error monitoring behavior on a nonlanguage task that was designed to elicit errors.

Short almost perfect scale. As a personality measure, we utilized the Short Almost Perfect Scale (SAPS; Rice, Richardson, & Tueller, 2014), which yields two subscale scores: Standards and Discrepancy. Standards is thought to reflect positive aspects of perfectionism related to conscientiousness, whereas Discrepancy is thought to reflect negative aspects of perfectionism related to neuroticism (Rice et al., 2014). This measure was administered in 40 participants.

Picture copy. As a measure of motivation to spontaneously self-monitor one's own errors, we designed a picture copy drawing task. In this task, participants used a touchscreen computer to recreate 10 trials of black-and-white line drawings composed of either one geometric shape or two concentric geometric shapes. To elicit increased errors in the drawings, during the final five trials, the pixels drawn on the screen were offset diagonally from the location touched on the touchscreen. All trials were presented with clear and erase buttons for participants to correct their drawing. Error monitoring behavior, called picture copy corrections, was scored as the total selections of the clear and erase buttons across all 10 trials. This measure was administered in 40 participants.

Experimental Design and Statistical Analyses

Motivational influences on SEM. We first set out to validate our measure of SEM. An oft-noted downside of the present measure of SEM is that it may be influenced by motivation or personality, because participants are not explicitly instructed to judge their accuracy on each trial. Consequently, it is possible that participants detect their errors but choose not to overtly display monitoring behavior. To investigate this possibility, we examined the relationship between SEM and scores on a personality scale as well as a spontaneous error monitoring task. This relationship was modeled using a mixed-effects logistic regression model examining error trials, with the dependent variable as a binary outcome of whether the error was detected, and the independent variable as fixed-effect measures of personality (i.e., SAPS Standards, and SAPS Discrepancy) and motivation (i.e., picture copy corrections). This analysis was run on 39 participants instead of 41, because SAPS scores were unavailable in two

participants. Subject ID was included as a random effect to model participant-wise random intercepts. Because this model only considered error trials, which could occur on any of the 120 items, there was insufficient sampling to consider item as a random effect.

Behavioral predictors of SEM in aphasia. We set out to examine behavioral predictors of SEM while accounting for differences between SEM of phonological errors versus SEM of semantic errors.

This examination involved two rounds of analysis (1) identifying speech and language measures that predict SEM, and (2) identifying cognitive control scores that predict SEM while accounting for the measures from (1).

Speech and language measures that predict SEM. To identify speech and language measures that predict SEM, we entered candidate speech and language measures into a statistical model of SEM and eliminated measures that did not inform the model. To model SEM, we designed a mixed-effects logistic regression examining error trials with the dependent variable set to a binary outcome of whether the error was detected. Trials with no response were not included in this model. The independent variables included subject-level fixed effects of scores from the speech and language measures (Table 4). To account for differences between SEM of phonological errors versus SEM of semantic errors, the phonological and semantic relatedness of the error to the target were included as separate item-level fixed effects and in interactions with select scores. These item-level fixed effects and interactions are listed in Table 4.

Akaike Information Criterion (AIC) values were used in the backward elimination of speech and language measures that were not informative in predicting SEM (Burnham & Anderson, 2004; Akaike, 1974). Starting with a model containing the full initial set of candidate predictors, AIC values were used to identify fixed terms to remove in the following manner. First, the AIC value was calculated for all the possible models that would result from dropping one of the fixed effect terms from the model (R function *drop1*). For example, if a model has five fixed effect terms, *drop1* calculates the AIC of the five possible models that would result from dropping one of those five terms. Next, of those possible models, we selected the model with the lowest AIC value, removing one fixed effect term. The selected model was then entered into *drop1* to proceed to selection of the next model. Iterations of *drop1*-based term removal continued until the model entered into *drop1* had an AIC at least two points below all possible models with one removed term. The two-point difference was selected as the cutoff because a difference of two points in AIC values is considered substantial support for the model with the lower AIC value (Burnham & Anderson, 2004). Thus, selection of a model without removed terms required substantial support, promoting parsimony of the final model. We aimed to minimize the

Table 4. Initial Set of Candidate Speech and Language Predictors of SEM

Effects	Model Terms
Main effects (fixed effects):	Semantic Error, Phonological Error, Word to Picture Matching, s weights, p weights, pseudoword discrimination, Apraxia of Speech Rating Scale 3 (ASRS3)
Interactions (fixed effects):	Semantic Error * Word to Picture Matching
	Semantic Error * s weights
	Phonological Error * p weights
	Phonological Error * ASRS3
	Phonological Error * Pseudoword Discrimination
Random effects	Subject code (random intercepts)

overall number of final predictors because our sample size is relatively modest ($n = 41$).

It can be tempting to interpret the elimination of model terms as indicating that the eliminated terms have no relationship with SEM. However, in the present model selection approach, terms might be eliminated if there is simply insufficient power to determine their relationship with SEM. To discern whether terms are eliminated because of a lack of a relationship versus a lack of power, a Bayes factor for each removed term was estimated based on each model's Bayesian Information Criterion before and after eliminating the term (Masson, 2011). Each Bayes factor was also used to compute a p value representing the posterior probability that the term does not relate to SEM (Masson, 2011).

Cognitive control scores that predict SEM. To identify cognitive control scores that relate to SEM, we started with the surviving model of speech and language predictors described above. On top of this model, we added cognitive control scores as fixed effect predictors (Table 5). To account for differences between SEM of phonological versus semantic errors, the phonological and semantic relatedness of the error to the target was included as separate item-level predictors and in interactions with select scores. The model terms for these candidate cognitive control predictors are listed in Table 5. Backward elimination based on model fit as described above allowed for backward elimination of cognitive control scores that did not predict SEM. Bayes factors for the elimination of terms were also calculated as described above.

Implementation of statistical analyses. Mixed-effects models were run using R Version 4.1.0 (https://cran.r-project.org/). Logistic mixed-effects models were estimated using the R function *glmer,* with default settings, except for: *family = binomial, control = glmerControl(optimizer = "bobyqa")*. Model summary tables (Table 6, 7, and 8) were generated using the R function *tab_model*. All other analyses were run in MATLAB R2018b (MathWorks, https://www.mathworks.com/products/matlab.html).

Table 5. Initial Set of Candidate Cognitive Control Predictors of SEM

Effects	Model Terms
Main effects (fixed effects):	Semantic Error, Phonological Error, Phonological Control, Semantic Control, Simon Cost, No-go Accuracy, Switching
Interactions (fixed effects):	Semantic Error * Phonological Control
	Semantic Error * Semantic Control
	Phonological Error * Phonological Control
	Phonological Error * Semantic Control
Random effects	Subject code (random intercepts)

Table 6. Logistic Mixed-Effects Regression of Motivational Influences on SEM

Predictors	Error Detection		
	Odds Ratios	CI	p
(Intercept)	0.86	0.50–1.47	.58
SAPS discrepancy	1.21	0.70–2.10	.49
SAPS standards	1.25	0.66–2.39	.49
Picture copy corrections	1.31	0.65–2.60	.45
Random Effects			
σ^2	3.29		
τ_{00} Subject Code	2.55		
ICC	0.44		
$N_{Subject\ Code}$	39		
Observations	1124		
Marginal R^2/conditional R^2	0.015/0.446		

Results

Motivational Influences on SEM

First, we sought to determine whether personality and motivation play a role in our measure of SEM. To that end, we tested the relationship between SEM and measures on the SAPS and Picture Copy using a mixed-effects logistic regression. We hypothesized that SEM is driven by specific linguistic or cognitive abilities, rather than personality or motivation. Therefore, we did not expect to find a strong relationship between scores on the SAPS or picture copy and SEM. As expected, SEM did not significantly relate to scores on the SAPS (SAPS Discrepancy: $OR = 1.21$, CI [0.70, 2.10], $p = .49$; SAPS Standards: $OR = 1.25$, CI [0.66, 2.39] $p = .49$), nor picture copy ($OR = 1.31$, CI [0.65, 2.60], $p = .45$; Table 6).

Behavioral Predictors of SEM in Aphasia

Speech and language predictors of SEM. To identify speech and language predictors of SEM, we used stepwise backward elimination of terms in a mixed-effects logistic regression, starting with the initial set of terms listed in Table 4. In the final model, the error type was a significant factor such that phonologically related errors were less likely to be detected than non-phonologically related errors (i.e., errors that are semantically but not phonologically related to the target, and errors that are neither semantically nor phonologically related to the target; $p = .041$; Table 7). In addition, there was a significant interaction between Word–Picture Matching and SEM of semantic errors ($p = .006$), which indicates that having an impairment in Word–Picture Matching relates to reduced semantic SEM (Table 7, Figure 2). This relationship was expected because Word–Picture Matching has

Table 7. Final Model Identifying Speech and Language Predictors of SEM

Predictors	Odds Ratios	CI	p
(Intercept)	1.23	0.56–2.70	.598
Semantic error	0.63	0.37–1.07	.090
Phonological error	0.71	0.51–0.99	**.041**
Word–picture matching impairment	1.50	0.52–4.32	.453
ASRS3	0.68	0.36–1.28	.235
Semantic Error * Word–Picture Matching Impairment	0.38	0.19–0.76	**.006**
Phonological Error * ASRS3	0.70	0.50–0.96	**.027**
Random Effects			
σ^2	3.29		
τ_{00} Subject Code	2.32		
ICC	0.41		
$N_{Subject\ Code}$	40		
Observations	1245		
Marginal R^2/conditional R^2	0.095/0.469		

Table 8. Final Model Identifying Cognitive Control Predictors of SEM

Predictors	Odds Ratios	CI	p
(Intercept)	1.14	0.52–2.48	.747
Semantic error	0.65	0.38–1.11	.111
Phonological error	0.57	0.40–0.82	**.002**
Word–picture matching impairment	2.22	0.76–6.54	.146
ASRS3	0.50	0.26–0.96	**.036**
Semantic control (cost)	0.41	0.24–0.70	**.001**
Semantic Error * Word–Picture Matching Impairment	0.32	0.15–0.65	**.002**
Phonological Error * ASRS3	0.71	0.51–0.99	**.042**
Semantic Error * Semantic Control (Cost)	1.68	1.16–2.44	**.006**
Random Effects			
σ^2	3.29		
$\tau_{00 \ Subject \ Code}$	2.10		
ICC	0.39		
$N_{Subject \ Code}$	39		
Observations	1139		
Marginal R^2/conditional R^2	0.193/0.507		

previously been shown to relate to SEM (Dean et al., 2017) and specifically to semantic SEM (Mandal et al., 2020). There was also a significant interaction between ASRS3 and phonological errors ($p = .027$), which indicates that worse apraxia of speech relates to reduced detection of phonological errors (Table 7, Figure 3). This relationship between apraxia of speech and phonological SEM is in the opposite direction of what we would expect from sensori-motor models of SEM (see Introduction section; e.g., Hickok, 2012; Tourville & Guenther, 2011).

Notably, stepwise model selection eliminated three variables: s weights, p weights, and pseudoword discrimination. For each of these eliminated variables, we used Bayes factors to assess the evidence for an absence of a
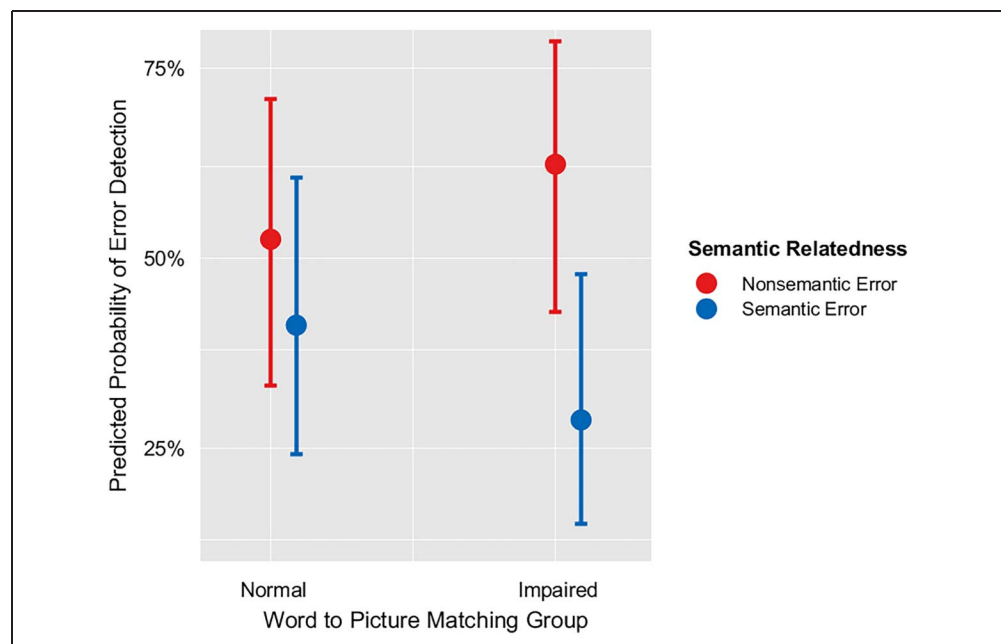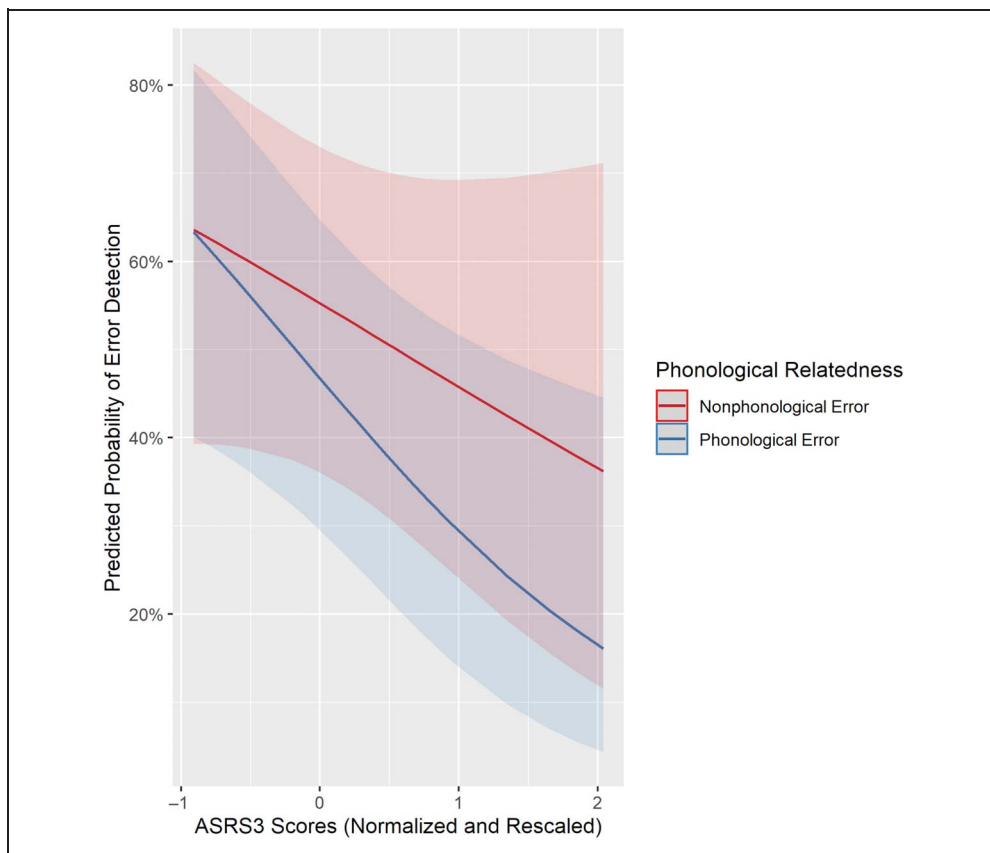

Figure 2. Interaction between word–picture matching and SEM of semantic errors.

Figure 3. Interaction between ASRS3 and SEM of phonological errors. Higher ASRS3 scores indicate greater severity of apraxia of speech.

relationship with SEM. Bayesian analysis on the elimination of s weights during model selection supported an absence of a relationship between s weights and SEM (Bayes factor = 0.028; $p = .028$). On the other hand, Bayesian analysis on the elimination of p weights did not strongly support an absence of a relationship between p weights and SEM (Bayes factor = 0.16; $p = .14$). Finally, Bayesian analysis on the elimination of pseudoword discrimination did support an absence of a relationship between pseudoword discrimination and SEM (Bayes factor = 0.073; $p = .068$).

Cognitive control predictors of SEM. Next, we assessed the relationship between cognitive control and SEM, building on the model above using speech and language predictors of SEM. In the final model, the error type was a significant factor such that phonologically related errors were less likely to be detected than non-phonologically related errors ($OR = 0.57$, CI [0.40, 0.82], $p = .002$; Table 8) . Because we hypothesized that cognitive control deficits impair SEM in aphasia, we expected at least one cognitive control measure would relate to SEM. Indeed, one cognitive control measure (i.e., the semantic control measure) survived backward elimination (Table 8), supporting that hypothesis. In addition, we hypothesized that if SEM relied on language-specific cognitive control, then phonological control and semantic control would relate to SEM. In support of that hypothesis, there was a significant main effect of semantic control, which indicates that worse semantic control (in the form of increased cost) was

related to reduced SEM ($OR = 0.41$, CI [0.24, 0.70], $p = .001$). However, we did not find evidence supporting a role of phonological control in SEM, because the phonological control measure was eliminated during stepwise selection. On the other hand, Bayes factor analysis of the removal of the phonological control measure did not strongly support an absence of a relationship between phonological control and SEM (Bayes factor = 0.16; $p = .14$). Therefore, we cannot strongly conclude that phonological control does not relate to SEM.

Furthermore, we predicted that phonological control would relate to detection of phonological errors and semantic control would relate to detection of semantic errors. Although there was a significant interaction between semantic control and SEM of semantic errors ($p = .006$), this finding does not support our prediction. Surprisingly, semantic control influenced SEM of nonsemantic errors to a greater extent than SEM of semantic errors (Table 8, Figure 4).

We also hypothesized that if domain-general cognitive control was used in SEM, then nonlanguage cognitive control scores would relate to SEM. That prediction was not supported because the nonlanguage cognitive control scores of no-go accuracy, Simon cost, and task switching were all eliminated during stepwise selection. Furthermore, Bayes factor analyses on the removal of Simon cost, and no-go accuracy during model selection supported an absence of a relationship between these factors and SEM (Simon cost: Bayes factor = 0.046; $p = .044$, no-go accuracy: Bayes factor =

Figure 4. Interaction between semantic control and SEM of semantic errors. Higher values of semantic control (cost) indicate *worse* semantic control.
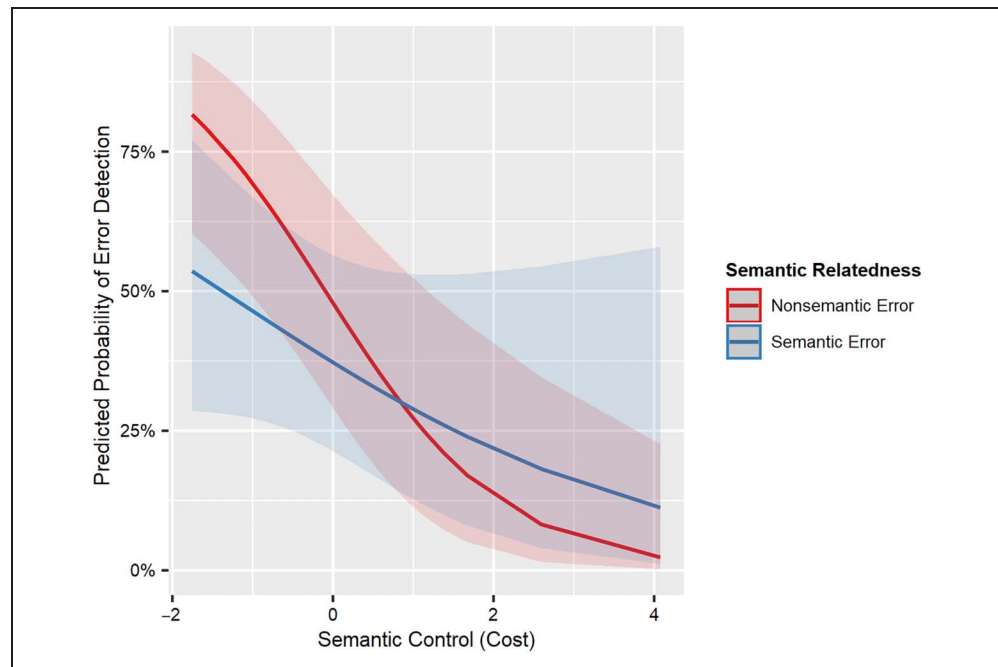
0.049; *p* = .046). However, Bayes factor analysis of the removal of task switching during model selection did not strongly support an absence of a relationship between task switching and SEM (Bayes factor = 0.15; *p* = .13).

Experiment 1 interim summary. The first experiment examined how speech, language, and cognitive control abilities relate to SEM of phonological and semantic errors in people with aphasia. As predicted by comprehension-based models, we found that worse auditory comprehension abilities were associated with reduced detection of semantic errors. We also found that worse motor speech deficits in the form of apraxia of speech were associated with reduced SEM of phonological errors. Although the association of motor deficits and phonological SEM is predicted by sensorimotor models, we found the direction of the association was the opposite of expected. Consistent with both production and comprehension models, we found that worse cognitive control abilities were associated with reduced SEM. As predicted by comprehension-based models, semantic control, a language-specific cognitive control ability was associated with SEM. Intriguingly, we found an interaction effect between semantic control and SEM of semantic errors that was opposite of what we expected. This interaction effect is further elaborated in the Discussion section. Next, Experiment 2 builds upon these findings by examining the brain structures critical for SEM of phonological and semantic errors.

EXPERIMENT 2

Methods

Experiment 2 investigated lesion–behavior relationships for SEM from participants across three studies on language

and stroke (R01DC014960, R03DC014310, Doris Duke Charitable Foundation 2012062).

Participants

As in Experiment 1, all participants were native speakers of English and did not have any psychiatric or neurological disorders aside from stroke. These participants were also screened for perceptual deficits such as hearing or vision loss. This study was approved by Georgetown University's Institutional Review Board, and all participants provided written informed consent.

Here in Experiment 2, we analyzed lesion–behavior relationships in a data set of 76 stroke survivors (Table 9; McCall, Dickens, et al., 2022; Mandal et al., 2020). Of these 76 participants, 30 were common to Experiment 1, and 43 were common to McCall, Dickens, and colleagues (2022), which examined a subset of Mandal and colleagues (2020). In addition, four participants intersected across all studies: Experiment 1, Experiment 2, and McCall, Dickens, and colleagues (2022). Power analyses of lesion–behavior investigations indicate that sample sizes of 60 or fewer can overestimate effect sizes, whereas sample sizes of 120 or greater can identify weak, trivial relationships (Lorca-Puls et al., 2018). All participants were at least 6 months poststroke. Participants were included based on the types of errors they committed during confrontation picture naming. Specifically, participants were included if they committed at least five semantic errors or at least five phonological errors. As described in Experiment 1, this cutoff has been adopted in prior studies of SEM to promote precision in the calculation of SEM abilities.

SEM measurement. In Experiment 2, the only behavioral measure analyzed was SEM. As in Experiment 1,

SEM was scored as detections of errors committed during picture naming. Picture naming task instructions were identical to those in Experiment 1. Here, participants named all or a subset (i.e., 60) of the 120 images from Experiment 1 presented on a computer. Because Experiment 2 used data pooled across multiple studies, several participants who were included in multiple studies completed multiple administrations of the picture naming task. For those participants, data from all of their administrations were combined for the present experiment.

Error scoring.   Errors were identified as belonging to standard categories using coding rules adapted from the Philadelphia Naming Test. This can be contrasted with Experiment 1, where errors were coded irrespective of whether they fit into a Philadelphia Naming Test category. Furthermore, a semantic error was considered an error that had a semantic but not a phonological relationship to the target word, whereas a phonological error was considered an error that had a phonological but not semantic relationship to the target word. Thus, errors that were both semantically and phonologically related to the target were not considered for analyses.

Error detection scoring.   Trained researchers scored error detection following the same procedure as in Experiment 1. Error detection was summarized across error types for each participant. Error detection scores for each error type equaled the proportion of detected errors to committed errors. For example, phonological error detection equaled the number of detected phonological errors divided by the number of committed phonological errors. Error detection was only calculated where a participant committed a minimum of five errors of the respective type. That is, phonological error detection was not calculated for participants with fewer than five phonological errors, and semantic error detection was not calculated for participants with fewer than five semantic errors.

Brain Imaging Measures

Brain imaging acquisition.   Participants from Experiment 2 underwent MRI scans including structural T1-weighted magnetization prepared rapid gradient echo (MPRAGE) scans and structural T2-weighted scans, which were collected using Siemens 3.0 Tesla scanners. T2-weighted images where available were used to assist in the manual segmentation of stroke lesions by a board-certified neurologist (author P.E.T). These segmented lesions were then normalized to the same template as the T1 images. Before analysis, all lesions were transformed into Montreal Neurological Institute (MNI) space, with a voxel size of 2.5 × 2.5 × 2.5 mm.

Scans collected in Experiment 2 followed the following scanning sequence.

T1-weighted MPRAGE. One hundred sixty sagittal slices, voxels =1 × 1 × 1 mm, matrix = 246 × 256, field of view = 250 mm, slice thickness = 1 mm, repetition time (TR) = 1900 msec, echo time (TE) = 2.56 msec.

T2-weighted. One hundred seventy-six sagittal slices, voxels = 0.625 × 0.625 × 1.25 mm, matrix = 384 × 384, field of view = 240 mm, slice thickness = 1.25 mm, TR = 3200 msec, variable TE and flip angle.

The scanning sequence for participants who were scanned as part of enrollment in Experiment 1 is listed below.

T1-weighted MPRAGE. 176 sagittal slices, voxels = 1 × 1 × 1 mm, matrix = 256 × 256, field of view = 256 mm, GeneRalized Autocalibrating Partially Parallel Acquisition (GRAPPA) = 2, slice thickness = 1 mm, TR = 1900 msec, TE = 2.98 msec.

T2-weighted (Fluid-Attenuated Inversion Recovery). One hundred ninety-two sagittal slices, voxels = 1 × 1 × 1 mm, matrix = 256 × 256, field of view = 256 mm, slice thickness = 1 mm, TR = 5000 msec, TE = 38.6 msec, TI = 1800 msec, flip angle = 120.

Experimental design and statistical analyses.   To determine the brain bases of phonological SEM and

Table 9. Demographics for Experiment 2

	LH Stroke Survivors
Total	76
Sex	47 M, 29 F
Age	$M = 60.11$, $SD = 9.45$ [43–84]
Education (years)	$M = 16.04$, $SD = 2.99$ [10–24]
Picture naming accuracy	$M = 0.57$, $SD = 0.26$ [0.08–0.98]
Picture naming error detection rate	$M = 0.41$, $SD = 0.26$ [0–0.97]
Time since stroke months	$M = 47.41$, $SD = 49.81$ [6.2–255.7]
Lesion volume (cm^3)	$M = 114$, $SD = 84$ [1–415]

M = mean; SD = standard deviation; [min–max] = range.

semantic SEM, we followed two SVR-LSM approaches (Zhang, Kimberg, Coslett, Schwartz, & Wang, 2014). First, we replicated the lesion–symptom mapping study of phonological SEM and semantic SEM in Mandal and colleagues (2020) on a cohort that includes participants from Experiment 1 in addition to participants from Mandal and colleagues (2020; Table 9). Second, we investigated lesions that selectively reduce phonological SEM more than semantic SEM, as well as the reverse, by running a variant of SVR-LSM referred to below as Dissociation SVR-LSM.

The first set of SVR-LSM analyses was performed using the SVR-LSM GUI (https://github.com/atdemarco/svrlsmgui; DeMarco & Turkeltaub, 2018). The Dissociation SVR-LSM analyses were performed in MATLAB using a customized version of the SVR-LSM GUI program. All lesion–symptom mapping analyses were conducted in MATLAB R2018b.

Standard SVR-LSM. The first SVR-LSM approach we took involved analyses based on Mandal and colleagues (2020),

serving as a replication of their study. These analyses are referred to here as "standard" SVR-LSM analyses to contrast with Dissociation SVR-LSM analyses described in the following section. The SVR-LSM GUI toolbox was used to run two standard SVR-LSM analyses investigating the effect of stroke lesions on SEM. One SVR-LSM analysis examined SEM of phonological errors, and the other analysis examined SEM of semantic errors. In both analyses, voxels were considered if they were lesioned in at least 10% of participants. Before modeling, lesion volume was regressed out of both the lesion images and the behavioral scores. Voxel-wise p values were derived via permutation testing using 10,000 permutations. Next, voxel-level multiple comparisons correction was applied using continuous family-wise error rate (CFWER) method, with an FWE rate of .05 and v set to 100 (Mirman et al., 2018). The same 10,000 permutations were also used for cluster-level correction with a p value threshold of .05.

Dissociation SVR-LSM. To determine how lesions in the brain differentially affect behaviors, we utilized an extension
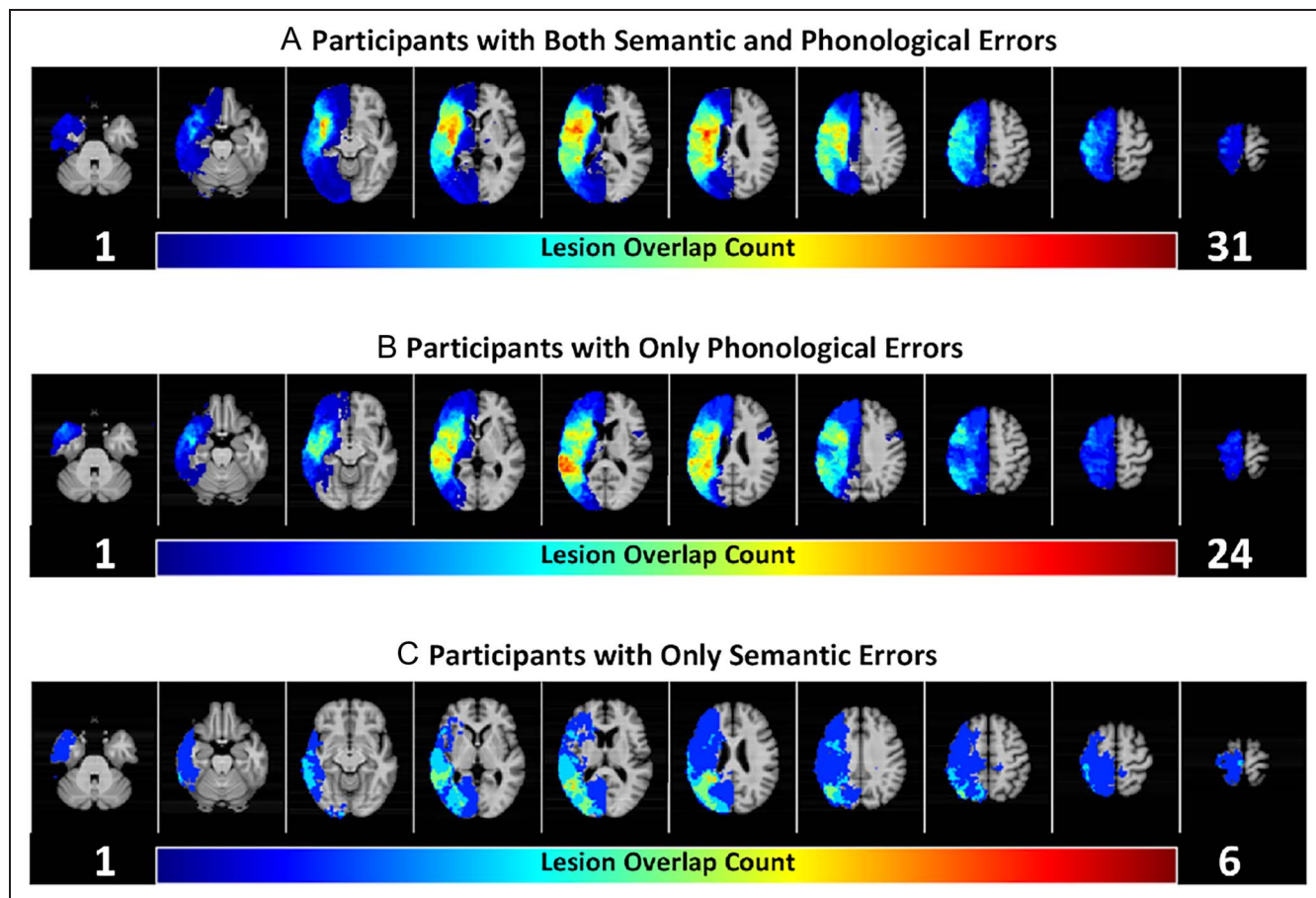

Figure 5. Lesion coverage for included participants. Participants are grouped by whether they committed sufficient (i.e., at least five) phonological and/or semantic errors, a requirement for inclusion in analyses of phonological SEM and/or semantic SEM. (A) Those with both sufficient phonological errors and sufficient semantic errors ($n = 36$). In other words, individuals in (A) committed at least five phonological errors and at least five semantic errors. (B) Those with sufficient phonological errors but not semantic errors ($n = 34$). (C) Those with sufficient semantic errors but not phonological errors ($n = 6$). Phonological SEM is analyzed in participants from (A) and (B), totaling to 70 participants. Semantic SEM is analyzed in participants from (A) and (C) totaling to 42 participants.

of SVR-LSM, here referred to as Dissociation SVR-LSM. Dissociation SVR-LSM is a whole-brain technique designed to map lesions that selectively impair one behavior more than another behavior. As in standard SVR-LSM, Dissociation SVR-LSM observes lesion–behavior relationships in the form of voxel-wise SVR-β values. Critically, Dissociation SVR-LSM observes lesion–behavior relationships for two behaviors consecutively, resulting in two sets of voxel-wise SVR-β values. Next, the observed difference in lesion–behavior relationships is calculated by subtracting the two sets of voxel-wise SVR-β values from one another. The difference value resulting from this subtraction is called the ΔSVR-β value. Finally, voxel-wise p values are calculated by comparing the observed ΔSVR-β values to a null distribution of ΔSVR-β values derived from permutations using shuffled behaviors. The two compared behaviors are identically shuffled on each permutation. In Dissociation SVR-LSM, both tails of the p values are interrogated, as each tail indicates the opposite selective impairment.

All other settings in the Dissociation SVR-LSM analyses were identical to the standard SVR-LSM analyses. That is, lesion volume was regressed from both voxels and behavior, voxels were only considered if they were lesioned in 10% of participants, voxel-wise p values were generated using 10,000 permutations, and multiple comparisons corrections on the vowel-wise values included CFWER $v = 100$ as well as cluster correction with FWER p threshold of .05.

Results

Lesion Symptom Mapping Analyses

Lesion symptom mapping analyses were used to investigate the brain basis of phonological SEM and semantic SEM. Lesion–behavior relationships for phonological SEM were investigated in participants who committed at least five phonological errors ($n = 70$). Lesion overlap maps revealed broad anatomical lesion coverage of the left middle cerebral artery territory for participants included in phonological SEM analyses (Figure 5). Similarly, lesion–behavior relationships for semantic SEM were investigated in participants who committed at least five semantic errors ($n = 42$). Lesion overlaps maps also revealed broad anatomical lesion coverage of the left middle cerebral artery territory for participants included in semantic SEM analyses (Figure 5A, C). Across the groups, participants had an average lesion volume of 114 cm^3 ($SD = 84$ cm^3). On average, participants detected 44% ($SD = 29\%$) of their phonological speech errors; and 34% ($SD = 24\%$), of their semantic speech errors.

Standard Support Vector Lesion Symptom Mapping of SEM

Lesion symptom map of phonological SEM. The SVR-LSM symptom map for phonological errors contained 230 surviving voxels, surpassing the CFWER threshold of $v = 100$ voxels (Figure 6). As the resultant SVR-LSM map
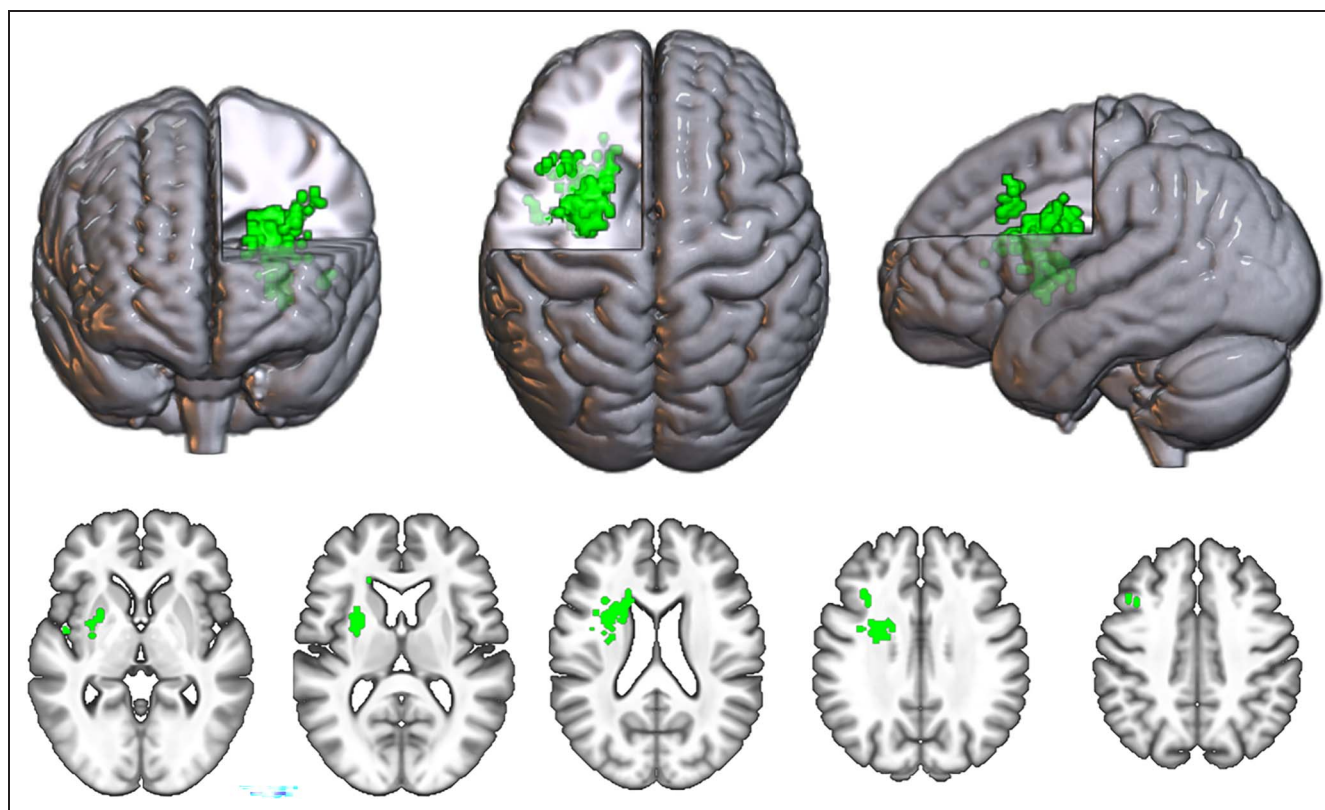

Figure 6. Standard SVR-LSM map for phonological SEM. All significant voxels at CFWER $v = 100$ are shown in green.

exceeded the threshold of 100 voxels, we reject the null hypothesis that there is no lesion–symptom association for phonological SEM.

The map contained one significant cluster at FWE < .05 that consisted of 199 voxels with a center of mass at MNI coordinates −27, 3, 20. As expected, this cluster was located in and around the frontal white matter, consistent with the findings in Mandal and colleagues (2020).

Lesion symptom map of semantic SEM. The lesion symptom map for semantic errors contained zero voxels that surpassed the CFWER threshold at $v = 100$ voxels.

Dissociation Lesion Symptom Mapping of SEM

Next, we used Dissociation SVR-LSM to determine whether lesions reveal a double-dissociation between phonological SEM and semantic SEM. On the basis of the results of Experiment 1, we expected lesions affecting brain regions associated with apraxia of speech (e.g., ventral motor and premotor cortex; see Basilakos, Rorden, Bonilha, Moser, & Fridriksson, 2015; Graff-Radford et al., 2014) to be more related to phonological SEM than semantic SEM. On the basis of the results of Experiment 1, we expected ventral stream lesions associated with auditory comprehension or semantic deficits to be related to semantic SEM more than phonological SEM.

Lesions that reduce phonological SEM more than semantic SEM. The Dissociation map for lesions that reduce phonological SEM more than semantic SEM contained 421 surviving voxels, surpassing the CFWER threshold of $v = 100$ voxels (Figure 7), which allows us to reject the null hypothesis that there is no lesion–symptom association. The map contained one significant cluster at FWE < .05, consisting of 414 voxels. This cluster was predominantly localized to the ventral precentral gyrus with a center of mass at MNI coordinates −49, 2, 19 (Figure 7).

Lesions that reduce semantic SEM more than phonological SEM. The Dissociation map for lesions that reduce semantic SEM more than phonological SEM contained 71 surviving voxels and so did not pass the CFWER threshold of 100 voxels (Figure 7), indicating there is a plausible chance (i.e., greater than 5%) that this result could be produced under the null hypothesis. The map did not contain any significant clusters at FWE < .05, but one cluster centered at MNI coordinates −31, −66, 8 approached significance (56 voxels, $p = .052$. This cluster was located posteriorly in the brain, within white matter beneath the temporo-parieto-occipital (TPO) junction.
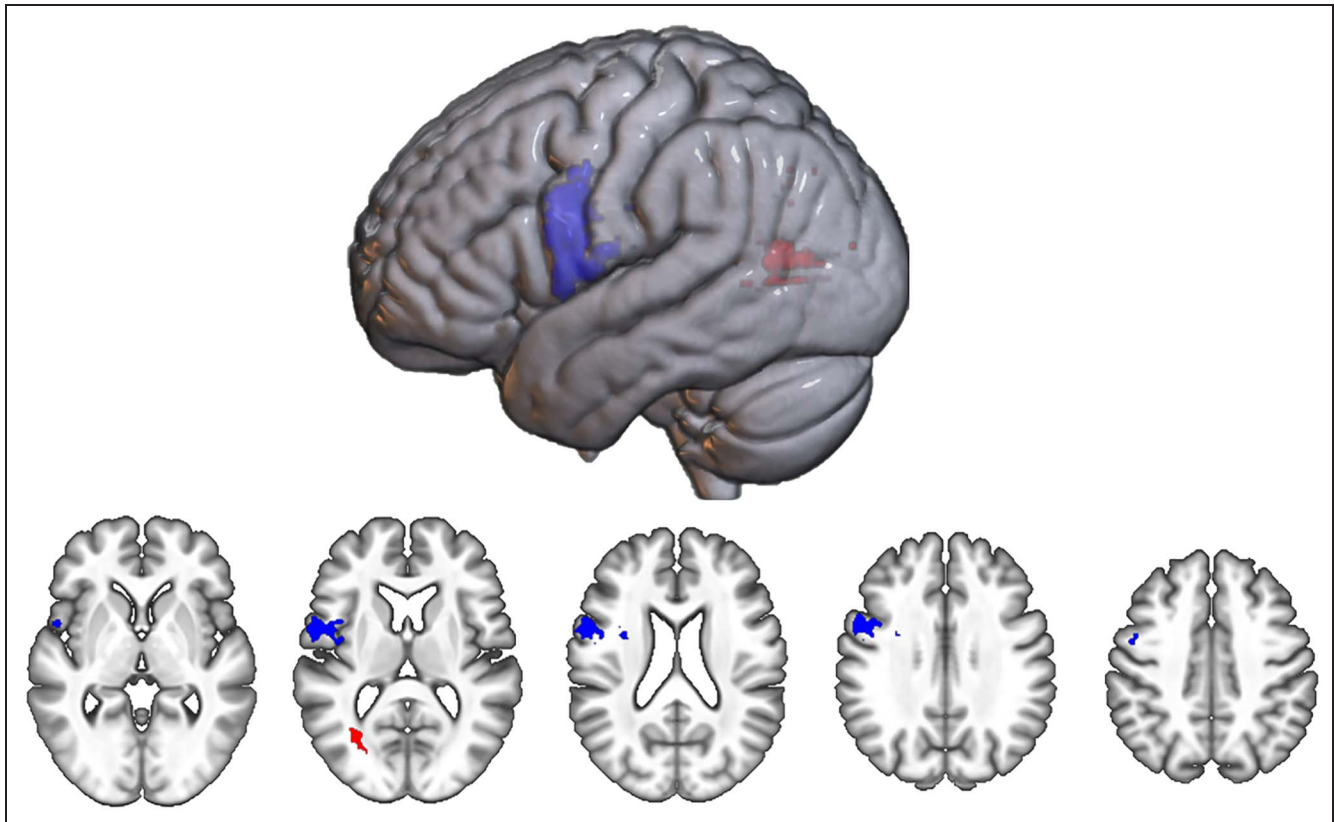

Figure 7. Dissociation SVR-LSM maps. Includes map for lesions that reduce phonological SEM more than semantic SEM (blue) and that reduce semantic SEM more than phonological SEM (red). All significant voxels at CFWER $v = 100$ are shown.

GENERAL DISCUSSION

Main Findings

We investigated critical behavioral and neural substrates supporting SEM of phonological and semantic errors in people with aphasia. Altogether, the findings indicate that SEM of phonological errors involves different neurocognitive mechanisms relative to SEM of semantic errors. Motor speech deficits as well as lesions to the ventral motor cortex, a region of the brain involved in motor speech (Conant, Bouchard, & Chang, 2014), selectively reduced SEM of phonological errors relative to SEM of semantic errors. Furthermore, deficits in auditory word comprehension related to reduced SEM of semantic errors. We then found a trending relationship between reduced semantic SEM and white matter lesions beneath the TPO junction. Finally, we found that semantic control deficits reduced SEM across all error types, supporting the proposition that cognitive control is a critical underlying ability for SEM.

Distinct Mechanisms for Monitoring Phonological Errors and Monitoring Semantic Errors

Critically, our neural and behavioral investigations differentiated SEM of phonological errors from SEM of semantic errors. In particular, it appeared that the motor cortex and motor abilities support SEM of phonological errors, whereas auditory comprehension abilities support SEM of semantic errors. Consequently, in line with proposals by various researchers (Nozari, 2020; Postma, 2000), we

conclude that there are multiple SEM mechanisms at play during speech. That is, throughout speaking, people monitor their motor planning for phonological errors (Figure 8, green), while also monitoring their speech output for semantic errors (Figure 8, orange). Importantly, we found evidence that there are mechanisms common to both SEM of semantic errors as well as SEM of phonological errors. One such common mechanism we identified was the language-specific cognitive control ability of semantic control, which supported SEM irrespective of the error type (Figure 8, Box 8B). In the following sections, we elaborate and propose finer details for these distinct and common mechanisms.

Phonological Error Monitoring Leverages Conflict Monitoring of Motor Speech

Although our findings suggest that motor speech is monitored for phonological errors, it remains unclear *how* motor speech is monitored. To date, mechanisms for monitoring motor speech have only been proposed by sensorimotor models. For example, as mentioned above, sensorimotor models in their simplest interpretation indicate that auditory perception of speech as well as somatosensory perception of speech are important for SEM. However, our Bayesian analysis supported an absence of a relationship between SEM and auditory speech perception. Notably, our measure of auditory speech perception, pseudoword discrimination, required the explicit comparative judgment of two pseudowords
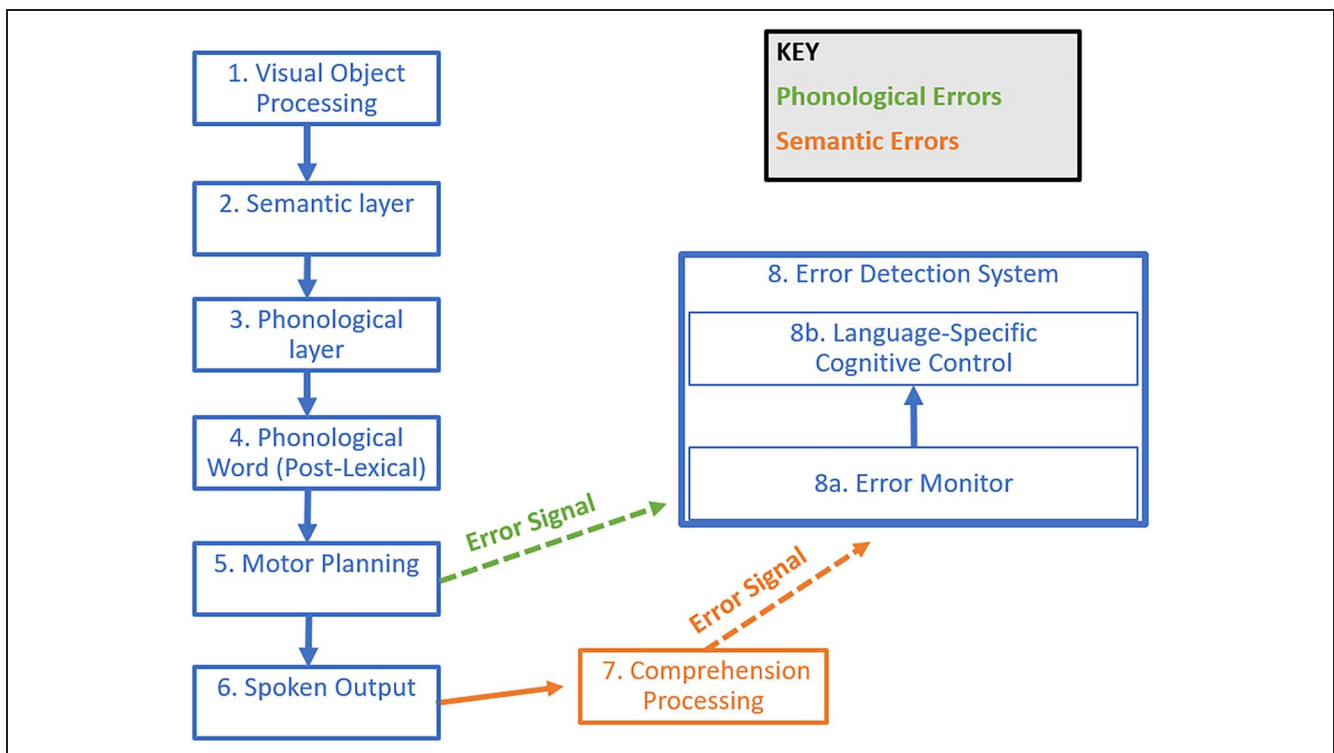

Figure 8. Hypothesized accounts for SEM of phonological errors (green) versus SEM of semantic errors (orange).

that differ on a single phoneme, and as such may not perfectly approximate the SEM mechanisms proposed by sensorimotor models. This measure might also rely on sensorimotor processing in addition to the sensory-perceptual processing described in sensorimotor models. Nevertheless, sensorimotor models would still predict a relationship between SEM and this measure to the extent that they both rely on phoneme perception. Moreover, the relationship we found between motor speech ability and phonological SEM was in the opposite direction from our sensorimotor model predictions. In summary, although we did not examine the contribution of somatosensory perception to SEM, our findings largely oppose sensorimotor model predictions of SEM.

One potential alternative mechanism we propose is a conflict-based monitoring account of motor speech. Such an account could utilize conflict monitoring in a similar fashion to the conflict-based account as well as Gauvin and Hartsuiker's Account. In support of a conflict-based monitoring account of motor speech, in McCall, Dickens, and colleagues (2022), we found that phonological SEM relies in part on connections between the ventral motor cortex and the pMFC (McCall, Dickens, et al., 2022), a region implicated in conflict monitoring as well as error detection (Yeung, 2014; Carter & van Veen, 2007). Along the same lines, in our noncomparative standard SVR-LSM analysis of phonological SEM, we replicated the finding from Mandal and colleagues (2020), in that frontal white matter damage reduced SEM. Such frontal white matter lesions may sever tracts that connect the pMFC to brain regions supporting speech and language. Therefore, we propose that a common error detection system monitors speech for errors (Figure 8, green), as opposed to the speech perception system proposed by sensorimotor models (Figure 1, purple).

It is important to identify the aspects of motor speech that might be monitored for conflict. One potential source of conflict could arise from competition between consecutive motor plans in a motor speech sequence. For example, a motor plan in progress might conflict with residual activation from a completed motor plan, or conflict with anticipatory activation for the next motor plan in the sequence. This conflict may be especially increased if neighboring items in a sequence share articulatory features (e.g., in tongue twisters). We refer to this type of motor speech conflict as *sequencing conflict*.

Another potential source of conflict in motor speech could arise at the stage of phonetic selection. Multiple competing nodes during phonetic selection could result from spreading activation (Dell, 1986), or perhaps other forms of interactivity (for a review of interactivity in word production, see Rapp & Goldrick, 2000). Then, the amount of competition between the activated nodes could be computed as conflict. We refer to this type of motor speech conflict as *phonetic selection conflict*. Unlike sequencing conflict, phonetic selection conflict would allow for detection of errors during the production of single-phoneme utterances irrespective of neighboring motor plans in a sequence.

Our finding that the ventral motor cortex supports phonological SEM is more consistent with phonetic selection conflict than sequencing conflict. Neither DIVA nor HSFC incorporate the ventral motor cortex in motor sequencing. Instead these models propose that motor sequencing relies on more anterior structures such as the IFG, SMA, and PreSMA, as well as subcortical structures including the basal ganglia (Miller & Guenther, 2021; Hickok, 2012). Moreover, DIVA proposes that the ventral motor cortex supports phonetic selection (Tourville & Guenther, 2011).

One way to differentiate whether SEM is based on sequencing conflict versus phonetic selection conflict would be through examining SEM of different subtypes of phonological errors. For example, we would expect that monitoring sequencing conflict would preferentially support SEM of phonological transposition errors (e.g., Spoonerisms), whereas monitoring phonetic selection conflict would preferentially support SEM of phonological distortion errors or SEM of phonological substitution errors.

Altogether, Experiment 2 indicates that the ventral motor cortex is only a small part of the neural story for SEM of phonological errors. This indication is evident from the minimal overlap between the standard lesion symptom map for phonological SEM (Figure 7) and the dissociation map for phonological SEM relative to semantic SEM (Figure 6). The exclusion of the ventral motor cortex from the standard lesion symptom map suggests that the ventral motor cortex is not the most important neural substrate for SEM of phonological errors.

Semantic Error Monitoring Leverages Auditory Comprehension

Consistent with comprehension-based accounts, this study confirmed that auditory word comprehension related to SEM of semantic errors, aligning with similar results from two other studies (Mandal et al., 2020; Dean et al., 2017). The present findings therefore suggest that SEM of semantic errors relies on comprehension processing (Figure 8, orange). In other words, while speaking, we listen to what we are saying to identify semantic speech errors. Interestingly, our findings disconfirmed the conflict-based account's prediction that poorer semantic encoding during lexical access would relate to worse semantic SEM. Although a prior study found this predicted relationship (Nozari et al., 2011), our Bayesian analysis found evidence against this predicted relationship. In summary, our findings confirm comprehension-based monitoring of semantic errors and refute production-based monitoring of semantic errors.

While the behavioral predictors of semantic SEM in this study were readily identified, the lesions that reduce SEM of semantic errors remain elusive. Similar to Mandal and

colleagues (2020), our standard lesion–symptom mapping analysis for SEM of semantic errors was not significant. However, the dissociation map for lesions that selectively reduce SEM of semantic errors relative to SEM of phonological errors was trending toward significance. The lesions in this trending dissociation map were in the white matter beneath the TPO junction, a location far away from the brain regions discussed in the introduction that are predicted by production-based models (e.g., anterior temporal cortex) as well as by speech comprehension-based models (e.g., posterior superior temporal cortex).

Interestingly, this white matter finding is near some gray matter regions described in ventral stream models that broadly describe language comprehension. In particular, regions of the angular gyrus have been proposed to process argument structure of words (Pylkkänen, 2019), combinatorial semantics (Friederici, 2017), as well as semantic-conceptual event knowledge (Matchin & Hickok, 2020). In the context of producing semantic errors, the angular gyrus is also implicated in thematic semantic processing (Schwartz et al., 2011). The role of the angular gyrus in language comprehension is often proposed in the context of syntactic processing, whereas angular gyrus is sometimes not included in ventral stream models that focus on single-word processing (e.g., Hickok & Poeppel, 2004). Syntactic processing is not relevant for the present study, which examined SEM in the context of spoken production of single word nouns. Comprehension functions ascribed to the angular gyrus aside from syntactic processing include semantic aspects of comprehension (Matchin & Hickok, 2020; Friederici, 2017), which is consistent with our finding that the underlying white matter is associated with SEM of semantic errors. Further investigation on this finding is therefore warranted to better understand the neural basis of semantic SEM. For example, a structural connectome-based lesion symptom mapping approach could reveal pairs of brain regions that are anatomically disconnected by this lesion, which might include ventral stream regions.

SEM across all Error Types Relies on Semantic Control

Consistent with production-based and comprehension-based models, we found that a cognitive control deficit related to reduced SEM in aphasia. Specifically, we found that worse semantic control related to poorer SEM across all error types (Figure 4). This relationship supports the comprehension-based model prediction that language-specific cognitive control abilities are important for SEM. Notably, semantic control has been theorized to rely on both domain-general as well as language-specific cognitive control processes (Lambon Ralph et al., 2017). Thus, one could argue that domain-general cognitive control abilities consistent with a conflict-based account contributed to the relationship between semantic control and SEM. However, we examined several cognitive control abilities that

do not rely on language (inhibitory control, conflict resolution, task switching), and did not find a relationship between any of them and SEM. Furthermore, Bayesian analyses supported the absence of relationships between inhibitory control and SEM, as well as between conflict resolution and SEM. Therefore, this study indicates that SEM is supported by semantic control in a language-specific manner. As such, it appears that language-specific semantic control is important for SEM of all error types, serving as an integral part of the error detection system (Figure 8, box 8B).

There are several possible accounts through which semantic control might support SEM. For instance, semantic control may contribute to the comparison between one's intended message and comprehension of one's own speech. In particular, this comparison process may benefit from controlled semantic retrieval, as well as maintenance and manipulation of retrieved semantic representations. Even nonsemantic errors might result in the activation of semantic representations that contrast or conflict with the target word. In such a case, semantic control could contribute to SEM of nonsemantic errors as well, which is consistent with our finding that semantic control was important for SEM across all errors. This relationship might also make sense if the semantic control measure actually reflected conflict monitoring of a broader array of language representations (e.g., semantics, phonetics, morphology). It is conceivable that the present measure of semantic control reflects semantic conflict monitoring because the measure is based on one's ability to ignore conflicting semantic distractors. If the measure were to also reflect conflict monitoring of phonological and phonetic representations, then it could account for SEM of multiple error types.

Intriguingly, the present study revealed a seemingly paradoxical relationship between semantic control and SEM of semantic errors. Although semantic control deficits decreased SEM of all error types (Table 8), we found that semantic control deficits decreased SEM of nonsemantic errors to a greater extent than SEM of semantic errors (Figure 4). We hypothesize that this paradoxical relationship reflects the fact that semantic control supports the correction of semantic errors internally before they are committed. As our measure of SEM is potentially biased against corrections of errors before they are committed (see Discussion: Limitations section), we may have observed a slightly lower SEM score in individuals who frequently correct semantic errors before they are committed. For example, if an individual with strong semantic control internally corrects several semantic errors before they are committed, then their semantic SEM ability will be underestimated. Consequently, this individual's overall SEM of semantic errors could be observed as paradoxically lower than their SEM for nonsemantic errors. Consistent with this interpretation, we previously found that LH stroke survivors with worse semantic control made more semantic errors during

picture naming than those with greater semantic control, in a study including most of the participants from Experiment 1 (McCall, van der Stelt, et al., 2022).

Clinical Considerations

The present findings demonstrate that SEM of phonological errors and SEM of semantic errors are differentially reduced by certain brain lesions and aphasic deficits. Therefore, considering SEM of phonological errors separately from SEM of semantic errors may add useful precision to clinical assessment. Furthermore, the distinction between SEM of phonological versus semantic errors may support the identification of therapeutic targets for improving SEM. For instance, because motor speech deficits reduced SEM of phonological errors, one might improve SEM of phonological errors by targeting motor speech abilities, rather than targeting error awareness (SEM) directly. Similarly, because auditory comprehension deficits reduced SEM of semantic errors, one might target auditory comprehension abilities to improve SEM of semantic errors. Another potential target for therapies is honing semantic control ability to improve SEM across all error types.

Considering the location of brain damage may also be informative for prognosis. Consistent with Mandal and colleagues (2020), we found frontal white matter lesions are associated with poorer phonological SEM. Knowledge of this association may help clinicians identify individuals at risk for poor phonological SEM early in their clinical course. Similarly, assessing damage to the ventral motor cortex may help identify those with disproportionately worse SEM of phonological errors relative to SEM of semantic errors. Identifying those at risk of poor SEM based on lesion location might motivate clinicians to thoroughly assess SEM impairments and to consider therapy targeting SEM.

Limitations

In this study, we measured SEM as spontaneous error detection during a confrontation naming task. One oft noted limitation of this measurement of SEM is that the measure depends on participants' overt acknowledgment that they made a mistake. This limitation is particularly salient when task instructions do not explicitly ask participants to identify their errors, as was the case in our study. Therefore, it is possible that participants detect some of their errors, but choose not to overtly acknowledge them. To investigate this possibility, we examined the relationship between SEM performance and motivational and personality measures. Here, we did not find that motivation or personality predicted SEM ability. Therefore, our findings do not indicate that this measure of SEM is confounded by one's general desire or motivation to correct oneself. Of course, the absence of proof does not disconfirm the potential relationship between SEM and one's

overall desire to overtly acknowledge errors. For example, SEM might still relate to other motivational measures or aspects of individual identity that we did not examine.

One aspect of motivation we did not examine is whether specific deficits impact one's motivation to acknowledge different error types. Therefore, it remains untested whether apraxia of speech related to SEM of phonological errors simply because people with apraxia of speech have decreased motivation to monitor phonological errors. Furthermore a deficit-specific motivation loss could occur with other speech and language deficits. For instance, those with semantic deficits affecting their comprehension could correspondingly have decreased motivation to detect semantic errors. This potential deficit-specific effect of motivation on error monitoring remains an open question in the field of SEM, and is important to address to strengthen interpretations of why particular deficits are associated with reduced SEM.

Motivational factors on SEM might also vary based on the context in which speech is produced. For example, during discourse, one may be more motivated to correct semantic errors, which can cause a greater disruption in transmitting the intended message. On the other hand, during the present picture naming task, participants might be more motivated to correct phonological errors, as the task goal is to produce a specific word form rather than convey an intended message.

We also acknowledge that our measure is perhaps biased against some covert aspects of SEM, such as corrections of errors before they are committed. For example, a participant might internally correct a speech error before speech is overtly produced, resulting in uninterrupted accurate speech with no overt acknowledgment of SEM. Notably, our measure is still able to observe some cases of SEM that occur before speech is overtly produced, as some participants verbalize their internal thought process during naming. For instance, some participants in the present study negated an incorrect response before producing it (e.g., "I know it's not garlic…" [Target: Onion]), and these would be counted as successful detections of an error.

This study was also limited in its ability to measure phonological aspects of lexical access in individuals with motor speech deficits. Here, we measured lexical access by calculating p weights, which are computed based on the error types that participants committed during picture naming (Dell et al., 2007). However, it is possible that the incidence of phonologically related naming errors can be artificially increased in participants who have apraxia of speech. Indeed, apraxia of speech can be difficult to distinguish from aphasia with phonemic paraphasias (e.g., Haley, Jacks, Richardson, & Wambaugh, 2017). In that way, motor speech deficits could have partially contributed to the calculation of p weights. It is therefore challenging to examine phonological aspects of lexical access in isolation in the presence of apraxia of speech. This limitation has important implications for our interpreting

the neural and behavioral findings in this study. For example, our lesion associations with SEM of phonological errors may in part reflect SEM of phonetic errors. In addition, we cannot strictly preclude the potential role of phonological aspects of lexical access, as measured by p weights, in SEM. Although the present behavioral analysis did not find a significant relationship between p weights and phonological SEM, such a relationship was identified in Nozari and colleagues (2011), where participants with moderate to severe apraxia were excluded.

Experiments 1 and 2 focused on divergent predictions between models of SEM, but were limited in their examination of common predictions between models. One step for future research is to investigate the role of shared processes between speech production and comprehension in SEM. For example, both production and comprehension models might predict that amodal semantic deficits could impair SEM of semantic errors. Future work on SEM could benefit by examining how representational deficits affect SEM.

Conclusions

This study provides evidence that SEM of phonological errors and SEM of semantic errors rely on distinct and shared neurocognitive mechanisms. Specifically, our findings suggest that SEM of phonological errors distinctly relies on motor speech processes, whereas SEM of semantic errors distinctly relies on auditory comprehension processes. Correspondingly, lesions to ventral motor cortex are more associated with reduced phonological SEM than semantic SEM. Across all error types, we found that SEM relies on semantic control abilities. These findings advance our understanding of the neurocognitive architecture of SEM and might suggest clinical avenues to improve identification of and treatment for SEM deficits in people with aphasia.

Reprint requests should be sent to Peter E. Turkeltaub, Center for Brain Plasticity and Recovery, Neurology Department, Georgetown University Medical Center, Washington, District of Columbia, USA, or via e-mail: turkeltp@georgetown.edu.

Data Availability Statement

Select data, materials, and/or analysis scripts may be made available on request to corresponding author.

Author Contributions

Joshua D. McCall: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Software; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Andrew T. DeMarco: Data curation; Methodology; Resources; Software; Writing—Review & editing. Ayan S. Mandal: Data curation; Investigation; Writing—Review & editing. Mackenzie E. Fama: Data curation; Investigation; Writing—Review & editing. Candace M. van der Stelt: Data curation; Investigation; Project administration; Supervision; Validation; Writing—Review & editing. Elizabeth H. Lacey: Data curation; Investigation; Validation. Alycia B. Laks: Data curation; Investigation; Writing—Review & editing. Sarah F. Snider: Data curation; Investigation; Supervision; Validation. Rhonda B. Friedman: Conceptualization; Supervision; Writing—Review & editing. Peter E. Turkeltaub: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing—Review & editing.

Funding Information

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .46; W/M = .286; M/W = .159; W/W = .095.

REFERENCES

Abel, S., Dressel, K., Kümmerer, D., Saur, D., Mader, I., Weiller, C., et al. (2009). Correct and erroneous picture naming responses in healthy subjects. *Neuroscience Letters*, *463*, 167–171. https://doi.org/10.1016/j.neulet.2009.07.077, PubMed: 19647038

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: A computational model of anterior cingulate and dorsolateral prefrontal cortex. *Neural Computation*, *27*, 2354–2410. https://doi.org/10.1162/NECO_a_00779, PubMed: 26378874

Basilakos, A., Rorden, C., Bonilha, L., Moser, D., & Fridriksson, J. (2015). Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke*, *46*, 1561–1566. https://doi.org/10.1161/STROKEAHA.115.009211, PubMed: 25908457

Behroozmand, R., Phillip, L., Johari, K., Bonilha, L., Rorden, C., Hickok, G., et al. (2018). Sensorimotor impairment of speech auditory feedback processing in aphasia. *Neuroimage*, *165*, 102–111. https://doi.org/10.1016/j.neuroimage.2017.10.014, PubMed: 29024793

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796. https://doi.org/10.1093/cercor/bhp055, PubMed: 19329570

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*, 1504–1529. https://doi.org/10.1162/jocn.2009.21306, PubMed: 19583476

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. https://doi.org/10.1037/0033-295X.108.3.624, PubMed: 11488380

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, *103*, 3153–3161. https://doi.org/10.1121/1.423073, PubMed: 9637026

Burnham, K. P., & Anderson, D. R. (Eds.). (2004). *Model selection and multimodel inference*. New York: Springer. https://doi.org/10.1007/b97636

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 367–379. https://doi.org/10.3758/cabn.7.4.367, PubMed: 18189010

Cespón, J., Hommel, B., Korsch, M., & Galashan, D. (2020). The neurocognitive underpinnings of the Simon effect: An integrative review of current research. *Cognitive, Affective, & Behavioral Neuroscience*, *20*, 1133–1172. https://doi.org/10.3758/s13415-020-00836-y, PubMed: 33025513

Conant, D., Bouchard, K. E., & Chang, E. F. (2014). Speech map in the human ventral sensory-motor cortex. *Current Opinion in Neurobiology*, *24*, 63–67. https://doi.org/10.1016/j.conb.2013.08.015, PubMed: 24492080

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*, 179–229. https://doi.org/10.1016/j.cognition.2002.07.001, PubMed: 15037130

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, *2*, 233. https://doi.org/10.3389/fpsyg.2011.00233, PubMed: 21954390

Dean, M. P., Della Sala, S., Beschin, N., & Cocchini, G. (2017). Anosognosia and self-correction of naming errors in aphasia. *Aphasiology*, *31*, 725–740. https://doi.org/10.1080/02687038.2016.1239014

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. https://doi.org/10.1037/0033-295X.93.3.283, PubMed: 3749399

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, *56*, 490–520. https://doi.org/10.1016/j.jml.2006.05.007, PubMed: 21085621

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314. https://doi.org/10.1016/0010-0277(92)90046-K, PubMed: 1582160

DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion–symptom mapping. *Human Brain Mapping*, *39*, 4169–4182. https://doi.org/10.1002/hbm.24289, PubMed: 29972618

Fama, M. E., Henderson, M. P., Snider, S. F., Hayward, W., Friedman, R. B., & Turkeltaub, P. E. (2019). Self-reported inner speech relates to phonological retrieval ability in people with aphasia. *Consciousness and Cognition*, *71*, 18–29. https://doi.org/10.1016/j.concog.2019.03.005, PubMed: 30921682

Fama, M. E., Snider, S. F., Henderson, M. P., Hayward, W., Friedman, R. B., & Turkeltaub, P. E. (2019). The subjective experience of inner speech in aphasia is a meaningful reflection of lexical retrieval. *Journal of Speech, Language, and Hearing Research*, *62*, 106–122. https://doi.org/10.1044/2018_JSLHR-L-18-0222, PubMed: 30950758

Fama, M. E., & Turkeltaub, P. E. (2020). Inner speech in aphasia: Current evidence, clinical implications, and future directions. *American Journal of Speech-Language Pathology*, *29*, 560–573. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0212, PubMed: 31518502

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

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity* (1st ed.). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262036924.001.0001

Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *Neuroimage*, *126*, 96–105. https://doi.org/10.1016/j.neuroimage.2015.11.037, PubMed: 26608243

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition*, *3*, 17. https://doi.org/10.5334/joc.81, PubMed: 32944680

Ghaleh, M., Lacey, E. H., Fama, M. E., Anbari, Z., DeMarco, A. T., & Turkeltaub, P. E. (2019). Dissociable mechanisms of verbal working memory revealed through multivariate lesion mapping. *Cerebral Cortex*, *30*, 2542–2554. https://doi.org/10.1093/cercor/bhz259, PubMed: 31701121

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*, 219–260. https://doi.org/10.1016/j.cognition.2005.12.010, PubMed: 16483561

Graff-Radford, J., Jones, D. T., Strand, E. A., Rabinstein, A. A., Duffy, J. R., & Josephs, K. A. (2014). The neuroanatomy of pure apraxia of speech in stroke. *Brain and Language*, *129*, 43–46. https://doi.org/10.1016/j.bandl.2014.01.004, PubMed: 24556336

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *72*, 43–53. https://doi.org/10.1007/bf00206237, PubMed: 7880914

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*, 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013, PubMed: 16887139

Haley, K. L., Jacks, A., Richardson, J. D., & Wambaugh, J. L. (2017). Perceptually salient sound distortions and apraxia of speech: A performance continuum. *American Journal of Speech-Language Pathology*, *26*, 631–640. https://doi.org/10.1044/2017_AJSLP-16-0103, PubMed: 28654944

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145. https://doi.org/10.1038/nrn3158, PubMed: 22218206

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99. https://doi.org/10.1016/j.cognition.2003.10.011, PubMed: 15037127

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402. https://doi.org/10.1038/nrn2113, PubMed: 17431404

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255, PubMed: 22016740

Johnson, L. P., Sangtian, S., Johari, K., Behroozmand, R., & Fridriksson, J. (2020). Slowed compensation responses to altered auditory feedback in post-stroke aphasia: Implications for speech sensorimotor integration. *Journal of Communication Disorders*, *88*, 106034. https://doi.org/10.1016/j.jcomdis.2020.106034, PubMed: 32919232

Kertesz, A. (2007). *WAB-R: Western Aphasia Battery–Revised*. PsychCorp. https://doi.org/10.1037/t15168-000

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42–55. https://doi.org/10.1038/nrn.2016.150, PubMed: 27881854

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104. https://doi.org/10.1016/0010-0277(83)90026-4, PubMed: 6685011

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75. https://doi.org/10.1017/S0140525X99001776, PubMed: 11301520

Lorca-Puls, D. L., Gajardo-Vidal, A., White, J., Seghier, M. L., Leff, A. P., Green, D. W., et al. (2018). The impact of sample size on the reproducibility of voxel-based lesion-deficit mappings. *Neuropsychologia*, *115*, 101–111. https://doi.org/10.1016/j.neuropsychologia.2018.03.014, PubMed: 29550526

Mandal, A. S., Fama, M. E., Skipper-Kallal, L. M., DeMarco, A. T., Lacey, E. H., & Turkeltaub, P. E. (2020). Brain structures and cognitive abilities important for the self-monitoring of speech errors. *Neurobiology of Language*, *1*, 319–338. https://doi.org/10.1162/nol_a_00015, PubMed: 34676371

Marshall, R. C., Neuburger, S. I., & Phillips, D. S. (1994). Verbal self-correction and improvement in treated aphasic clients. *Aphasiology*, *8*, 535–547. https://doi.org/10.1080/02687039408248680

Martin, N., Minkina, I., Kohen, F. P., & Kalinyak-Fliszar, M. (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neurolinguistics*, *48*, 199–225. https://doi.org/10.1016/j.jneuroling.2018.02.006, PubMed: 30220790

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679–690. https://doi.org/10.3758/s13428-010-0049-5, PubMed: 21302025

Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *30*, 1481–1498. https://doi.org/10.1093/cercor/bhz180, PubMed: 31670779

McCall, J. D., Dickens, J. V., Mandal, A. S., DeMarco, A. T., Fama, M. E., Lacey, E. H., et al. (2022). Structural disconnection of the posterior medial frontal cortex reduces speech error monitoring. *NeuroImage: Clinical*, *33*, 102934. https://doi.org/10.1016/j.nicl.2021.102934, PubMed: 34995870

McCall, J. D., van der Stelt, C. M., DeMarco, A., Dickens, J. V., Dvorak, E., Lacey, E., et al. (2022). Distinguishing semantic control and phonological control and their role in aphasic deficits: A task switching investigation. *Neuropsychologia*, *173*, 108302. https://doi.org/10.1016/j.neuropsychologia.2022.108302, PubMed: 35718138

Meekings, S., & Scott, S. K. (2021). Error in the superior temporal gyrus? A systematic review and activation likelihood estimation meta-analysis of speech production studies. *Journal of Cognitive Neuroscience*, *33*, 422–444. https://doi.org/10.1162/jocn_a_01661, PubMed: 33326327

Miller, H. E., & Guenther, F. H. (2021). Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework. *Aphasiology*, *35*, 424–441. https://doi.org/10.1080/02687038.2020.1765307, PubMed: 34108793

Mirman, D., Landrigan, J.-F., Kokolis, S., Verillo, S., Ferrara, C., & Pustina, D. (2018). Corrections for multiple comparisons in voxel-based lesion-symptom mapping. *Neuropsychologia*, *115*, 112–123. https://doi.org/10.1016/j.neuropsychologia.2017.08.025, PubMed: 28847712

Nozari, N. (2020). A comprehension- or a production-based monitor? Response to Roelofs (2020). *Journal of Cognition*, *3*, 19. https://doi.org/10.5334/joc.102, PubMed: 32944682

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*, 1–33. https://doi.org/10.1016/j.cogpsych.2011.05.001, PubMed: 21652015

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*, 97–132. https://doi.org/10.1016/S0010-0277(00)00090-1, PubMed: 10986364

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*, 62–66. https://doi.org/10.1126/science.aax0050, PubMed: 31604303

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499. https://doi.org/10.1037/0033-295X.107.3.460, PubMed: 10941277

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724. https://doi.org/10.1038/nn.2331, PubMed: 19471271

Rice, K. G., Richardson, C. M. E., & Tueller, S. (2014). The short form of the revised almost perfect scale. *Journal of Personality Assessment*, *96*, 368–379. https://doi.org/10.1080/00223891.2013.838172, PubMed: 24090303

Ridderinkhof, K. R. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447. https://doi.org/10.1126/science.1100301, PubMed: 15486290

Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, *111*, 561–572. https://doi.org/10.1037/0033-295X.111.2.561, PubMed: 15065924

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, *59*, 33–48. https://doi.org/10.1016/j.cortex.2014.07.001, PubMed: 25128898

Roelofs, A. (2020). Self-monitoring in speaking: In defense of a comprehension-based account. *Journal of Cognition*, *3*, 18. https://doi.org/10.5334/joc.61, PubMed: 32944681

Rogers, S. L., & Friedman, R. B. (2008). The underlying mechanisms of semantic memory loss in Alzheimer's disease and semantic dementia. *Neuropsychologia*, *46*, 12–21. https://doi.org/10.1016/j.neuropsychologia.2007.08.010, PubMed: 17897685

Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., et al. (2021). Cerebellar and cortical correlates of internal and external speech error monitoring. *Cerebral Cortex Communications*, *2*, tgab038. https://doi.org/10.1093/texcom/tgab038, PubMed: 34296182

Sangtian, S., Wang, Y., Fridriksson, J., & Behroozmand, R. (2021). Impairment of speech auditory feedback error detection and motor correction in post-stroke aphasia. *Journal of Communication Disorders*, 94, 106163. https://doi.org/10.1016/j.jcomdis.2021.106163, PubMed: 34768093

Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, 135, 3799–3814. https://doi.org/10.1093/brain/aws300, PubMed: 23171662

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., et al. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences, U.S.A.*, 108, 8520–8524. https://doi.org/10.1073/pnas.1014935108, PubMed: 21540329

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., et al. (2009). Anterior temporal involvement in semantic word retrieval: Voxel-based lesion–symptom mapping evidence from aphasia. *Brain*, 132, 3411–3427. https://doi.org/10.1093/brain/awp284, PubMed: 19942676

Schwartz, M. F., Middleton, E. L., Brecher, A., Gagliardi, M., & Garvey, K. (2016). Does naming accuracy improve through self-monitoring of errors? *Neuropsychologia*, 84, 272–281. https://doi.org/10.1016/j.neuropsychologia.2016.01.027, PubMed: 26863091

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215. https://doi.org/10.1037//0278-7393.6.2.174, PubMed: 7373248

Specht, K. (2014). Neuronal basis of speech comprehension. *Hearing Research*, 307, 121–135. https://doi.org/10.1016/j.heares.2013.09.011, PubMed: 24113115

van der Stelt, C. M., Fama, M. E., McCall, J. D., Snider, S. F., & Turkeltaub, P. E. (2021). Intellectual awareness of naming abilities in people with chronic post-stroke aphasia. *Neuropsychologia*, 160, 107961. https://doi.org/10.1016/j.neuropsychologia.2021.107961, PubMed: 34274379

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26, 952–981. https://doi.org/10.1080/01690960903498424, PubMed: 23667281

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge University Press.

Walker, G. M., & Schwartz, M. F. (2012). Short-form Philadelphia naming test: Rationale and empirical evaluation. *American Journal of Speech-Language Pathology*, 21, S140–S153. https://doi.org/10.1044/1058-0360(2012/11-0089), PubMed: 22294412

Wambaugh, J. L., Bailey, D. J., Mauszycki, S. C., & Bunker, L. D. (2019). Interrater reliability and concurrent validity for the apraxia of speech rating scale 3.0: Application with persons with acquired apraxia of speech and aphasia. *American Journal of Speech-Language Pathology*, 28, 895–904. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0099, PubMed: 31306600

Yeung, N. (2014). Conflict monitoring and cognitive control. In K. N. Ochsner & S. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience: The cutting edges* (Vol. 2, pp. 275–299). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199988709.013.0018

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion–symptom mapping using support vector regression. *Human Brain Mapping*, 35, 5861–5876. https://doi.org/10.1002/hbm.22590, PubMed: 25044213