

Research Article

The Subjective Experience of Inner Speech in Aphasia Is a Meaningful Reflection of Lexical Retrieval

Mackenzie E. Fama,^{a,b} Sarah F. Snider,^{b,c} Mary P. Henderson,^a William Hayward,^b
Rhonda B. Friedman,^{b,c} and Peter E. Turkeltaub^{a,b,c,d}

Purpose: Individuals with aphasia often report that they feel able to say words in their heads, regardless of speech output ability. Here, we examine whether these subjective reports of successful “inner speech” (IS) are meaningful and test the hypothesis that they reflect lexical retrieval.

Method: Participants were 53 individuals with chronic aphasia. During silent picture naming, participants reported whether or not they could say the name of each item inside their heads. Using the same items, they also completed 3 picture-based tasks that required phonological retrieval and 3 matched auditory tasks that did not. We compared participants’ performance on these tasks for items they reported being able to say internally versus those they reported being unable to say internally. Then, we examined the relationship of psycholinguistic word features to self-reported IS and spoken naming accuracy.

Results: Twenty-six participants reported successful IS on nearly all items, so they could not be included in the item-level

analyses. These individuals performed correspondingly better than the remaining participants on tasks requiring phonological retrieval, but not on most other language measures. In the remaining group ($n = 27$), IS reports related item-wise to performance on tasks requiring phonological retrieval, but not to matched control tasks. Additionally, IS reports were related to 3 word characteristics associated with lexical retrieval, but not to articulatory complexity; spoken naming accuracy related to all 4 word characteristics. Six participants demonstrated evidence of unreliable IS reporting; compared with the group, they also detected fewer errors in their spoken responses and showed more severe language impairments overall.

Conclusions: Self-reported IS is meaningful in many individuals with aphasia and reflects lexical phonological retrieval. These findings have potential implications for treatment planning in aphasia and for our understanding of IS in the general population.

Aphasia is a language disorder, acquired through stroke or other brain injury, that typically has chronic effects and a significant negative impact on long-term quality of life (Berthier, 2005; Engelter et al., 2006; Hilari et al., 2010). The specific language difficulties associated with aphasia can vary from person to person,

but a relatively universal deficit is anomia, an impairment of naming and word-finding (Laine & Martin, 2006; Maher & Raymer, 2004). Here, we are interested in a common clinical phenomenon in which some individuals with aphasia and anomia report that they can say words in their head that they cannot say out loud. This anecdotal sense of “inner speech” (IS) is supported by some objective prior work suggesting that IS can exceed overt speech abilities in people with aphasia (Fama, Hayward, Snider, Friedman, & Turkeltaub, 2017; Feinberg, Rothi, & Heilman, 1986; Geva, Bennett, Warburton, & Patterson, 2011; Geva, Jones, et al., 2011; Goodglass, Kaplan, Weintraub, & Ackerman, 1976; Hayward, Snider, Luta, Friedman, & Turkeltaub, 2016; Stark, Geva, & Warburton, 2017). In this study, we will examine the relationship between self-reported IS and objective measures of the mental process of naming in order to better understand the validity and meaning of the experience of IS in aphasia.

The phenomenon of IS has been intermittently studied in aphasia, beginning with a study that examined a related

^aCenter for Brain Plasticity and Recovery, Georgetown University Medical Center, Washington, DC

^bDepartment of Neurology, Georgetown University Medical Center, Washington, DC

^cCenter for Aphasia Rehabilitation and Research, Georgetown University Medical Center, Washington, DC

^dResearch Division, MedStar National Rehabilitation Hospital, Washington, DC

Correspondence to Mackenzie E. Fama: mackenziefama@gmail.com

Editor-in-Chief: Sean Redmond

Editor: Carl Coelho

Received June 3, 2018

Revision received July 26, 2018

Accepted August 17, 2018

https://doi.org/10.1044/2018_JSLHR-L-18-0222

Disclosure: The authors have declared that no competing interests existed at the time of publication.

experience, tip-of-the-tongue (Goodglass et al., 1976). Tip-of-the-tongue is a phenomenon in which an individual feels close to retrieving a target word, which is different from a sense of successful IS (sIS) as we have defined it: the feeling of being able to say a word correctly in one's head. Despite this important difference, this particular study on tip-of-the-tongue in aphasia warrants consideration because it is the only prior study (other than our own) that examines participants' subjective perceptions of performance on a silent picture-naming task. In Goodglass et al. (1976), patients with aphasia completed a picture-naming task and, when unable to name target words aloud, were asked whether they had an "idea of the word," followed by objective tests for phonological knowledge: first-letter identification (ID), syllable counting, and, finally, target word ID from multiple choices. Results showed that individuals with Broca's and conduction aphasia both demonstrated relatively intact phonological knowledge along with their frequent tip-of-the-tongue experiences (Goodglass et al., 1976). These findings were confirmed by a later study, which used objective measures such as picture-based rhyme and homophone judgments as a proxy for IS and showed that performance was preserved in four of five individuals with conduction aphasia, despite general difficulties in spoken naming (Feinberg et al., 1986). More recently, studies have used written word-based rhyme and homophone judgments to demonstrate relatively preserved IS in individuals with production deficits consistent with either conduction aphasia or motor planning impairments (Geva, Bennett, et al., 2011; Stark et al., 2017).

Although these studies have consistently demonstrated relatively preserved IS alongside speech production deficits, several questions remain unanswered. These prior studies have drawn conclusions about discrepancies between inner and overt speech without matching the assessments utilized, using different task designs and/or different stimuli across the inner/overt tasks. The consistency of the relationship between inner and overt speech could be more closely examined by comparing matched tasks using identical stimuli. Furthermore, many of these studies have defined IS as the ability to perform objective tasks, such as silent rhyme or homophone judgment, based on either pictures or written words (Geva, Bennett, et al., 2011; Geva, Jones, et al., 2011; Stark et al., 2017). An alternative approach to the study of IS is to ask language users directly about their own experience of it. In fact, this method is already employed in the study of IS in healthy language users, where questionnaires and other subjective approaches are used to elicit self-reports of the experience (Alderson-Day & Fernyhough, 2015; Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016; Morin, Uttl, & Hamper, 2011). Moreover, many individuals with aphasia spontaneously provide evidence for some level of metalinguistic awareness, making comments such as "I know it, but I can't say it" (Blanken, Dittmann, Haas, & Wallesch, 1987; Martin & Dell, 2007), strongly suggesting that their internal experiences may provide a novel, rich source of information about IS.

In our work, we address a question that is fundamentally different from prior studies on IS in aphasia, asking whether the subjective experience of IS commonly reported by individuals with aphasia relates to objective evidence of word retrieval and production (Fama et al., 2017; Hayward et al., 2016). Here, we define the subjective experience of sIS as the sense of being able to accurately say a word in one's head, with all the right sounds in the right order. Using this approach in the context of naming, we have previously demonstrated in two individuals with aphasia that self-report of sIS at the item level relates to subsequent success of spoken naming or, in the event of incorrect naming, the likelihood of the error being phonologically related to the target word (Hayward et al., 2016). These findings were replicated in a group of six participants, in which self-reported sIS was again related to evidence of phonological knowledge (Hayward, 2016). In a larger participant group, we have previously shown that the general experience of sIS followed by overt anomia is common in aphasia, distinct from other anomia experiences (e.g., a vaguer sense of "knowing it"), and associated with lesions primarily in the ventral sensorimotor cortex, a brain region that supports speech output processes (Fama et al., 2017).

In the current study, we aim to further elucidate the validity of self-reported sIS in people with aphasia by comparing subjective reports of the experience with objective behavioral measures, in order to learn more about the subjective experience of IS and its potential implications for understanding anomia. We frame our hypotheses in the context of processing models of naming, which universally describe naming as requiring multiple stages, including several steps for lexical retrieval as well as for postlexical output processing (Dell & O'Seaghdha, 1992; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Indefrey & Levelt, 2004; Levelt, Roelofs, & Meyer, 1999; Walker & Hickok, 2015). Specifically, we test the hypothesis that self-reported IS reflects successful lexical access (including both semantics and phonology) and that IS does not rely on output processes such as articulatory motor planning.

Unlike most previous studies on IS in aphasia, this study utilizes a subjective measure of IS alongside carefully matched tasks of objective language abilities to examine the phenomenon in a large group of participants with aphasia. To test the relationship between self-reported IS and lexical access, we compare self-reports of IS on a silent picture-naming task with performance on picture-based tasks that require lexical retrieval, using matched auditory tasks that do not require lexical retrieval as control tasks. If a participant reports the ability to say a particular word in his or her head during silent picture naming, he or she should be more successful when performing tasks that rely on accurate retrieval of that word, so we predict a specific relationship between self-reported IS and performance for the same items on the picture-based tasks. For additional evidence regarding IS and lexical retrieval and to test whether IS relies on speech output processes, we examine the relationship of IS reports to psycholinguistic features

of word stimuli. These include features that are more strongly associated either with retrieval (frequency and age of acquisition [AoA]) or with production (length and articulatory complexity). We predict that IS reports will relate to the word features affecting retrieval processes, but not to those affecting production.

Method

Participants

Participants for this study included adults in the chronic stage of recovery from left-sided stroke (> 6 months) and healthy, age-matched control subjects for task norming. All participants underwent an informed consent process approved by the Georgetown University Institutional Review Board.

Patient Participants

In our post-stroke group, all participants had a history of left-sided stroke at least 6 months prior to enrollment. Several of these participants had evidence of prior small, incidental strokes that were asymptomatic: one in the right putamen, two in the right cerebellum, one in the left cerebellum, and two in right-hemisphere cortical areas. All participants were native English speakers and were required to demonstrate adequate sentence-level auditory comprehension by a minimum score of 48 out of 60 on the Yes/No Questions subtest of the Western Aphasia Battery–Revised (Kertesz, 2006). Individuals who perform in this range typically exhibit good auditory comprehension for task instructions and conversational speech given cues, as needed, by a speech-language pathologist. After initially enrolling 65 participants, two were unable to complete both sessions of the language testing, and nine failed to meet the comprehension cutoff. One additional participant was excluded from the analysis due to having scored near-floor performance on all tasks in the battery (despite meeting the comprehension cutoff). Thus, the final participant group was composed of 53 participants (22 women, 31 men), with a mean age of 60.2 years ($SD = 9.8$, range 40–80), mean education of 16 years ($SD = 2.8$, range 12–24), and mean time since stroke of 5 years ($SD = 4.8$, range 0.5–22.9) and with handedness as follows: 46 right-handed, six left-handed, and one ambidextrous. All participants in this final group presented with adequate single-word intelligibility; that is, no participants had significant evidence of dysarthria, as assessed by two certified speech-language pathologists (the first and second authors).

Control Participants

Participants also included a set of 20 healthy older adult controls who were native English speakers and reported no history of developmental learning disability, neurological disorder, or major psychiatric illness. In this control group, participants were an average age of 65.7 years ($SD = 8.4$), with an average education of 17.2 years ($SD = 2.6$).

Session Format

The testing battery was administered over the course of two testing sessions, lasting approximately 2 hr each. These testing sessions included the tasks described below as well as other language or cognitive measures that are not essential to the questions at hand. The two sessions occurred at least 10 days apart ($M = 18.7$ days). One participant required two additional sessions in order to complete the battery due to a slow pace during testing.

Language Testing Battery

Task Stimuli and Norming Procedure

The primary stimuli were 60 words selected to vary in length, frequency, AoA, and articulatory complexity (see Table 1). The list included 20 each of one-, two-, and three-syllable words. Length was measured in phonemes. Frequency was calculated as the \log_{10} of the frequency identified in the SUBTLEX-US database, which is based on spoken English (Brysaert & New, 2009). AoA was drawn from a database of 30,000 English words with self-reported AoA obtained via Amazon Mechanical Turk (Kuperman, Stadthagen-Gonzalez, & Brysaert, 2012). Articulatory complexity was defined here using the Word Complexity Measure, which was generated in the context of developmental phonology to calculate which complex articulatory structures young speakers are able to produce successfully (Stoel-Gammon, 2010). The measure is calculated based on the presence of features that relate to word patterns (more than two syllables, noninitial stress), syllable structures (word-final consonants, consonant clusters), and specific sound classes (velar consonants such as k/g, liquids, rhotic vowels, fricatives/affricates, and voiced fricatives/affricates; Stoel-Gammon, 2010).

For first-letter identification (ID) and syllable counting (see below), the 60-item list was split into two matched sets (Sets A and B; see Table 1) for the picture-based and auditory versions of the tasks. For IS report and spoken naming (see below), we included an additional 60 items from the Philadelphia Naming Test (Roach, Schwartz, Martin, Grewal, & Brecher, 1996), for a total of 120 items on each (see Table 1 for the word features of this 120-item stimulus list). A large number of naming items were necessary in order to allow for a detailed error analysis beyond a numerical accuracy score; these analyses are not included in the current study. These 120 items were split into two sets of 60 items (Sets 1 and 2), each including 30 in-house items and 30 Philadelphia Naming Test items (see Figure 1). These sets were matched on all four word features: frequency, AoA, length, and articulatory complexity. If Set 1 was used for IS report on Day 1 of testing, Set 2 would be used for spoken naming, with the opposite sets then being used on Day 2 of testing. This design ensured that, after 2 days of testing, each item had been tested exactly once for IS report and once for spoken naming, with each item encountered only once per day in either task. The order was counterbalanced across participants (see Figure 1).

Table 1. Word features for the 60- and 120-item stimulus lists.

Stimulus list	Descriptive statistic	Length (phonemes)	Frequency	Age of acquisition	Articulatory complexity
60-item list	<i>M</i> (<i>SD</i>) Range	5.47 (1.65) 2–9	2.69 (0.61) 1.2–4.1	5.75 (1.26) 3.3–8.7	3.55 (1.41) 1–7
Set A (30-item subset)	<i>M</i> (<i>SD</i>) Range	5.33 (1.52) 2–9	2.78 (0.68) 1.2–4.1	5.54 (1.21) 3.9–8.4	3.37 (1.38) 1–6
Set B (30-item subset)	<i>M</i> (<i>SD</i>) Range	5.6 (1.79) 3–8	2.6 (0.53) 1.5–3.6	5.97 (1.28) 3.3–8.7	3.73 (1.47) 1–7
120-item list (spoken naming and IS report)	<i>M</i> (<i>SD</i>) Range	5.01 (1.75) 2–10	2.86 (0.61) 1.2–4.4	5.22 (1.31) 2.5–8.72	3.41 (1.55) 0–8

Note. The 60-item list was the main stimulus set for all tasks. This list was split into two matched sets, Set A and Set B, for the first-letter identification and syllable-counting tasks. An additional 60 items from the Philadelphia Naming Test (Roach et al., 1996) were added to the full 60-item list to form a 120-item list for use in spoken naming and inner speech (IS) report.

For the picture stimuli, we utilized black-and-white line drawings that were previously normed at Georgetown University in a set of 24 healthy older adult controls (primarily drawn from the International Picture-Naming Project database; Szekely et al., 2004). These control participants completed a confrontation picture-naming task to identify items with high name agreement; that is, at least 70% of the control participants produced the same single-word name for the picture. Novel tasks in the language battery using these preexisting picture stimuli were normed in the 20 healthy control participants (described above in the Participants section); these tasks included the first-letter ID and syllable-counting tasks (both picture-based and auditory versions). The primary purpose of this norming was to determine whether there were any problematic items on the tests, as judged by incorrect performance by a large proportion of healthy controls; all individual items

were answered correctly by at least 17 of 20 control participants, and all items were maintained for patient testing.

IS Report

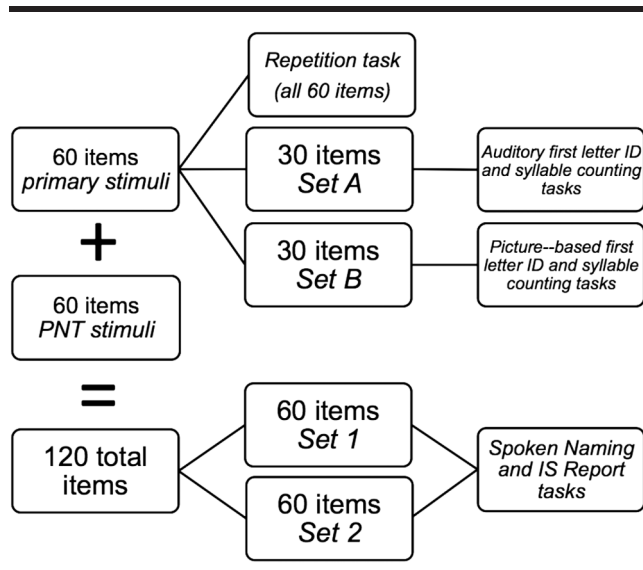
In this task, items were presented one at a time for 20 s on a laptop screen and participants were instructed to name the picture in their heads without moving their lips or tongue. They then pressed a button on the keyboard (labeled with the written words “yes” and “no”) to report whether they could say the word in their head, with all the right sounds in the right order. There were practice items to ensure participants were performing the task appropriately, and the instructions were repeated if necessary. The test items advanced automatically upon key press. This task was administered using PsychoPy presentation software (Peirce, 2009), as were all other tasks in the battery with the exception of repetition; for all tasks, the picture stimuli measured approximately 2 in. in height on the laptop screen.

In designing the IS task, we considered the use of a concatenated task in which the participant would be asked to name the picture aloud immediately after providing a yes/no report of sIS; in fact, we have used such a paradigm in previous work in our laboratory (Hayward, 2016). We elected not to do so here because of concerns that IS reports could be influenced by the upcoming spoken naming task; for example, a person with good self-monitoring of his or her spoken output might answer “no” during the IS report portion of the task simply due to awareness that he or she will not be able to speak that word aloud. Administering the IS report and spoken naming tasks separately thus helps to differentiate retrieval from output and minimize any potential influence across tasks.

Spoken Naming

The spoken naming task was a traditional confrontation naming task, in which participants named a set of stimulus pictures that were presented one at a time for 20 s on a laptop screen. Participants were given instructions to “please use only one word” but were not given any explicit instructions regarding self-monitoring or self-correction.

Figure 1. Task stimuli. This figure illustrates a breakdown of the stimulus lists and the tasks for which they were used. PNT = Philadelphia Naming Test; ID = identification; IS = inner speech.



Participants advanced the test items by pressing the space bar, either once they were satisfied with their response or once the time limit was reached. All sessions were video-recorded.

Repetition

The repetition task was performed using prerecorded stimuli in a natural speaker's voice (the first author), played through QuickTime software on a laptop computer. Participants used high-fidelity headphones to complete the repetition task, which utilized the main stimulus list (60 items) to elicit single-word repetition. Each item was presented a single time with a 5-s intertrial interval. When participants demonstrated a need for additional time, the audio recording was paused manually.

First-Letter ID and Syllable Counting: Picture Based

In this task, participants were presented with 30 picture stimuli, one at a time, and were asked to name the picture in their heads, without moving their lips or tongue, and then indicate the first letter of the word and the number of syllables. A response page was provided with the numbers 1–5 in a single row at the top of the page and the alphabet (in order, in lowercase letters in Arial, a Sans-Serif font) presented in four rows: three rows of seven letters and one row of five letters. Using this response page, the participant could point to the correct answer rather than give a verbal response as desired, although verbal responses were also accepted.

First-Letter ID and Syllable Counting: Auditory

In this task, participants were presented with 30 different stimuli (matched to the picture-based task on all four word features), but an auditory recording of the target word in a natural speaker's voice (the first author) accompanied the pictures (simultaneous presentation). On the same support page as described above, participants pointed to the number of syllables and the first letter of the word. They were allowed to repeat the word aloud to themselves if they did so spontaneously.

Picture Description

Participants were shown the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination—Third Edition (Goodglass, Kaplan, & Barresi, 2001) and were given unlimited time and the following instructions: “Tell me everything you see going on in this picture.” The picture description task was included in order to obtain overall measures of fluency in our participants, because prior studies on IS in individuals with aphasia have shown differences between individuals with fluent versus nonfluent aphasia subtypes (Geva, Bennett, et al., 2011; Goodglass et al., 1976).

Auditory Comprehension Tasks

For sentence-level comprehension, we used the Yes/No Questions subtest of the Western Aphasia Battery—Revised (Kertesz, 2006). This task requires a yes/no response

to 20 items including questions that are biographical, environmental, and noncontextual/grammatically complex in nature. For word-level comprehension, we used the lexical comprehension task, a 48-item auditory word-to-picture matching task developed by Martin, Minkina, Kohen, and Kalinyak-Fliszar (2018), adapted originally from the Philadelphia Comprehension Battery for Aphasia (Martin et al., 2018; Saffran, Schwartz, Linebarger, Martin, & Bochetto, 1988). In this task, participants hear a word and point to the target item in a field of four semantically related pictures.

Hearing Screening

All participants underwent pure-tone audiometry threshold testing. All participants met a minimum threshold of at least 40 dB at 1000 and 2000 Hz in the better ear (Ventry & Weinstein, 1983), except for three patient participants and one control participant. We assessed these participants' functional hearing ability for speech using phoneme discrimination tasks (see the next section). In these tasks, two words or pseudowords are presented and participants provide a yes/no response to indicate whether the two items are identical. Each task is composed of 44 pairs of one- or two-syllable words/pseudowords, with nonmatching stimuli differing by a single phoneme (Martin et al., 2018). All four participants who failed to meet the minimum pure-tone hearing threshold performed significantly above chance on these functional speech perception tasks (all at $p < .0001$), demonstrating adequate hearing ability for the tasks utilized in this study.

Motor Speech Examination

We used a standard motor speech examination protocol to test speech production ability (Haley, Jacks, de Riesthal, Abou-Khalil, & Roth, 2012). This examination requires participants to perform diadochokinetic tasks (alternating motion rates and sequential motion rates [SMRs]) and to repeat a series of multisyllabic words in various task contexts (e.g., repeated productions, increasing length).

Scoring Procedure

Spoken Naming

Two independent raters (the first and third authors) coded naming accuracy, and in the case of coding discrepancies, a consensus was reached via video review and discussion. Specific error codes were assigned to each incorrect response, but for the purposes of the analyses presented here, all responses were coded simply as correct/incorrect. No leniency was given in the case of distortions or single sound substitutions, so accuracy required the attempt to be completely identical to the target word. Finally, errors made within the 20-s stimulus presentation were scored for the presence of spontaneous detection and correction. A spontaneous detection was defined as verbal rejection of a response (e.g., “apple – no”) or production of a second naming attempt that differed from the initial response.

Correction was noted if any subsequent attempts at naming resulted in production of the correct target.

Repetition

Participants' verbal responses were scored as correct/incorrect. Video review was used as needed.

First-Letter ID and Syllable Counting

Participants' responses were scored as correct/incorrect. If participants self-corrected spontaneously, their final answer was accepted.

Picture Description

Participants' spoken narratives were recorded with a video camera and then transcribed into a text file. The work of the primary transcriber (the third author) was checked in its entirety by a second rater (the first author). Following the guidelines of quantitative production analysis (Saffran, Berndt, & Schwartz, 1989), non-narrative words were deleted, including starters, fillers, responses to leading questions asked by the interviewer, and commentary on the task itself. The transcription was then divided into discrete utterances based on grammar, prosody, and length of pause (greater than 2 s indicates an utterance break). An automated script analyzed each plain text file by extracting the mean number of words per utterance and the mean number of words spoken per minute, along with a count of the total real words, unique real words, total nonwords, and unique nonwords produced.

Motor Speech Examination

Analysis of the motor speech examination was performed in order to identify participants with evidence of possible apraxia of speech. Evidence for possible apraxia of speech was assigned to a participant if he or she met at least two of the following criteria, which are particularly characteristic of apraxia of speech (Jacks & Haley, 2015): segment prolongation, ambiguous/distorted consonant production, and inadequate SMR production. Each single-word production was assessed for the presence of ambiguous consonant production and/or segment prolongation by two independent raters (the first and second authors), discrepancies between the raters were resolved via mutual re-review, and a third rater resolved any further disagreement. Adequate SMR production was defined by at least three accurate repetitions of the syllable triad (/pʌtʌkʌ/).

Statistical Analyses

Because IS reports were provided on two different days (at least 10 days apart), we first assessed how consistent participants' subjective judgments were across the two sessions by comparing IS report scores from Days 1 and 2 of testing. We also examined differences in Day 1 and Day 2 scores on spoken naming, in order to compare IS report consistency with this more objective measure.

Our primary analysis examined whether individual participants had more information about the phonology of words they reported being able to say in their heads than words they did not. For this purpose, any participant who reported correct (or incorrect) IS for nearly all words on the IS report task was excluded from subsequent analyses. We set an arbitrary cutoff at five words, so that any participant who reported sIS for < 5 or > 115 items (of 120 in total) was excluded. Because this resulted in the exclusion of 26 participants, we examined differences between the included and excluded groups on their overall performance on the tests of phonological knowledge as well as other language abilities prior to completing the planned analyses. The final group of participants whose data were fit for analysis are characterized as follows: 27 participants (13 women, 14 men), mean age of 61.9 years ($SD = 9.9$, range 40–80), mean education of 15.4 years ($SD = 2.4$, range 12–20), and mean time since stroke of 4.5 years ($SD = 4.7$, range 0.5–22.9), with handedness as follows: 22 right-handed, four left-handed, and one ambidextrous.

In this group ($N = 27$), we used generalized linear mixed-effects models using the `glmer` command for binomial data in the `lme4` package in R (R Core Team, 2017) to examine which of the word features (frequency, AoA, articulatory complexity, and length) predicted performance on IS report and spoken naming across the final group of 27 participants. Then, we examined whether IS report predicted performance on the picture-based tasks and matched auditory versions: spoken naming, repetition, first-letter ID (IS based and auditory), and syllable counting (IS based and auditory). The use of generalized linear mixed-effects models was chosen in order to include random effects of item and participant as well as fixed effects from item-level IS report, participant features (age, education, and chronicity), and word features (frequency, AoA, length, and articulatory complexity). For all analyses, model fitting used a backward-stepwise iterative approach, followed by forward fitting of maximal random effects structure. Model fitting was independently supported by model fitness comparisons using Akaike information criterion (AIC) and Bayesian information criterion (BIC; Akaike, 1972; Schwarz, 1978).

Finally, to provide a measure of the reliability of IS reporting across the group of 27 participants who remained in the analysis, we calculated the difference in performance on the picture-based lexical retrieval tasks for items reported as successful versus unsuccessful on the IS report task; hereafter, these categories will be denoted as sIS and unsuccessful IS (uIS). Because subjects varied in their ability to perform these tasks even when the word was provided to them auditorily, we divided these difference scores by the total scores on the auditory version of each task. An average was taken across the three difference scores to establish a single value that, if positive, indicates better performance on sIS versus uIS items and suggests that responses on the IS report task are meaningful. In a final exploratory analysis, we examined patterns of behavioral performance across the participants who showed evidence of potential

unreliability in IS reporting, based either on abnormally high day-to-day variability or on a lack of better performance for sIS versus uIS words.

Results

Overall Performance on IS Report and the Language Tasks

In comparing IS report scores for Days 1 and 2 for each of the 53 participants, we found that, across the group, participants' raw scores changed very little of the total possible score of 60 points per day, with a median change of 1 point across days (interquartile range = 0–5; see Figure 2). This is comparable with the day-to-day performance on spoken naming, which had a median change of 3 items across days (interquartile range = 2–5). Notably, three outliers showed a difference of > 30 items across the 2 days on the IS report task, which suggests unreliable reporting on at least one of the 2 days for those three participants.

As noted above, 26 participants were excluded from the planned analyses because they reported sIS on nearly every trial (> 115/120). These participants might have reported a high level of sIS for two reasons: This could be a true reflection of their IS, or they might have erroneously overreported their IS compared with the group included in the analysis. Each of these possible explanations leads to specific predictions: In the case of accurate IS, we would expect the excluded group to have better overall phonological knowledge and naming scores, but in the case

of erroneous reporting, there should be no difference between groups.

When directly comparing the two groups (included/excluded), we found several important differences between them (see Table 2). First, the excluded group performed significantly better than the included group on the spoken naming task—by an average of 35 words, a difference that is strikingly similar to the difference in self-reported IS, for which the excluded group reported sIS for an average of 32 more words than the included group. This strongly supports a conclusion that the high rate of IS reports in the excluded group did in fact reflect an accurate self-assessment of relatively milder deficits. These parallel differences in IS report and naming, contrasted with comparable performance across groups on the repetition task, suggest that the main source of difference was in lexical retrieval ability, rather than speech output processes. Furthermore, the excluded group also performed significantly better on the other two picture-based tasks requiring phonological retrieval, but not on the matched control tasks in which the spoken word was presented, which again provides evidence that the main difference between groups is at the level of lexical retrieval. The two groups did not significantly differ in performance on general language measures of auditory comprehension, fluency, or spontaneous error detection/correction and did not differ in incidence of apraxia of speech (see Table 2).

Relationship Between Word Features and Task Performance

In the final group of included participants ($N = 27$), we were first interested in identifying which of the word features predicted performance on IS report and spoken naming. One participant was excluded from the spoken naming analysis because she produced zero correct responses on the task. To assess statistical differences, we used generalized linear mixed-effects models to assess the contributions of each of the word features (fixed effects) while also incorporating the random effects of item and participant. We found that IS report related to AoA, frequency, and length but not to articulatory complexity. Specifically, prior to removing the articulatory complexity from the model as a nonsignificant predictor, its z score value was $z = 0.842$, $p = .399$. In contrast, spoken naming accuracy related to all four of the word features we examined, including articulatory complexity (see Table 3).

Relationship Between IS Report and Task Performance

Next, we examined whether participants demonstrated more evidence of lexical retrieval for items they reported as being able to say correctly using IS compared with items they reported not being able to say with IS. We examined the relationship between IS report (sIS vs. uIS) and performance on the picture-based tasks and the matched auditory versions. On average, participants'

Figure 2. Day-to-day variability of scores on inner speech (IS) report and spoken naming. IS report and spoken naming tasks were administered in two 60-item sets across two testing sessions at least 10 days apart. The horizontal axis represents the difference in scores across these 2 days.

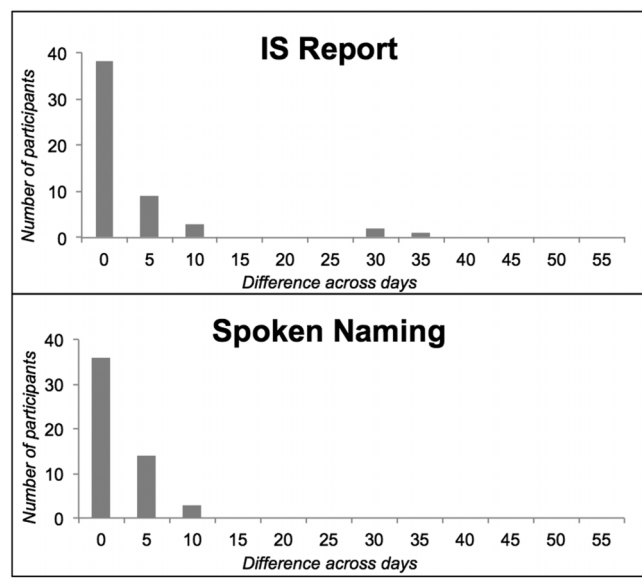


Table 2. Task scores for general language measures.

Task	Possible score	Participants included in the analyses ($N = 27$)	Participants excluded due to sIS > 115 ($N = 26$)	Group comparison	
IS Report, M (SD)	120	86.33 (26.9)	118.85 (1.38)	$t(26.141) = 6.266, p < .001$	
Spoken naming	120	49.19 (34.33)	84.35 (33.48)	$t(51) = 3.773, p < .001$	
Error detection (proportion out of total errors made)	1	0.28 (0.24)	0.31 (0.25)	$t(51) = 0.565, p = .575$	
Error correction (proportion out of total errors detected)	1	0.35 (0.31)	0.52 (0.33)	$t(48) = 1.826, p = .074$	
WAB yes/no questions	60	55.78 (3.56)	56.19 (3.56)	$t(51) = 0.424, p = .673$	
Lexical comprehension (auditory word-to-picture matching)	48	43.52 (5.42)	45.92 (4.45)	$t(49.777) = 1.760, p = .083$	
Repetition (single words)	60	41.41 (15.0)	47.42 (16.02)	$t(51) = 1.412, p = .164$	
Number of participants with evidence of possible apraxia	N/A	14/27	10/26	$\chi^2(1, N = 53) = 0.958, p = .328$	
Fluency measures from picture description task	Average words per minute	Unlimited	39.55 (27.4)	51.22 (27.96)	$t(51) = 1.534, p = .131$
	Average words per utterance	Unlimited	4.12 (2.27)	4.86 (2.31)	$t(51) = 1.158, p = .252$
Primary IS tasks	First-letter identification: picture based	30	17.3 (9.0)	24.58 (6.94)	$t(48.694) = 3.303, p = .002$
	First-letter identification: auditory	30	21.85 (9.96)	26.12 (7.24)	$t(47.5) = 1.788, p = .080$
	Syllable counting: picture based	30	17.15 (7.68)	23.77 (5.46)	$t(51) = 3.606, p = .001$
	Syllable counting: auditory	30	22.30 (7.76)	25.81 (5.59)	$t(51) = 1.884, p = .065$

Note. The table shows overall task results for the group of participants who were included in further analyses ($N = 27$) versus the group who was excluded due to a score > 115 on the inner speech (IS) report task. Independent-samples t tests were performed to compare scores for the two main participant groups. Note that sIS refers to successful inner speech, as reported by participants during a silent picture-naming task. IS = inner speech; sIS = successful inner speech; WAB = Western Aphasia Battery; N/A = not applicable.

Table 3. Word feature comparison across tasks.

Word feature	IS self-report			Naming aloud		
	sIS items	uIS items	LMEM z score	Correct items	Incorrect items	LMEM z score
Frequency	2.93 (0.02)	2.64 (0.03)	3.661**	3.08 (0.03)	2.69 (0.02)	5.40**
Age of acquisition	5.06 (0.02)	5.72 (0.08)	-4.32**	4.78 (0.07)	5.54 (0.06)	-6.14**
Length	4.85 (0.03)	5.49 (0.12)	-2.33*	4.49 (0.08)	5.35 (0.06)	-3.47**
Articulatory complexity	3.36 (0.02)	3.59 (0.08)	<i>ns</i>	3.04 (0.07)	3.62 (0.04)	-2.45*

Note. The table shows the means (standard deviations) for successful inner speech (sIS) versus unsuccessful inner speech (uIS) items on the inner speech (IS) report task and correct versus incorrect items on the spoken naming task. Values represent the grand mean across $N = 27$ for IS report and $N = 26$ for spoken naming (one participant was excluded from the analysis here due to a score of 0 on this task). LMEM = linear mixed effects models. z Scores are provided with significance reported as follows:

* $p < .05$. ** $p < .001$. *ns* = not statistically significant.

accuracy was more than 25% higher for sIS versus uIS items on the three picture-based tasks, but this raw difference was less than 4% for the matched auditory tasks (see Figure 3, with means and standard deviations provided in Table 4). Average performance on sIS items on the first-letter ID (67%) and syllable-counting (64%) tasks was relatively close to performance on sIS items on the matched auditory tasks (76% and 74%, respectively). In contrast, however, naming performance for sIS items (47%) was lower than repetition performance on those same items (73%).

Using generalized linear mixed-effects models to examine the role of IS report in task performance while also considering other stimulus features (length, frequency, AoA, and articulatory complexity), we found that IS report contributed to performance on each of the picture-based

tasks but none of the matched auditory tasks, confirming our predictions (see Table 4). In this set of analyses, we found that IS report was a significant predictor of spoken naming in addition to the four word features previously identified (see Table 3). A chi-square analysis comparing the two models shows that the model including IS report as a predictor (AIC = 2931.9, BIC = 2980.6) is significantly better at predicting spoken naming than the model without IS report (AIC = 2990.5, BIC = 3033.1), $\chi^2(1) = 60.573$, $p < .001$.

Individual Reliability of IS Reporting

After finding that IS report was a significant predictor of group level performance (in $N = 27$) on the picture-based tasks in our battery, we then examined performance at the

Figure 3. Item-level performance for words reported as successful or unsuccessful inner speech (IS) across other tasks. The legend shown in Panel A applies to the entire figure. (A) Tasks relying on phonological retrieval, for which we predicted a relationship between IS report and performance. (B) Matched auditory tasks, for which we did not predict a relationship between IS report and performance. ID = identification.

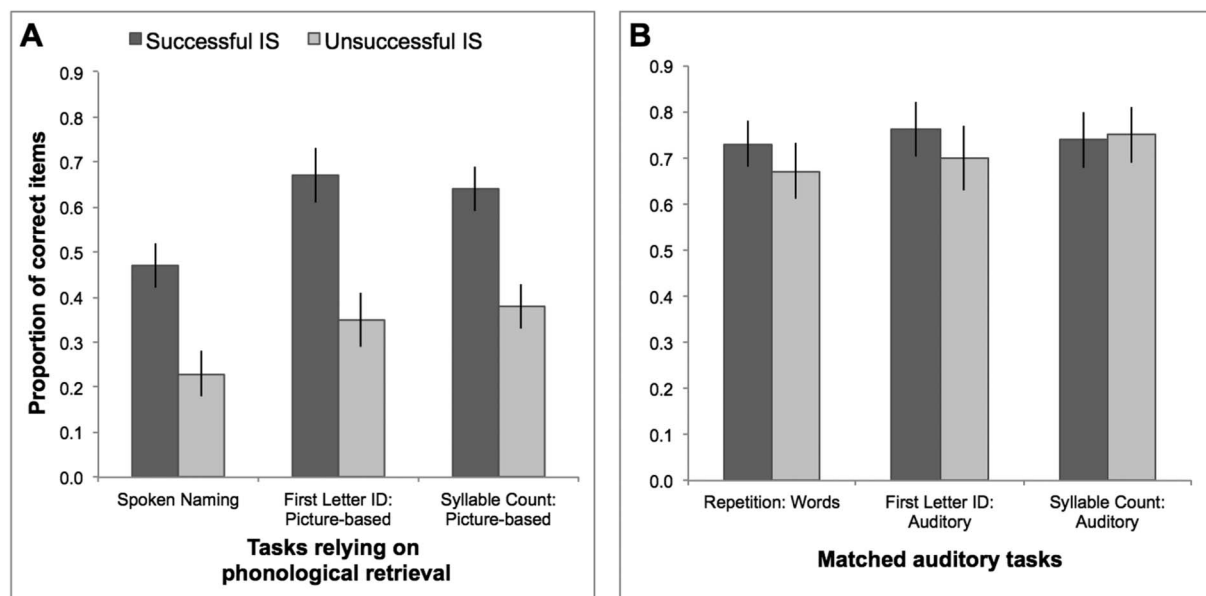


Table 4. Results of generalized linear mixed-effects models predicting task performance.

Task		<i>M (SD) Accuracy</i>		LMEM <i>z</i> scores for predictors				
		sIS Items	uIS Items	IS Report	Articulatory complexity	AoA	Frequency	Length
Picture-based tasks of lexical retrieval	Spoken naming	0.47 (0.05)	0.23 (0.05)	7.70**	-2.63*	-5.82**	4.95**	-3.41**
	First-letter ID	0.67 (0.06)	0.35 (0.06)	5.37**	-2.93*	-3.73**	<i>ns</i>	<i>ns</i>
	Syllable counting	0.64 (0.05)	0.38 (0.05)	4.10**	<i>ns</i>	<i>ns</i>	<i>ns</i>	-4.21**
Matched auditory control tasks	Repetition	0.73 (0.05)	0.67 (0.06)	<i>ns</i>	-2.93*	<i>ns</i>	5.72**	-3.28**
	First-letter ID	0.76 (0.06)	0.70 (0.07)	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
	Syllable counting	0.74 (0.06)	0.75 (0.06)	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	-2.69*

Note. For each task, means and standard deviations are shown for the average proportion of correct performance for successful inner speech (sIS) items and unsuccessful inner speech (uIS) items. Model fitting was performed in a backward-stepwise iterative fashion, followed by forward fitting of maximal random effects structure. Model fitting was independently supported by model fitness comparisons using Akaike information criterion and Bayesian information criterion. LMEM = linear mixed effects models; IS = inner speech; AoA = age of acquisition; ID = identification; *ns* = not statistically significant.

* $p < .01$. ** $p < .001$.

subject level to identify how many individual participants in this group showed evidence of reliability in IS reporting. Using the item-level responses on the IS report task, we calculated a difference score to compare performance for sIS versus uIS words on the three picture-based tasks requiring lexical retrieval. Overall, 24 of 27 participants showed better performance for sIS words than for uIS words, as indicated by a positive value of the average difference score as described in the Method section, providing good evidence for reliability of item-level IS reporting in all but three participants.

Thus, there are six participants with some evidence of unreliable IS reporting: the three outliers with high day-to-day variability in IS reports (see Figure 2) and the three participants who did not consistently show better performance for sIS versus uIS words. When examining performance of these participants compared with the rest of the group, there were a few notable differences (see Table 5). The participants with some evidence of unreliability ($n = 6$) showed lower accuracy on the spoken naming task as well as less frequent spontaneous detection and correction of their naming errors. They also showed poorer performance on the single-word auditory comprehension task as well as the first-letter ID (both picture-based and auditory) and syllable counting (picture-based only) tasks. There were no significant differences in overall IS report, sentence-level auditory comprehension, repetition, or fluency measures.

Discussion

The primary aim of this study was to determine whether self-reported IS is meaningful at the level of individual items in people with aphasia. By demonstrating that IS report relates to performance on objective language tasks, we have shown that at least some people with aphasia appear to be reliable in reporting the success of their IS and that self-reported IS relates to lexical retrieval ability. We will discuss the implications of our findings in relationship to processing stages of the naming process (i.e., lexical access vs. output) and to theories of self-monitoring

in aphasia. Finally, we will describe the clinical relevance of self-reported IS in aphasia, giving consideration to possible future directions for this line of work.

The Subjective Experience of IS Relates to Lexical Retrieval

In examining the validity of IS reports, we were particularly interested in identifying how this experience can be understood in the context of processing models of naming. Our results confirmed the hypothesis that self-reported IS relates to successful lexical access, as reports of sIS related specifically to performance on picture-based tasks that depend on lexical retrieval. Also, on the picture-based first-letter ID and syllable counting tasks, participants performed nearly as well on items that they reported as sIS as they did on the matched auditory tasks in which the words were presented to them, which suggests that the experience of sIS reflects retrieval of at least some information about the word in a great majority of cases.

However, when comparing naming and repetition performance for sIS items, we found that average naming accuracy was lower than repetition accuracy of the same words. It is important to note that the picture-based/auditory tasks were not administered concurrently but instead given separately during a testing session or even across two different testing sessions. Given the probabilistic nature of word retrieval (Freed, Marshall, & Chuhlantseff, 1996; Howard, Patterson, Franklin, Morton, & Orchard-Lisle, 1984), we expected that there would likely be an impact of day-to-day variability on our findings. Accurate retrieval in one instance does not guarantee success in another instance, so we did not predict that sIS items would always be successfully retrieved during the spoken naming task, even if self-reported sIS accurately reflects complete retrieval of a lexical phonological form. If an individual retrieves a word correctly on one occasion, however, it is likely that he or she will be able to retrieve at least some of the phonology at another time (Howard et al., 1984). Thus, because first-letter ID and

Table 5. Task scores for general language measures: Reliability analysis.

Task	Possible score	Participants with some evidence of unreliable IS reporting (<i>n</i> = 6)	Remaining participants in the included group (<i>n</i> = 21)	Group comparison	
IS Report, <i>M</i> (<i>SD</i>)	120	74.5 (26.28)	89.71 (26.76)	$t(25) = -1.233, p = .229$	
Spoken naming	120	10.5 (5.75)	60.24 (30.77)	$t(23.77) = -6.993, p < .001$	
Error detection (proportion out of the total errors made)	1	0.09 (0.07)	0.33 (0.25)	$t(25) = -2.208, p = .037$	
Error correction (proportion out of the total errors detected)	1	0.03 (0.05)	0.44 (0.29)	$t(23.19) = -6.269, p < .001$	
WAB yes/no questions	60	55 (3.63)	56 (3.59)	$t(25) = -0.600, p = .554$	
Lexical comprehension (auditory word-to-picture matching)	48	38.33 (6.28)	45 (4.24)	$t(25) = -3.050, p = .005$	
Repetition (single words)	60	39.67 (15.58)	41.9 (15.18)	$t(25) = -0.317, p = .754$	
Number of participants with possible evidence of apraxia	N/A	4/6	10/21	$\chi^2(1, N = 27) = 0.678, p = .410$	
Fluency measures from the picture description task	Average words per minute	Unlimited	25.45 (21.26)	43.58 (28.04)	$t(25) = -1.460, p = .157$
	Average words per utterance	Unlimited	3.31 (2.43)	4.36 (2.22)	$t(25) = -1.008, p = .323$
Primary IS tasks	First-letter identification: picture based	30	5.5 (4.51)	20.67 (6.84)	$t(25) = -5.089, p < .001$
	First-letter identification: auditory	30	9.33 (9.22)	25.43 (6.87)	$t(25) = -4.7, p < .001$
	Syllable counting: picture based	30	9.33 (6.86)	19.38 (6.43)	$t(25) = -3.330, p = .003$
	Syllable counting: auditory	30	17.33 (10.39)	23.71 (6.46)	$t(25) = -1.860, p = .075$

Note. The table shows overall task results for the group of participants who revealed some evidence of unreliable IS reporting (*n* = 6) compared with the remainder of the participants in the group included in all analyses (*n* = 21). Independent-samples *t* tests were performed to compare scores for the two groups. IS = inner speech; WAB = Western Aphasia Battery; N/A = not applicable.

syllable counting can be performed with more limited lexical knowledge, we expected performance on these tasks to be more resilient to the variability of word retrieval.

There are at least two alternative interpretations of the discrepancy between naming and repetition scores for sIS items, in the context of comparable scores on the other picture-based/auditory task pairs. First, in some cases, participants may experience and report sIS when they are close to achieving complete phonological retrieval, rather than when retrieval is complete. A second alternative interpretation stems from processing models of naming that assume two levels of phonological representation: lexical phonological representations that are accessed during word retrieval and postlexical phonological representations that support articulatory output (Goldrick & Rapp, 2007). In such a model, an individual with aphasia could experience sIS given successful retrieval of the lexical phonological representation but fail to name an item correctly aloud due to impaired mapping between lexical and postlexical phonology. An impairment at that level would spare repetition ability, because repetition can occur in a nonlexical route based on activation of postlexical phonological representations through acoustic–phonological conversion (Goldrick & Rapp, 2007).

Generally, the data showing some spared ability to perform tasks of phonological knowledge are consistent with prior work on IS in aphasia (Feinberg et al., 1986; Goodglass et al., 1976). These prior studies either examined tip-of-the-tongue rather than a more strictly defined experience of IS (Goodglass et al., 1976) or compared IS-based performance to tasks using different stimuli (Feinberg et al., 1986). Our work therefore extends these prior findings by showing that an individual's perception of the success of IS on individual items is predictive of performance on those same items on other tasks requiring retrieval. Our findings also build on prior item-level work in which we showed in six individuals with aphasia that self-reported sIS related to phonological retrieval, based on specific associations with naming accuracy as well as certain error types and word features (Hayward, 2016; Hayward et al., 2016).

A separate source of support for our hypothesis regarding lexical access comes from the relationship that was identified between self-reported IS and specific word features that relate to word retrieval. Frequency and AoA are features that are closely related to the efficiency and/or integrity of word retrieval in healthy speakers and in individuals with aphasia. Frequency effects can predict naming accuracy and error types in individuals with aphasia (Butterworth, Howard, & Mcloughlin, 1984; Kittredge, Dell, Verkuilen, & Schwartz, 2008). Similar patterns are observed for AoA (Brysbaert & Ellis, 2016; Jescheniak & Levelt, 1994; Morrison, Ellis, & Quinlan, 1992; Snodgrass & Yuditsky, 1996), and the effects of AoA have actually been suggested to supersede frequency effects, for example, when both variables are included simultaneously in a regression model predicting naming accuracy (Nickels & Howard, 1995). In this study, we found that both frequency and AoA

were significant predictors in our models predicting IS report (as well as spoken naming), which supports our hypothesis, our own prior work (Hayward, 2016), and the work of others (Oppenheim & Dell, 2010) suggesting that IS arises from lexical (phonological) retrieval.

The Subjective Experience of IS May Not Require Articulation

We have discussed the support for our hypothesis that self-reported IS reflects lexical access/retrieval. An important extension of this hypothesis is that IS does not rely on the postlexical output processes that prepare a lexical item for spoken production. These output processes can be characterized as involving the sensorimotor interface, in which auditory representations are converted to motor-related representations, and motor programming for articulation (Indefrey, 2011; Walker & Hickok, 2015). Based on our hypothesis, we initially predicted that IS report would relate uniquely to features primarily influencing word retrieval (e.g., frequency, AoA) and not features primarily influencing word production (e.g., length, articulatory complexity). As predicted, we found a lack of relationship between IS report and articulatory complexity, which directly supports our hypothesis that IS does not relate to output processing; however, we also found a significant relationship between IS report and word length, measured in phonemes. Although we did not predict it, this finding is not incompatible with our hypothesis. Whereas most studies describe the word length effect as postlexical in nature, prior work has suggested that word length's effect on naming may be relevant during retrieval itself. In testing the dual-origin theory of phonological errors in naming, Schwartz et al. found that word length relates to prevalence of errors with high phonological overlap with the target ("proximate" errors) as well as errors with low overlap ("remote" errors; Schwartz, Wilshire, Gagnon, & Polansky, 2004). Because remote errors commonly arise during lexical retrieval, the authors conclude that length effects can arise during lexical retrieval in addition to postlexical processing (the more common origin of proximate errors; Schwartz et al., 2004). Thus, the relationship between self-reported IS and length may not necessarily represent a challenge for our hypothesis about the role of output processes in IS, but further research is needed to clarify this association.

Additional support for the lack of relationship between IS and output processing comes from the comparison of the group that remained in the main analyses, who reported variable levels of sIS, and the group of participants who were excluded due to self-reported sIS above 95% on the silent picture-naming task. There was a large difference between groups on average naming accuracy but similar repetition accuracy, suggesting that sIS is more closely related to retrieval than to output processing. If postlexical output processing is in fact unnecessary for the experience of IS, a person with deficits in output processing (e.g., someone with conduction aphasia or apraxia of speech) could experience and report sIS based on intact

retrieval ability, despite subsequent failure of spoken naming. This possibility aligns with our prior finding that a failure of spoken naming following an experience of sIS is common in people with aphasia and is associated with lesions primarily in the left ventral sensorimotor cortex, which supports speech output processes (Fama et al., 2017).

Importantly, the role of articulatory processing in IS not only is relevant in the context of aphasia but is also discussed within the general literature on IS, where the question remains open as to whether IS in healthy individuals involves prearticulatory motor planning processes. Early models described IS as including all stages of speech production up to overt articulation, thus including a fully specified articulatory plan (Levelt, 1983; Postma & Noordanus, 1996). Since then, many theorists have shifted toward models of IS that are more abstract in nature, without particular articulatory features (Indefrey & Levelt, 2004; Levelt, 2001; Oppenheim & Dell, 2008). Some recent work takes a more intermediate stance, however, suggesting that abstract phonology is the primary level at which IS is achieved but that it can be affected by articulatory factors under certain circumstances (Oppenheim & Dell, 2010). Additional work has used neuroimaging approaches in healthy adults to reexamine earlier accounts of IS as having a necessary component of articulatory specificity. In studies of motor imagery of speech, evidence has been provided for a theory in which efference copies from the motor system provide feedback to sensory regions, allowing for monitoring of IS prior to overt articulation (Tian & Poeppel, 2013, 2015). Our findings would not be consistent with such a model, which requires motor processes for IS monitoring. In general, our results align with theories of IS in which monitoring can rely on earlier stages of processing and articulatory planning is not necessary for IS. Our findings, however, do not rule out the possibility that experiences of IS can reflect speech production processes in some circumstances, as discussed below.

Reliability and the Role of Self-Monitoring in the Experience of IS in Aphasia

The study presented here examines a phenomenon that is assessed primarily through self-report by the participants. The ability to accurately report on one's internal experience of language relies on self-monitoring, several theories of which will be described later in this section. In general, self-monitoring of spoken output does vary across individuals with aphasia, but there is ample prior literature suggesting that many individuals with aphasia are able to monitor their own speech errors, even without explicit instruction to do so (Marshall, Neuburger, & Phillips, 1994; Nickels & Howard, 1995; Schuchard, Middleton, & Schwartz, 2017; Schwartz, Middleton, Brecher, Gagliardi, & Garvey, 2016).

We have shown, by finding significant predictive relationships between subjective reports and objective measures, that the self-report data regarding the success of IS

are meaningful to some extent. Importantly, however, we excluded a large proportion of our original participant group from the main item-level analysis due to individual reports of nearly 100% success on the IS report task; thus, because our measure of reliability requires an adequate number of both sIS and uIS items, we were only able to assess individual differences in IS reliability in those participants who were included in the final analyses. Within that group, it appears that a relatively small proportion of individual participants (3/27) may have been unreliable in their IS reporting, as evidenced by a lack of better performance on items reported as sIS versus those reported as uIS. Three additional participants showed significant variability across the 2 days of IS report testing, which could also be evidence of unreliability. These six participants with some evidence for unreliable IS reporting were less likely than other participants to spontaneously detect and correct their errors during spoken naming, which suggests that reduced self-monitoring ability similarly impacts both inner and overt speech.

Participants excluded due to reporting near 100% success on the IS report task demonstrated objective task performance, suggesting that their word retrieval was likely preserved relative to the other participants in our sample. Within each group, however, self-monitoring ability certainly varies across individual participants, which represents a potential limitation in the scope of conclusions that can be drawn. In the remainder of this section, we provide detailed consideration of theories of self-monitoring in order to understand (a) how processing models account for the monitoring of IS in general and (b) the potential impacts of aphasia on self-monitoring ability.

There is ongoing debate as to the processing mechanisms and neural bases through which self-monitoring of spoken output is achieved. A long-standing theory, proposed first by Levelt in 1983 and refined over the following decades, suggests that self-monitoring is performed by the comprehension system and can occur based on one's inner or overt speech, using the same mechanisms by which language users understand the speech of others (Levelt, 1983). In contrast to this comprehension-based monitor, several models have been proposed in which self-monitoring is achieved based on the production process itself, either through comparisons of actual versus target output or through recognition of unfamiliar patterns of activation (Laver, 1980; MacKay, 1987). Recently, a conflict-based account built upon other production-based models of self-monitoring, proposing that error detection occurs via response conflict during the production stage of speech (Nozari, Dell, & Schwartz, 2011). This model was developed based on the two-step interactive processing model of word retrieval (Dell & O'Seaghdha, 1992), which does not account for postretrieval speech production, so the conflict-based monitor is acting within the stages of word retrieval (Nozari et al., 2011). Each of these major theories of self-monitoring accounts for mechanisms of both external and internal self-monitoring, the latter of which is crucial to our study.

Critically, our interpretation of IS as a reflection of lexical retrieval does not require fidelity to one specific theory of self-monitoring but rather is consistent with both comprehension- and production-based monitors. If a sense of sIS arises in conjunction with successful lexical (i.e., phonological) retrieval, then this phonological form would be available prior to spoken output either to the auditory processing system for comprehension-based monitoring or to the domain-general error detection system for conflict-based monitoring. One exception to this compatibility is the theory of IS monitoring that relies on efference copies produced by the motor system during production (Tian & Poeppel, 2013, 2015), but there are several reasons our specific findings need not be compatible with this theory. First, IS monitoring was performed during a silent naming task in which participants were explicitly instructed not to perform any silent articulation of any mouth, lip, or tongue movements, inherently limiting the role of the motor system in the task. Second, to the degree that motor planning and articulatory programming systems are damaged in these patients, they may not be reliable for monitoring, and participants may favor prearticulatory monitoring systems, as observed in a prior study directly comparing prearticulatory and postarticulatory monitoring in individuals with Broca's aphasia (Oomen, Postma, & Kolk, 2001). Finally, our study assesses self-monitoring in the context of asking participants to perform a conscious judgment (rather than by observing natural monitoring in the context of error repair, for instance), which may change the nature of the self-monitoring process. Our findings show simply that IS can be monitored based on prearticulatory, phonological representations in the context of conscious judgments during a naming task. It remains possible that IS could evoke articulatory processes in certain situations, as necessitated by task context or demands, such as when IS is silently mouthed (Oppenheim & Dell, 2010) or during silent rehearsal, particularly for upcoming speech output.

Clinical Implications and Future Directions

To date, research on IS in aphasia has universally concluded that this ability can be preserved in some individuals with aphasia, above and beyond spoken language ability (Feinberg et al., 1986; Geva, Bennett, et al., 2011; Goodglass et al., 1976; Stark et al., 2017). Moreover, there is growing evidence for validity to the subjective experience and self-report of IS by individuals with aphasia (Fama et al., 2017; Hayward et al., 2016). Taken together with the results of the current study, findings suggest that IS is a valuable and clinically relevant topic in aphasia. Behavioral language therapy with a speech-language pathologist is the most common treatment for anomia, and despite evidence supporting its efficacy, there is no universally successful evidence-based approach to the selection of treatment stimuli or even treatment paradigm (Brady, Kelly, Godwin, & Enderby, 2012). An attractive future direction for this line of work, therefore, is to investigate whether

the subjective experience of IS is a useful avenue for selecting effective approaches to, or specific stimuli for, anomia treatment.

There is preliminary evidence for a relationship between self-reported IS and treatment outcomes, from a study in which item-level IS reports predicted response to treatment in two participants with aphasia (Hayward et al., 2016). A prospectively designed treatment study in a larger group of people with aphasia would help determine whether IS is a reliable predictor of therapeutic outcomes on a larger scale. We have shown here that item-level IS report predicts performance across other retrieval-based tasks, even when those tasks may occur during a different testing session on a different day. This suggests that there is enough stability to retrieval probabilities across time such that IS report on an item-by-item level could be informative for the selection of treatment stimuli. There are two main ways in which this could be useful. First, self-reported IS could predict which words would be learned faster, thereby informing stimulus selection for treatment. For instance, a clinician could focus on stimuli that were reported as sIS in order to make treatment more efficient overall, or he or she might select a mix of sIS and uIS words to stagger success throughout a longer treatment course. Another benefit of using self-reported IS might be to determine treatment approach at the individual item level: One could target sIS words using an output-focused approach (Kendall et al., 2008; Raymer, Thompson, Jacobs, & Le Grand, 1993), because these words are already being successfully retrieved (at least in part), but then target uIS words with a more retrieval-focused approach, such as semantic feature analysis (Boyle & Coelho, 1995).

Knowledge of the subjective experience of IS may be useful in a more general sense, beyond specific treatment planning at the item level. Understanding how often an individual with aphasia is experiencing sIS during anomia could contribute to determining the main cause of anomia overall, which in turn could inform overall treatment approaches. Although cueing has not been previously studied specifically in the context of IS, understanding an individual's overall experience of IS during anomia could be useful in identifying what types of cues might be most appropriate (Linebaugh, Shisler, & Lehner, 2005; Wambaugh et al., 2001) or even in helping determine whether self-cueing might be possible (DeDe, Parris, & Waters, 2003). Additionally, an individual whose experience of sIS suggests that his or her retrieval is relatively intact might benefit from the use of specific compensatory strategies for conveying words for which output fails, such as writing/finger-tracing the first letter of the word or pointing to the first letter on a letter board, both of which are less likely to be useful to someone whose anomia results mostly from word retrieval failures. In these ways, understanding the subjective experience of IS in patients with aphasia has the potential to validate a common patient experience and inform clinicians' decision-making process at all levels of treatment planning, although more research is clearly

needed to determine the specific impacts that self-reported IS could have on clinical practice.

Conclusions

In this study, we have demonstrated that self-reported IS can predict item level performance on objective language tasks that rely on word retrieval. Furthermore, we have shown that articulatory complexity is a significant predictor of spoken naming, but not of self-reported IS. Taken together, these findings support a theory in which the subjective experience of sIS arises in association with lexical retrieval and that output processes (i.e., articulation) are not required for the experience of IS. These conclusions have the potential to impact treatment approaches for naming deficits in aphasia and also help further our understanding of IS in the general population.

Acknowledgments

This work was funded by National Institute on Deafness and Other Communication Disorders Grants F31DC014875 (awarded to Mackenzie E. Fama) and R03DC014310 (awarded to Peter E. Turkeltaub). Mackenzie E. Fama received additional training support through the ASHFoundation New Century Scholars Doctoral Scholarship. We are grateful to the individuals with aphasia who gave their time and effort to participate in this study. Thank you also to Kathryn Schuler for providing support during task programming in PsychoPy and to Maryam Ghaleh for her assistance with statistical analyses in R.

References

- Akaike, H.** (1972). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Caski (Eds.), *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.
- Alderson-Day, B., & Fernyhough, C.** (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*(5), 931–965. <https://doi.org/10.1037/bul0000021>
- Berthier, M. L.** (2005). Poststroke aphasia: Epidemiology, pathophysiology and treatment. *Drugs & Aging, 22*(2), 163–182. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15733022>
- Blanken, G., Dittmann, J., Haas, J.-C., & Wallesch, C.-W.** (1987). Spontaneous speech in senile dementia and aphasia: Implications for a neurolinguistic model of language production. *Cognition, 27*, 247–274.
- Boyle, M., & Coelho, C.** (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*(4), 94–98.
- Brady, M., Kelly, H., Godwin, J., & Enderby, P.** (2012). Speech and language therapy for aphasia following stroke (review). *The Cochrane Library, 5*, 1–229.
- Bryshaert, M., & Ellis, A. W.** (2016). Aphasia and age-of-acquisition: Are early-learned words more resilient? *Aphasiology, 30*(11), 1240–1263.
- Bryshaert, M., & New, B.** (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Butterworth, B., Howard, D., & McLoughlin, P.** (1984). The semantic deficit in aphasia: The relationship between semantic errors in auditory comprehension and picture naming. *Neuropsychologia, 22*(4), 409–426. [https://doi.org/10.1016/0028-3932\(84\)90036-8](https://doi.org/10.1016/0028-3932(84)90036-8)
- DeDe, G., Parris, D., & Waters, G.** (2003). Teaching self-cues: A treatment approach for verbal naming. *Aphasiology, 17*(5), 465–480. <https://doi.org/10.1080/02687030344000094>
- 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](https://doi.org/10.1016/0010-0277(92)90046-K)
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A.** (1997). Lexical access in aphasic and non-aphasic speakers. *Psychological Review, 104*, 801–838. <https://doi.org/10.1037/0033-295X.104.4.801>
- Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., . . . Lyrer, P. A.** (2006). Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency, etiology, and thrombolysis. *Stroke, 37*(6), 1379–1384. <https://doi.org/10.1161/01.STR.0000221815.64093.8c>
- Fama, M. E., Hayward, W., Snider, S. F., Friedman, R. B., & Turkeltaub, P. E.** (2017). Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates. *Brain and Language, 164*, 32–42. <https://doi.org/10.1016/j.bandl.2016.09.009>
- Feinberg, T., Rothi, L., & Heilman, K.** (1986). “Inner speech” in conduction aphasia. *Archives of Neurology, 43*, 591–593. Retrieved from <http://archneur.ama-assn.org/cgi/reprint/43/6/591.pdf>
- Freed, D., Marshall, R., & Chuhlantseff, E.** (1996). *Picture naming variability: A methodological consideration of inconsistent naming responses in fluent and nonfluent aphasia*. Paper presented at the Clinical Aphasiology Conference, Traverse City, MI. Retrieved from <http://aphasiology.pitt.edu/id/eprint/221>
- Geva, S., Bennett, S., Warburton, E. A., & Patterson, K.** (2011). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology, 25*(3), 323–343. <https://doi.org/10.1080/02687038.2010.511236>
- Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A.** (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain, 134*(Pt. 10), 3071–3082. <https://doi.org/10.1093/brain/awr232>
- Goldrick, M., & Rapp, B.** (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition, 102*(2), 219–260. <https://doi.org/10.1016/j.cognition.2005.12.010>
- Goodglass, H., Kaplan, E., & Barresi, B.** (2001). *BDAE: The Boston Diagnostic Aphasia Examination—Third Edition*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Goodglass, H., Kaplan, E., Weintraub, S., & Ackerman, N.** (1976). The “tip-of-the-tongue” phenomenon in aphasia. *Cortex, 12*, 145–153. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0010945276800184>
- Haley, K. L., Jacks, A., de Riesthal, M., Abou-Khalil, R., & Roth, H. L.** (2012). Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(5), S1502–S1517. [https://doi.org/10.1044/1092-4388\(2012\)11-0318](https://doi.org/10.1044/1092-4388(2012)11-0318)
- Hayward, W.** (2016). *Objective support for the subjective report of successful inner speech in aphasia* (Doctoral dissertation). Georgetown University, Washington, DC.
- Hayward, W., Snider, S. F., Luta, G., Friedman, R. B., & Turkeltaub, P. E.** (2016). Objective support for subjective reports of successful inner speech in two people with aphasia. *Cognitive Neuropsychology, 33*, 299–314. <https://doi.org/10.1080/02643294.2016.1192998>

- Hilari, K., Northcott, S., Roy, P., Marshall, J., Wiggins, R. D., Chataway, J., & Ames, D.** (2010). Psychological distress after stroke and aphasia: The first six months. *Clinical Rehabilitation, 24*(2), 181–190. <https://doi.org/10.1177/0269215509346090>
- Howard, D., Patterson, K., Franklin, S., Morton, J., & Orchard-Lisle, V.** (1984). Variability and consistency in picture naming by aphasic patients. *Advances in Neurology, 42*, 263–276.
- Hurlburt, R. T., Alderson-Day, B., Kühn, S., & Fernyhough, C.** (2016). Exploring the ecological validity of thinking on demand: Neural correlates of elicited vs. spontaneously occurring inner speech. *PLoS One, 11*(2), e0147932. <https://doi.org/10.1371/journal.pone.0147932>
- 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>
- Indefrey, P., & Levelt, W. J. M.** (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Jacks, A., & Haley, K. L.** (2015). Auditory masking effects on speech fluency in apraxia of speech and aphasia: Comparison to altered auditory feedback. *Journal of Speech, Language, and Hearing Research, 58*, 1670–1686.
- Jescheniak, J. D., & Levelt, W. J. M.** (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 824–843. <https://doi.org/10.1037/0278-7393.20.4.824>
- Kendall, D. L., Rosenbek, J. C., Heilman, K. M., Conway, T., Klenberg, K., Gonzalez Rothi, L. J., & Nadeau, S. E.** (2008). Phoneme-based rehabilitation of anomia in aphasia. *Brain and Language, 105*(1), 1–17. <https://doi.org/10.1016/j.bandl.2007.11.007>
- Kertesz, A.** (2006). *Western Aphasia Battery-Revised (WAB-R)*. San Antonio, TX: Pearson.
- Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F.** (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25*(4), 463–492. <https://doi.org/10.1080/02643290701674851>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M.** (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990. <https://doi.org/10.3758/s13428-012-0210-4>
- Laine, M., & Martin, N.** (2006). *Anomia: Theoretical and clinical aspects*. New York, NY: Psychology Press.
- Laver, J.** (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 287–305). New York, NY: Academic Press.
- Levelt, W. J. M.** (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104. <https://doi.org/10.1111/0023-8333.00090>
- Levelt, W. J. M.** (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 13464–13471. <https://doi.org/10.1073/pnas.231459498>
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S.** (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences, 22*, 1–38; discussion 38–75. <https://doi.org/10.1017/S0140525X99001776>
- Linebaugh, C. W., Shisler, R. J., & Lehner, L.** (2005). CAC classics: Cueing hierarchies and word retrieval: A therapy program. *Aphasiology, 19*(1), 77–92. <https://doi.org/10.1080/02687030444000363>
- MacKay, D. G.** (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York, NY: Springer-Verlag.
- Maher, L. M., & Raymer, A. M.** (2004). Management of anomia. *Topics in Stroke Rehabilitation, 11*(1), 10–21. <https://doi.org/10.1310/318R-RMD5-055J-PQ40>
- Marshall, R. C., Neuburger, S. I., & Phillips, D. S.** (1994). Verbal self-correction and improvement in treated aphasic clients. *Aphasiology, 8*(6), 535–547. <https://doi.org/10.1080/02687039408248680>
- Martin, N., & Dell, G. S.** (2007). Common mechanisms underlying perseverative and non-perseverative sound and word substitutions. *Aphasiology, 21*(10–11), 1002–1017. <https://doi.org/10.1080/02687030701198346>
- Martin, N., Minkina, I., Kohen, P., & Kalinyak-Fliszar, M.** (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neuro-linguistics, 48*, 199–225.
- Morin, A., Utzl, B., & Hamper, B.** (2011). Self-reported frequency, content, and functions of inner speech. *Procedia - Social and Behavioral Sciences, 30*, 1714–1718. <https://doi.org/10.1016/j.sbspro.2011.10.331>
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T.** (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*(6), 705–714. <https://doi.org/10.3758/BF03202720>
- Nickels, L. A., & Howard, D.** (1995). Aphasic naming: What matters? *Neuropsychologia, 33*(10), 1281–1303.
- Nozari, N., Dell, G. S., & Schwartz, M. F.** (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63*(1), 1–33. <https://doi.org/10.1016/j.cogpsych.2011.05.001>
- Oomen, C. C., Postma, A., & Kolk, H. H.** (2001). Prearticulatory and postarticulatory self-monitoring in Broca's aphasia. *Cortex, 37*(5), 627–641. [https://doi.org/10.1016/S0010-9452\(08\)70610-5](https://doi.org/10.1016/S0010-9452(08)70610-5)
- Oppenheim, G. M., & Dell, G. S.** (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528–537. <https://doi.org/10.1016/j.cognition.2007.02.006>
- Oppenheim, G. M., & Dell, G. S.** (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition, 38*(8), 1147–1160. <https://doi.org/10.3758/MC.38.8.1147>
- Peirce, J. W.** (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 1–8. <https://doi.org/10.3389/neuro.11.010.2008>
- Postma, A., & Noordanus, C.** (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39*(4), 375–392. <https://doi.org/10.1177/002383099603900403>
- Raymer, A. M., Thompson, C. K., Jacobs, B., & Le Grand, H. R.** (1993). Phonological treatment of naming deficits in aphasia: Model-based generalization analysis. *Aphasiology, 7*(1), 27–53. <https://doi.org/10.1080/02687039308249498>
- R Core Team.** (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A.** (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology, 24*, 121–133.
- Saffran, E. M., Berndt, R. S., & Schwartz, M. F.** (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440–479. [https://doi.org/10.1016/0093-934X\(89\)90030-8](https://doi.org/10.1016/0093-934X(89)90030-8)

- Saffran, E. M., Schwartz, M. F., Linebarger, M., Martin, N., & Bochetto, P. (1988). *Philadelphia Comprehension Battery for Aphasia*. Unpublished manuscript.
- Schuchard, J., Middleton, E. L., & Schwartz, M. F. (2017). The timing of spontaneous detection and repair of naming errors in aphasia. *Cortex*, *93*, 79–91. <https://doi.org/10.1016/j.cortex.2017.05.008>
- 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>
- Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology*, *21*(2–4), 159–186. <https://doi.org/10.1080/02643290342000519>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 516–536. <https://doi.org/10.3758/BF03200741>
- Stark, B. C., Geva, S., & Warburton, E. A. (2017). Inner speech's relationship with overt speech in poststroke aphasia. *Journal of Speech, Language, and Hearing Research*, *60*, 2406–2415.
- Stoel-Gammon, C. (2010). The Word Complexity Measure: Description and application to developmental phonology and disorders. *Clinical Linguistics & Phonetics*, *24*(4–5), 271–282. <https://doi.org/10.3109/02699200903581059>
- Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250. <https://doi.org/10.1016/j.jml.2004.03.002>
- Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*(7), 1020–1036. https://doi.org/10.1162/jocn_a_00381
- Tian, X., & Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience*, *27*(2), 352–364. https://doi.org/10.1162/jocn_a_00692
- Ventry, I., & Weinstein, B. (1983). Identification of elderly people with hearing problems. *Asha*, *25*(7), 37–42.
- Walker, G. M., & Hickok, G. (2015). Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM). *Psychonomic Bulletin & Review*, *23*, 339–352. <https://doi.org/10.3758/s13423-015-0903-7>
- Wambaugh, J. L., Linebaugh, C. W., Doyle, P. J., Martinez, A. L., Kalinyak-Fliszar, M., & Spencer, K. A. (2001). Effects of two cueing treatments on lexical retrieval in aphasic speakers with different levels of deficit. *Aphasiology*, *15*(10–11), 933–950. <https://doi.org/10.1080/02687040143000302>