THE SUBJECTIVE EXPERIENCE OF INNER SPEECH IN APHASIA

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Neuroscience

By

Mackenzie Elizabeth Fama, M.S.

Washington, DC
May 14, 2018
THE SUBJECTIVE EXPERIENCE OF INNER SPEECH IN APHASIA

Mackenzie Elizabeth Fama, M.S.

Thesis Advisor: Peter E. Turkeltaub, M.D., Ph.D.
Thesis Advisor: Rhonda B. Friedman, Ph.D.

ABSTRACT

All individuals with aphasia experience some level of anomia, or difficulty with naming and word finding, but many report that their internal word knowledge exceeds their spoken output, making comments like, “I know it but I can’t say it” or even more specifically, “I can say it in my head” (but not aloud). Such comments align with processing models of naming, which include multiple stages of lexical retrieval and post-lexical output processing. Given that anomia can arise from impairments at any one (or more) of the processing stages, it is conceivable that feelings of intact lexical knowledge reported by people with aphasia reflect success in the early stages, despite overt anomia resulting from impairment at a later stage.

In this dissertation, I examine the specific experience of “inner speech” (IS), defined in this context as a sense of being able to say a word in one’s head, regardless of subsequent ability to say that word aloud. Our hypotheses are that (1) the subjective experience of IS in aphasia reflects the status of lexical (phonological) retrieval and (2) IS itself does not rely on post-lexical stages of naming, so a failure of naming following successful IS results from deficits at the level of post-lexical output processing. Two approaches are used to elicit self-reports of IS from individuals with aphasia: an interview-based approach regarding the general experience of IS and an item-by-item approach in the context of silent naming. Self-reported IS is then compared to structural brain imaging and objective language measures related to the mental processes of naming (e.g., tasks emphasizing retrieval vs. output, error patterns, and psycholinguistic
properties of words). The results of these studies consistently support the hypotheses, showing that self-reported IS relates to measures of phonological retrieval but not articulatory processing.

Overall, the findings of these studies support a theory in which the experience of successful IS arises in conjunction with lexical retrieval and does not require articulatory processing. This work has potential implications for clinical decision-making in the treatment of anomia as well as for our general scientific understanding of the nature of IS.
ACKNOWLEDGMENTS

First and foremost, I will be forever grateful to the individuals with aphasia who put forth the time and effort to participate in these studies and, more generally, have taught me about the value of patience, communication, and connection – without you, this work would not be possible. Thank you!

There are so many people who have supported me over the past five years (and beyond) and have helped me get to where I am today. Thank you to my thesis mentors Peter Turkeltaub and Rhonda Friedman, for making me a stronger researcher and a better thinker. You have shaped my career path and my life in immeasurable ways and I hope (and plan) to never stop learning from you.

Throughout my thesis work, I have received input and feedback from my committee members, who have deepened and broadened my thinking in important ways: Elissa Newport, Michael Ullman, Michelle Harris-Love, and Jamie Reilly. I am thankful to have worked alongside all current and previous members of the Cognitive Recovery Lab and Center for Aphasia Research and Rehabilitation, with special gratitude to William Hayward, Sarah Snider, Elizabeth Lacey, Kelly Michaels, and Mary Henderson.

For the past decade, I have been part of a community of speech-language pathologists who have taught me to value creativity, learning, and compassion. Thank you to my colleagues at MedStar National Rehabilitation Hospital and the Stroke Comeback Center. Thank you especially to Anne H. Charity Hudley, Maura Silverman, and Amy Georgeadis, three long-term mentors who each instilled in me a deep enthusiasm for what we do, which continues to move me forward every step of the way.
I am so grateful to my amazing family, including the best parents in the world Ron and Carol (and now Steve and Deb too). You have been a constant source of support to me as I have worked toward this goal, just as you have been at all the stages of my life. To my chosen family, Meagan, Samantha, Jacqueline, Katherine, Jennifer, Sara, and Gloria: thank you, thank you for always lifting me up in ways I didn’t even know I needed and keeping me laughing every step of the way.

And finally, to my husband Matthew: I wouldn’t be where I am without you and I am so incredibly lucky to have you by my side. I can’t wait to start this next chapter with you.

This work was directly supported by funding from the NIH/NIDCD grants F31DC014875 and R03DC014310, with additional support through the ASHFoundation New Century Scholars Doctoral Scholarship.
# TABLE OF CONTENTS

## CHAPTER

<table>
<thead>
<tr>
<th>1</th>
<th>Introduction ........................................................................................................... 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>General introduction .............................................................................................. 1</td>
</tr>
<tr>
<td>1.2</td>
<td>Aphasia .................................................................................................................. 2</td>
</tr>
<tr>
<td>1.3</td>
<td>Models of word retrieval and speech production .................................................... 4</td>
</tr>
<tr>
<td>1.4</td>
<td>Self-monitoring ..................................................................................................... 10</td>
</tr>
<tr>
<td>1.5</td>
<td>Model of naming for the current studies ................................................................. 14</td>
</tr>
<tr>
<td>1.6</td>
<td>Model-based approaches to understanding anomia ..................................................... 15</td>
</tr>
<tr>
<td>1.7</td>
<td>Inner speech .......................................................................................................... 22</td>
</tr>
<tr>
<td>1.8</td>
<td>Feasibility of using self-report in aphasia ................................................................. 30</td>
</tr>
<tr>
<td>1.9</td>
<td>Overview of the dissertation .................................................................................... 33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates ................................................................. 38</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction .......................................................................................................... 38</td>
</tr>
<tr>
<td>2.2</td>
<td>Materials and Methods ......................................................................................... 46</td>
</tr>
<tr>
<td>2.3</td>
<td>Results .................................................................................................................... 53</td>
</tr>
<tr>
<td>2.4</td>
<td>Discussion .............................................................................................................. 56</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusions .......................................................................................................... 64</td>
</tr>
<tr>
<td>2.6</td>
<td>Acknowledgements ............................................................................................... 64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>The subjective experience of inner speech in aphasia is a meaningful reflection of lexical retrieval ......................................................................................... 71</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction .......................................................................................................... 71</td>
</tr>
<tr>
<td>3.2</td>
<td>Method .................................................................................................................... 74</td>
</tr>
<tr>
<td>3.3</td>
<td>Results .................................................................................................................... 82</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion .............................................................................................................. 86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4</th>
<th>Self-reported inner speech relates to individual differences in phonological retrieval in people with aphasia ................................................................. 104</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction .......................................................................................................... 104</td>
</tr>
<tr>
<td>4.2</td>
<td>Method .................................................................................................................... 105</td>
</tr>
<tr>
<td>4.3</td>
<td>Results .................................................................................................................... 112</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion .............................................................................................................. 114</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions .......................................................................................................... 121</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Discussion .............................................................................................................. 127</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Summary of main findings .................................................................................... 127</td>
</tr>
<tr>
<td>5.2</td>
<td>Implications for self-monitoring of inner speech .................................................... 129</td>
</tr>
<tr>
<td>5.3</td>
<td>Alternative interpretations of the findings ............................................................. 130</td>
</tr>
<tr>
<td>5.4</td>
<td>Practical considerations regarding the use of subjective reports in individuals with aphasia ................................................................. 134</td>
</tr>
<tr>
<td>5.5</td>
<td>Limitations and unsuccessful elements of the studies ................................................. 140</td>
</tr>
<tr>
<td>5.6</td>
<td>Open questions and future directions ................................................................... 145</td>
</tr>
<tr>
<td>5.7</td>
<td>Conclusions .......................................................................................................... 153</td>
</tr>
</tbody>
</table>

## BIBLIOGRAPHY ........................................................................................................ 157
# LIST OF FIGURES

1.1 Aphasia subtypes ........................................................................................................ 36
1.2 A general model of naming ........................................................................................ 37

2.1 Simplified model of naming ........................................................................................ 65
2.2 Picture support for the concepts of sIS and IwW ....................................................... 66
2.3 sIS and other anomia ratings ...................................................................................... 67
2.4 Lesion-symptom mapping results ................................................................................ 68

3.1 Day to day variability of scores on IS Report and Spoken Naming ......................... 97
3.2 Item-level performance for words reported as successful or unsuccessful IS across other tasks .................................................................................................................. 98

4.1 Results of factor analysis ............................................................................................ 123

5.1 Response scales used during the interview ............................................................... 154
5.2 Interview picture support: page one (anomia vs. successful naming) ..................... 154
5.3 Interview picture support: page two (sIS vs. word retrieval failure) ......................... 156
LIST OF TABLES

2.1 Relationship between anomia ratings and behavioral language measures ............... 69
2.2 SVR-LSM results ........................................................................................................ 70

3.1 Word features for the 60-item and 120-item stimulus lists ....................................... 99
3.2 Task scores for general language measures ................................................................. 100
3.3 Word feature comparison across tasks ..................................................................... 101
3.4 Results of generalized linear mixed effects models predicting task performance .... 102
3.5 Task scores for general language measures: reliability analysis ............................. 103

4.1 Average performance on the language battery (N=53) ........................................... 124
4.2 Correlational analysis comparing logit-transformed IS Report, Spoken Naming, and factor scores ................................................................................................................. 125
4.3 Correlational analysis comparing IS Report and Spoken Naming to naming error types ......................................................................................................................... 126
CHAPTER 1

Introduction

1.1 General introduction

At least one-third of individuals who experience a stroke will develop aphasia, a language disorder that affects a person’s ability to communicate and has long-term effects on quality of life (Berthier, 2005; Engelter et al., 2006; Hilari et al., 2010). There is considerable variability across individuals in terms of the specific language deficits associated with aphasia, but a common thread among all individuals with the disorder is anomia, or a difficulty with naming and word-finding (Goodglass & Wingfield, 1997; Kohn & Goodglass, 1985; Laine & Martin, 2006; Maher & Raymer, 2004). In models outlining the mental process of naming, it is widely accepted that successful naming requires multiple processing stages and that anomia can arise from specific impairments at any one (or more) of these stages, which include semantic retrieval, phonological retrieval, and post-lexical/pre-articulatory output processes (Dell & O’Seaghdha, 1992b; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Indefrey & Levelt, 2004; Levelt, Roelofs, & Meyer, 1999; Walker & Hickok, 2015). Given the multi-step nature of naming, it is easily conceivable that success in the early stages of naming could give rise to a feeling of intact internal knowledge of the target word, despite overt anomia resulting from impairment at a later stage. In fact, in an interesting clinical phenomenon, many individuals with aphasia report that they have more knowledge of words in their head than they are able to demonstrate through speech, making comments like, “I know it but I can’t say it” (Blanken, Dittmann, Haas, & Wallesch, 1987; Martin & Dell, 2007). More specifically, individuals with aphasia commonly report that they are able to say words in their head that they cannot successfully say aloud (Fama, Hayward, Snider, Friedman, & Turkeltaub, 2017; Hayward, Snider, Luta, Friedman, &
Turkeltaub, 2016). These self-reports can be interpreted as a reflection of successful “inner speech” (IS) despite a failure of overt speech. This dissertation will examine the nature of self-reported IS and its relationship to objective measures of the various processes involved in word retrieval and production. In this introductory chapter, I will provide general background information on the major topics relevant to this work.

1.2 Aphasia

Language is a complex, symbolic system used by humans for functional communication. Aphasia, an acquired disorder of language, is equally complex, comprising a wide constellation of possible symptoms including difficulties in word-finding, producing grammatical sentences, understanding spoken language, reading, and writing (Berthier, 2005; McNeil & Pratt, 2001). Aphasia commonly results from left hemisphere stroke (Brady, Kelly, Godwin, & Enderby, 2012; Palle Møller Pedersen, Vinter, & Olsen, 2004; P M Pedersen, Jørgensen, Nakayama, Raaschou, & Olsen, 1995), as these language functions are subserved by regions throughout the dominant hemisphere, which is most often the left (Knecht et al., 2000; Pujol, Deus, Losilla, & Capdevila, 1999; Szafiarski et al., 2002). Historical models identified the essential areas to be Broca’s area in the frontal lobe and Wernicke’s area in the temporal lobe, connected by the arcuate fasciculus (a white matter tract), but more recent research demonstrates that the language network extends beyond these structures, including more widespread areas of the frontal, temporal, and parietal lobes in addition to subcortical gray matter structures (Fedorenko, Nieto-casta, & Kanwisher, 2012; Hickok & Poeppel, 2007; Price, 2012). A person with post-stroke aphasia is likely to experience some difficulty with language processing that will persist throughout his or her lifetime (Berthier, Pulvermüller, Dávila, Casares, & Gutiérrez, 2011;
Gottesman & Hillis, 2010), despite measurable improvement from the initial language impairment both through spontaneous recovery and in response to behavioral therapy with a speech-language pathologist (Brady et al., 2012; Lendrem & Lincoln, 1985; Anastasia M Raymer et al., 2008; Robey, 1998).

Historically, individuals with aphasia have been categorized into one of multiple subtypes of aphasia, based primarily on fluency of output, auditory comprehension, and repetition ability (Figure 1.1) (Goodglass, Kaplan, & Barresi, 2001; Helm-Estabrooks & Albert, 2004). The most common subtypes studied in the literature include the following:

- **Broca’s aphasia**: halting, agrammatic output (non-fluent), good comprehension and poor repetition; relatively preserved awareness of deficits
- **Wernicke’s aphasia**: fluent output containing semantic and/or phonological jargon, poor comprehension and poor repetition; generally poor awareness of deficits
- **Conduction aphasia**: fluent output, good comprehension and poor repetition; generally preserved awareness of deficits, as evidenced by frequent attempts to correct spoken errors that are predominantly phonologically related to the target

Additionally, some studies include individuals with global aphasia, who produce little spoken output and show impairments across all language functions, as well as individuals with anomic aphasia, who have impairments primarily in naming and word-finding. As mentioned above (Section 1.1), anomia is a common symptom across all subtypes of aphasia.

This system for categorizing aphasia subtypes is far from perfect, as there are some patients who do not easily fit one subtype and this system glosses over the more nuanced and specific aspects of language processing that may differ across patients. Although complete reliance on these subtypes can be limiting, it remains a useful vocabulary for interpreting prior
literature that often assumes familiarity with this diagnostic scheme. Since these terms will be used occasionally to describe patients elsewhere in the dissertation, they are presented here primarily as a point of reference. A more modern approach to understanding an individual’s deficits is to determine the specific processes disrupted by the stroke (or other brain injury) that lead to a person’s impairments in naming, grammar, comprehension, and other language abilities (Lacey, Skipper-Kallal, Xing, Fama, & Turkeltaub, 2017; Mirman et al., 2015). This more precise approach relies on models describing the processing steps necessary for performing various language functions.

1.3 Models of word retrieval and speech production

The set of studies presented here will interpret inner speech (IS) as it relates to the general process of spoken naming, so it is important to consider the existing major processing models and neuroanatomical models of naming. There are several caveats to be mentioned here. First, some of these models encompass the entire process of speech production, whereas others are focused specifically on word retrieval (a term which will be used interchangeably with lexical retrieval and lexical access). Second, the models discussed in this section were chosen as the most relevant models of naming, but they do only represent a subset. Finally, while many of these models touch on the nature of self-monitoring, a discussion of the specifics of self-monitoring will be reserved until the next section (1.4). After describing the relevant models and their accounts of self-monitoring, we will present a simplified, unified model that includes the most consistent features of the existing models. This model will serve as the theoretical context for our experimental work.
1.3.1 Processing models

1.3.1.1 The spreading activation model

In 1986, Dell proposed a two-step spreading activation model of word retrieval, which has been refined and expanded through decades of work examining speech production at the word- and sentence-level (Dell, 1986; Dell & O’Seaghdha, 1992b; Dell et al., 1997; Foygel & Dell, 2000). In this theory, there are three levels in the word retrieval network: semantic (concepts), lexical (lemmas, or abstract word forms), and phonological (sounds), each made up of many nodes. Word retrieval begins with a concept and continues via activation of the relevant nodes at the lexical level and then phonological level, through connections that are excitatory and bidirectional, ultimately leading to the target word (Dell, 1986; Dell et al., 1997). The spreading activation model successfully accounts for many aspects of the speech production of both healthy language users and those with aphasia. It is consistent with many of the natural phenomena observed in speech production, as demonstrated by evidence focused mostly on common speech errors in healthy speakers. Dell and colleagues later used the model to lend support to the continuity thesis, which proposes that speech errors in aphasia are not unlike slips of the tongue made by healthy speakers, since both are often related to the target word in meaning, sound, or both (Dell et al., 1997).

1.3.1.2 WEAVER++

In 1999, Levelt and colleagues summarized a decade of their work in the WEAVER++ model, a feed-forward, multistage theory of the process of spoken naming (Levelt, 1983, 1989; Levelt et al., 1999). Unlike other work in speech production which relied primarily on speech error data, the Levelt group based their theory primarily on reaction time experiments. The model encompasses the following stages: conceptual preparation, lexical selection,
morphological encoding, phonological encoding/syllabification, phonetic encoding, and finally articulation. It describes lexical selection as occurring via a statistical mechanism. The next process involves selection of the “morphological makeup, its metrical shape, and its segmental makeup” (p. 5, Levelt et al., 1999). Syllabification is described as a late process. The penultimate step is phonetic encoding, in which articulatory gestures are accessed – in part from the syllabary of “highly learned gestures” for common syllables in the language, which increases speech efficiency. The final step, articulation, is not described in detail (Levelt et al., 1999).

When Levelt’s WEAVER++ model is compared with Dell’s spreading activation model, both similarities and differences arise. In terms of similarities, both models purport the presence of three levels in the word retrieval network (concepts, lemmas, and word-forms) and a lack of inhibitory connections. Perhaps as a result of this similarity, both models can account for speech error data (e.g., anticipation errors, lexical bias) (Levelt et al., 1999). A primary difference, however, is that the Dell model is interactive in nature, with both feedforward and feedback mechanisms, whereas WEAVER++ is described as having no feedback except for monitoring of output (Dell & O’Seaghdha, 1992b; Dell et al., 1997; Levelt et al., 1999). Another distinction between the two theories is that WEAVER++ accounts for stages beyond word retrieval, including phonetic encoding (context-dependent sound processing) and articulation, both important stages of the speech production process.

1.3.1.3 Dissociating lexical and post-lexical phonology

Goldrick and Rapp (2007) do not propose a naming model per se, but contribute to the discussion of processing models of naming by specifying the existence of two levels of phonological representations: lexical and post-lexical (Goldrick & Rapp, 2007). At the lexical level, phonological representations are in a non-specific form, whereas post-lexical
representations are more fully specified, providing the necessary context for subsequent articulation. This appears consistent with prior models (e.g., WEAVER++) that draw a distinction between non-contextual phonological processing and context-based phonetic encoding (Levelt et al., 1999). By examining performance on naming and repetition tasks, Goldrick and Rapp illustrate the dissociation between individuals with impaired lexical processing who demonstrate poor naming ability but relatively intact repetition and individuals with impaired post-lexical processing who demonstrate comparable impairments across both naming and repetition (Goldrick & Rapp, 2007).

1.3.1.4 The semantic-lexical-auditory-motor model (SLAM)

The most recent of the prominent processing models of naming is the Semantic-Lexical-Auditory-Motor (SLAM) model of speech production (Walker & Hickok, 2015). It begins with Foygel and Dell’s 2000 semantic-phonological (SP) model, from the family of spreading activation models described in section 1.3.1.1, and builds on it through the application of motor control concepts via a hierarchical state feedback control model (HSFC) (Foygel & Dell, 2000; Hickok, 2012). The original SP model suggests that the two most salient features in successful word retrieval are the connection between semantics and lexical items (s weight) and between lexical items and phonology (p weight). The HSFC model contributes notions of motor control via three main characteristics: phonemes are dually encoded in the auditory and motor systems, there is a system of excitatory/inhibitory connections between these systems that is mediated by a sensorimotor translation area, and the system is organized hierarchically, where the auditory system manages acoustic targets and the motor system manages phonemic/phonetic targets (Hickok, 2012; Walker & Hickok, 2015). Thus, the SLAM model expands upon the SP model by proposing parallel lexical-auditory weights and lexical-motor weights rather than a single
phonological weight. The lexical-auditory weight is described as the primary route for naming and must be stronger than the lexical-motor weight. Auditory-motor weights are also relevant, drawing on the essential claim of an auditory-motor interface put forth by the dual-stream model (see section 1.3.2.2) (Hickok & Poeppel, 2007). The authors demonstrate that SLAM accounts for picture naming performance in individuals with aphasia equally well to the SP model, except in the case of conduction aphasia (i.e., low auditory-motor weights), in which SLAM explains the data better than SP in a comparative computational analysis (Walker & Hickok, 2015). The findings suggest that the application of motor control theory represents an important modification to two-level spreading activation models of word retrieval.

1.3.2 Neuroanatomical models

1.3.2.1 WEAVER++/LRM model

As with Dell’s spreading activation model, the WEAVER++ model has been refined over time. In 2004, Indefrey and Levelt used a meta-analysis approach to update their original model (which they refer to as LRM, for the authors’ initials) with the relevant brain regions and the time course of activation for each stage of lexical production (Indefrey & Levelt, 2004). The stages described map roughly onto the original model and the relevant brain structures are described as follows: conceptual preparation, lemma retrieval and selection (middle left middle temporal gyrus), phonological code retrieval (posterior middle/superior temporal gyri and possible left anterior insula/right supplementary motor area (SMA)), syllabification (left posterior inferior frontal gyrus (IFG)), phonetic encoding, and articulation (bilateral sensorimotor cortex, cerebellum, and thalamus along with the right SMA, midbrain. In 2011, Indefrey updated these findings in two primary ways: by giving evidence for a non-specific role of the inferior parietal cortex in word production and by dividing the role of the IFG into a dorsal region for
sylabification and a more ventral region for phonetic encoding/articulatory planning (Indefrey, 2011; Papoutsi et al., 2009).

1.3.2.2 The dual-stream model

In 2007, Hickok and Poeppel put forth a dual-stream model of language processing that involves two parallel sets of processing stages: a ventral stream for mapping sound to meaning and a dorsal stream for mapping sound to action (here, speech motor output) (Hickok & Poeppel, 2007). As this model was developed primarily to illustrate how we process auditory speech input, rather than as a model of word retrieval and speech output, we will consider only a few aspects of the model that are particularly relevant here. The dorsal stream is significantly lateralized to the left hemisphere (unlike their more bilateral ventral stream); it begins with bilateral posterior STS and connects to left temporal-parietal junction (here named area Spt) and left frontal areas. The authors cite the syndrome of conduction aphasia, in which damage to the temporal-parietal junction spares comprehension but causes deficits in speech production (see Figure 1.1), as evidence for the existence of an auditory-motor interface, which they localize to area Spt (Buchsbaum et al., 2011; Hickok & Poeppel, 2007). The dorsal stream ends with the articulatory network, including left posterior IFG, premotor areas, and anterior insula. Because of the nature of the dual-stream model as a more general model of speech processing, it does not address the nuances of the psycholinguistic aspects of speech production, but it is largely consistent with the Indefrey and Levelt (2004) model describing the brain regions underlying these processes (Indefrey & Levelt, 2004). Hickok and Poeppel add the important concept of auditory-motor integration, which appears to be essential for a full understanding of speech production itself as well as the interaction between production and perception (Hickok & Poeppel, 2007). The SLAM model described above (section 1.3.1.4) is a computational model
based on this general dual stream model and it updates the anatomy proposed here, accounting more specifically for speech output processing (Walker & Hickok, 2015). Area Spt serves as the interface between auditory syllable targets (in the superior temporal gyrus and superior temporal sulcus) and motor syllable programs (in BA44, pars opercularis of the IFG). These areas interact with the somatosensory phoneme targets (supramarginal gyrus and primary sensory cortex) and motor phonemes (ventral premotor cortex and primary motor cortex) for articulation (Walker & Hickok, 2015).

1.4 Self-monitoring

As mentioned above, the majority of processing and neuroanatomical models of speech production address the concept of self-monitoring, the essential ability of language users to identify and thus correct potential errors in speech output. The proposed theories fall into two main groups: comprehension-based and production-based monitoring systems. An important consistency across many models of self-monitoring is the inclusion of both internal and external monitoring systems; the internal system is fundamental to our examination of the subjective experience of IS, as it allows for monitoring prior to spoken output. Understanding the prevailing models of self-monitoring will improve our ability to interpret the experience of IS in aphasia by clarifying the underlying mental mechanisms by which self-reported IS could be meaningful in this population.

1.4.1 Comprehension-based monitors

1.4.1.1 Perceptual loop theory

In the perceptual loop theory, Levelt proposed that the auditory comprehension system is responsible for self-monitoring during speech production (Levelt, 1983, 1989). Specifically, he
suggests that the comprehension system receives input via two loops that provide unique information: first, an internal loop sends information from the phonological or phonetic processing stages, then the auditory loop sends the output of articulation, the acoustic speech signal (Indefrey & Levelt, 2004). The internal loop was motivated by error data indicating that speakers can self-interrupt in order to repair errors, indicating some ability to monitor internal speech prior to spoken output. In Levelt’s early work, the specific nature of the information carried within the internal loop includes articulatory gestures formulated just prior to speech (Levelt, 1983, 1989), but later work suggests that the internal loop monitors a more abstract phonological form (Wheeldon & Levelt, 1995). As this model was refined, Levelt and colleagues suggested the neural substrate to be the bilateral superior temporal gyri, consistent with neuroanatomical models of auditory comprehension (Indefrey & Levelt, 2004). Thus, a primary appeal of the perceptual loop theory is its parsimonious nature: the same system by which we understand the speech of others is utilized for monitoring speech of our own.

1.4.1.2 Criticisms of the perceptual loop theory

The perceptual loop theory has remained an influential theory of self-monitoring since its first inception in the early 1980s; however, many critiques of this theory have emerged over time based on both behavioral and neuroanatomical findings. In behavior, bidirectional dissociations have been demonstrated between self-monitoring and auditory comprehension (Maher, Gonzalez Rothi, & Heilman, 1994; J. Marshall, Robson, Pring, & Chiat, 1998; Nickels & Howard, 1995b). Furthermore, individuals with poor self-monitoring have shown relatively preserved ability to detect errors in other speakers (Maher et al., 1994; J. Marshall et al., 1998). Some authors have raised issue with the idea of simultaneous monitoring of internal and external speech in which the auditory comprehension system must manage both types of input (Huettig & Hartsuiker,
Finally, imaging findings have suggested that the auditory comprehension system does not act independently in the performance of speech monitoring. In the 2011 update to the neuroanatomical context of the WEAVER++/LRM model, Indefrey himself cites evidence that the auditory cortex interacts with many other structures in the performance of self-monitoring, including the cingulate cortex, the supplementary motor area, bilateral motor areas, and non-cortical structures including the cerebellum, the thalamus and the basal ganglia (Indefrey, 2011). Taken together, these critiques suggest a more nuanced basis of self-monitoring than the original perceptual loop theory.

1.4.2 Production-based monitors

1.4.2.1 General theories of production-based monitoring

In direct contrast to the perceptual loop theory, there is another set of theories that assert that it is the production system, not the comprehension system, that is responsible for error monitoring. The first wave of production-based monitors proposed a system by which the actual output of production is compared to the target/correct output (Laver, 1980; Schlenck, Huber, & Willmes, 1987). Laver (1980) proposed that there is a monitor at each stage of the production process that makes this comparison prior to the information moving to the next stage, ending in a sensory loop that parallels the auditory loop from the perceptual loop theory. A criticism of this type of theory is that a reduplication of information is required for comparing actual and correct output (MacKay, 1987). In node structure theory, which was devised to avoid such reduplication, monitoring occurs based on patterns of information flow with special sensitivity to unfamiliar patterns, wherein the speaker gains awareness of an error when novel, unexpected patterns occur (MacKay, 1987). A similar theory later proposed that monitoring occurs via comparison of activation that the production system sends forward to the amount of feedback it receives in
return (Postma & Kolk, 1993). Most recently, a motor-specific approach to a production-based
monitor proposed a direct motor-sensory link responsible for self-monitoring, where motor
speech regions produce efference copies that allow for internal monitoring of planned

1.4.2.2 Conflict-based theory of monitoring

Nozari et al. (2011) propose a novel type of production-based monitor that is rooted in
domain-general theories of conflict monitoring (Nozari, Dell, & Schwartz, 2011). In their model,
a domain-general monitoring system detects errors by identifying response conflict, i.e., the
amount of competition arising from words related to the target. They test this theory by
simulating response data using the two-step interactive SP model, which includes mapping from
concepts to abstract lemma and from abstract lemma to phonemes (Foygel & Dell, 2000). They
demonstrate that the conflict-based monitoring theory is sensitive to overall error rates as well as
the types of errors expected to result from conflict at either the first or second step of the word
retrieval process, i.e. at the time of production (not after a response has been generated) (Nozari
et al., 2011). Consistent with general models of conflict monitoring, the authors cite the anterior
cingulate cortex as the most likely neuroanatomical substrate for error monitoring in speech
production.

1.4.2.3 Criticisms of production-based monitors

As with the perceptual loop theory, the production-based monitors have been subject to
some criticism in the literature. Concerns about these theories include general questions of
whether they are detailed enough to be thoroughly tested (see Nozari et al., 2011 for a
discussion). Some of the production-based monitoring theories assume a level of shared
resources between the production and comprehension systems (e.g., node structure theory),
which leave them vulnerable to the same criticisms of the observable dissociations between error-monitoring and comprehension (see section 1.4.1.2 above). In the Laver model, it is proposed that monitoring at each stage of the process must be complete before production continues, but later theorists have called attention to that the fact that this seems detrimental to fluency in production (Laver, 1980; Postma, 2000). The conflict-monitoring approach claims to address some of these major concerns, but it too may not stand alone as a complete model of self-monitoring. As the authors themselves note, there must be some role of the perceptual system in monitoring, as there is empirical evidence for a decrease in error detection with auditory masking (Lackner & Tuller, 1979; Nozari et al., 2011). In light of these concerns, some type of hybrid system for self-monitoring that combines the strengths of both comprehension- and production-based monitors, including a role of domain-general conflict monitoring, may be the most appropriate. The specific application of theories of self-monitoring to our examination of IS in aphasia will be discussed in section 1.8.2 and revisited throughout the dissertation.

1.5 Model of naming for the current studies

The models described above each offer important contributions to our understanding of the mental process of naming. While it is not within the scope of this dissertation to propose my own thorough model, I do take this opportunity to join together the most consistent aspects of previous models into a single, general model of naming, while remaining agnostic about certain details that are not directly examined within these studies (Figure 1.2). Although this dissertation is focused on IS, we often consider IS in the context of the subsequent success or failure of spoken output; as such, it is essential for the model to account for both retrieval and production. The model begins with a feature that is consistent across models: lexical access
processes that include mapping of the semantic representation onto an abstract word-form, or lemma, and then mapping of the lemma onto the lexical phonological representation. The ensuing post-lexical output processes incorporate the notion of a more specified post-lexical phonological representation (Goldrick & Rapp, 2007) and conclude with motor programs for articulation, perhaps generated via the sensorimotor integration mechanism proposed by Hickok and Poeppel (2007) and Walker and Hickok (2015). Self-monitoring occurs internally at the level of the lexical phonological representation, consistent with most prior models that identify a mechanism for internal monitoring, although we remain agnostic as to whether the monitoring system relies on perception, production, conflict, or all of these mechanisms (Levelt et al., 1999; Nozari et al., 2011).

1.6 Model-based approaches to understanding anomia

The preceding sections present the essential aspects of naming, including the mental stages of processing and the neuroanatomical structures that underlie our ability to successfully retrieve and produce a word while also monitoring our own success in doing so. This next section is focused on anomia, a failure of naming and word-finding, which is typically considered to be the hallmark deficit of aphasia. The processing models of word retrieval and production described in the previous sections have been integral to the understanding of anomia. One universal aspect of naming models is that they require more than one stage of processing in order for speech production to occur, a feature that naturally lends itself to the notion that the process can fail in more than one way.

1.6.1 Subtypes of anomia

There is wide agreement that anomia can result from impairment at any one (or more) of the processing stages required for successful naming. In the literature, several subtypes of
anomia have been identified and labeled, based on some consistent aspects of processing and 
neuroanatomical models of word retrieval and production. A deficit at the level of 
semantic/conceptual representation or the abstract lemma is labeled “semantic anomia” (Laine & 
Martin, 2006). A deficit at the level of the lexical phonological representation, in accessing the 
specific sound-based information needed for production, is labeled “phonological anomia” or 
“word form anomia” (Laine & Martin, 2006). A deficit that lies in the connection between 
semantics and phonology, without specific impairment at either level, is labeled “classical 
anomia” (Geschwind, 1967; Lambon Ralph, Sage, & Roberts, 2000). More recently, researchers 
have also recognized deficits in post-lexical phonological output processing, i.e., disordered 
phoneme assembly, as another subtype of anomia (Howard, 2000; Laine & Martin, 2006). 
Failures at these later stages of the naming process can be difficult to distinguish from errors due 
to apraxia of speech, an acquired motor speech disorder characterized by poor motor planning 
and sequencing for articulation (McNeil, Robin, & Schmidt, 2009; Ogar et al., 2006).

Proponents of these anomia classifications do not claim them to be definitive nor entirely 
discrete, acknowledging that impairments in early stages can affect the entire naming process 
and that overt anomia can even arise from deficits not typically considered essential to naming 
(for example, attention impairments or dysarthria) (Laine & Martin, 2006). In the context of this 
dissertation, it is not essential to identify one specific set of anomia subtypes, but rather to 
appreciate the heterogeneity of specific causes of anomia across individuals. Also, since many 
individuals with aphasia do not have pure deficits confined to a single stage of the naming 
process, we treat their deficits at the various stages as continuous variables.

1.6.2 Model-based diagnostic approaches

Researchers who wish to identify the level(s) of processing that are impaired in a person
with anomia rely on measurable, overt behaviors as a window into the internal mental processes of naming. There are multiple methods by which anomia diagnosis has been attempted in the prior literature, each of which has benefits and disadvantages. Here we will consider analysis of naming errors, analysis of word features that affect naming ability, and the effects of cueing.

1.6.2.1 Analysis of naming errors

Anomia is frequently evident during confrontation naming tests, where an individual is asked to name a visually presented picture or object. While this is not the most natural form of language use, patients’ confrontation naming performance correlates with conversational language in aphasia (Goodglass & Wingfield, 1997), so these tasks are a common source of evidence for model-based treatments. Errors on these tasks may include words related to the target word in meaning (i.e., semantic errors: “pillow” for “bed”) or in sound (i.e., phonological errors: “beg” or “bet” for “bed”) (Kohn & Goodglass, 1985), or both. These error patterns are often used to inform model-based treatments. For instance, an error that is phonologically related to the target is often taken as evidence for failure at the level of lexical phonological representations (M. F. Schwartz, Wilshire, Gagnon, & Polansky, 2004) or during phonological assembly, an essential aspect of post-lexical processing (Howard, 2000).

1.6.2.2 Analysis of word features affecting naming ability

There are many psycholinguistic variables of words that affect how efficiently and effectively they are processed, with evidence from both healthy individuals and individuals with language impairments (e.g., aphasia). The nature of the words that present particular difficulties to a person with anomia may suggest impairment at specific processing levels. Here, I will discuss the word features that will be examined in this dissertation in the context of self-reported IS.
1.6.2.2.1 Lexical frequency

The word feature that has been examined most often is frequency, a measure of how often a particular word is used in either written or spoken language. The first study on word frequency demonstrated that it was strongly related to picture-naming response latencies in healthy language users, wherein more frequent words were produced more quickly (R. C. Oldfield & Wingfield, 1965). Although the evidence has been mixed, this frequency effect has been shown to be relevant in the context of aphasia, where naming accuracy and error types can be predicted by word frequency, an effect that derives from the word retrieval process prior to production (Butterworth, Howard, & McLoughlin, 1984; Kittredge, Dell, Verkuilen, & Schwartz, 2008). Some evidence localizes the effect specifically to phonological retrieval, due to the fact that low- and high-frequency homophones can be translated between languages with equal speed (Jescheniak & Levelt, 1994) and that low-frequency words are more likely than high-frequency words to elicit phonologically-related errors (Dell, 1990; M. F. Schwartz et al., 2004). Other work indicates that frequency effects may also be relevant earlier in the word retrieval process, in retrieving the abstract word-form from the concept (Kittredge et al., 2008).

There are a few caveats to the robust relationship between frequency and lexical retrieval that have been reported in the literature. First, there is some evidence that frequency does not have a strong, independent effect on naming in some individuals with aphasia; one such study shows that in a multiple regression analysis including other, related word variables like age of acquisition and familiarity, frequency is not a significant predictor of naming (Nickels & Howard, 1995a). Also, there is some evidence from non-naming tasks that frequency can affect word production, not just word retrieval, during delayed oral reading of single words (Balota &
Chumbley, 1985) and during spoken word repetition (Nozari, Kittredge, Dell, & Schwartz, 2010).

1.6.2.2.2 Age of acquisition

In examining the role of frequency and other psycholinguistic variables on word retrieval and production, some studies argue that there is another word feature that supersedes the effects of frequency, specifically arguing that the age at which a word is learned during initial language acquisition (age of acquisition (AoA)) is more predictive of naming ability (Brysbaert & Ellis, 2016; Jescheniak & Levelt, 1994; Morrison, Ellis, & Quinlan, 1992; Snodgrass & Yuditsky, 1996). AoA is often significantly correlated with word frequency and is also associated primarily with effects at the level of word retrieval. It is also difficult to objectively determine, so the most common way in which it is measured is via individual self-report (Brysbaert & Ellis, 2016; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Morrison et al., 1992).

1.6.2.2.3 Word length

The length of a word is another common psycholinguistic variable examined in this context, with length measured either in letters (graphemes), individual sounds (phonemes), or syllables. Length effects have been identified both in healthy adults and in individuals with aphasia, with shorter words being produced more quickly and accurately in confrontation naming (Levelt et al., 1999; Morrison et al., 1992; Nickels & Howard, 2004). The word length effect is not universally found in individuals with anomia and is most often noted in individuals with post-lexical processing impairments (Butterworth, 1992; Howard & Gatehouse, 2006; Nickels & Howard, 1995a). There is, however, some work indicating that the length effect is not purely a post-retrieval phenomenon, since increasing word length can relate to errors with low phonological overlap with the target word (remote errors), which are associated with retrieval
failure (M. F. Schwartz et al., 2004). Similarly, there is some evidence for a reverse length effect, where longer words are more often successfully named than short words, in some individuals with impairments in semantics and in accessing the phonological output lexicon, both of which are aspects of lexical retrieval, not post-lexical processing (Best, 1995; Howard & Gatehouse, 2006). Taken together, the evidence for length effects on naming suggests that this effect may be most robust at the level of post-lexical processing but is also relevant at earlier stages of the word-finding process.

1.6.2.2.4 Articulatory complexity

A final psycholinguistic variable to be considered here is articulatory complexity. Although it is quantified in different ways across the literature, articulatory complexity consistently represents a measure of the difficulty required for spoken production of a given word (Baldo, Wilkins, Ogar, Willock, & Dronkers, 2011; Stoel-Gammon, 2010). This psycholinguistic feature is uniquely related to post-retrieval production processing. There is widespread evidence for an increase in speech errors for words requiring more complex articulation, a difficulty that is often associated with acquired apraxia of speech, a disorder at the level of motor articulatory planning (Boyle et al., 2006; Haley, Jacks, de Riesthal, Abou-Khalil, & Roth, 2012; Strand, Duffy, Clark, & Josephs, 2014). Furthermore, in individuals with language impairments, the effects of articulatory complexity can be observed in the context of repetition tasks in addition to naming (Romani & Galluzzi, 2005).

1.6.2.3 Analysis of the effects of cueing on naming ability

Another potential method for diagnosing an individual’s specific type of anomia is to look at their responsiveness to cues during naming tasks. Individuals with any type of anomia can benefit from spoken phonological cues, but individuals with semantic anomia are more
susceptible to miscues, e.g., giving a cue of “d-“ when the target is “cat” may lead a patient with semantic anomia to say “dog” because their semantic representation is not strong enough for them to recognize the error (Howard & Gatehouse, 2006; Julie L. Wambaugh et al., 2001).

1.6.3 Limitations of these approaches to model-based diagnosis

As identified by Howard & Gatehouse, the preceding methods of identifying the level of breakdown in anomia have several limitations (Howard & Gatehouse, 2006). For example, in using error patterns to understand anomia, the simplest approach assumes that semantic errors reflect a semantic deficit and phonological errors reflect a phonological deficit (either in word form or post-lexical processing). This approach is inadequate, as a speaker may make different types of errors for different words or may even make different types of errors for the same word on different occasions. Many case studies of model-based treatment do show that phonological errors typically indicate that there has been at least some degree of phonological access to the target word, but this correlation may not be bidirectional, i.e., an individual who has at least partial access to the phonological form of the word (and thus intact semantics) may make an error that is either phonologically related, semantically related, neither, or both (Dell et al., 1997; Kendall, Pompon, Brookshire, Minkina, & Bislick, 2013; A. Raymer & Gonzalez Rothi, 2001; Wilshire, 2002). As described above (section 1.6.2.2), there is also a challenge in understanding anomia via the effects of psycholinguistic variables (e.g., frequency, length), since many of the variables’ effects have been shown to indicate impairments at more than one level of the naming process.

In general, previous research shows that the deficits suggested by error type patterns, effects of psycholinguistic variables, and response to cueing may not be consistent with the patient’s entire language profile (Goodglass & Wingfield, 1997). Thus, a richer language
assessment may be more useful in diagnosing anomia (Howard & Gatehouse, 2006). When naming performance is investigated in the context of other language tasks, such as comprehension, repetition, and overall fluency, this may reveal more general deficits in semantics or phonology that are not restricted to naming skills. This approach may be the most thorough but is also the most time-consuming. Even this in-depth approach to anomia diagnosis relies on the use of overt language behavior to understand an internal mental process, as is the case for the majority of scientific inquiry into language processing both in healthy individuals and in individual with aphasia. In this dissertation, we aim to combine objective assessment with subjective input to obtain a fuller characterization of IS, the phenomenon of interest.

1.7 Inner speech

Although this dissertation is focused on the experience of IS in aphasia, the phenomenon is relevant beyond this specific population. In this section, we will consider the literature on IS in healthy speakers before turning to the prior research on IS in aphasia.

1.7.1 Inner speech in healthy language users

The relationship between thought and language has been a topic of philosophical inquiry since the ancient Greek concept of logos, representing a unity of thought and word. The specific phenomenon of IS is familiar to most language users and has been studied scientifically since at least the late 19th century (see Sokolov, 1972 for a review). Early theorists suggested that IS serves as the basis of thinking and that IS requires the same articulatory movements required for overt speech, albeit to a lesser degree (Watson, 1913). This strong position regarding the role of articulation in IS was later dispelled by a case study in which temporary total paralysis, induced by curare in a healthy adult, did not affect the participant’s “consciousness, memory, or
sensorium” (p. 11, Smith, Brown, Toman, & Goodman, 1947), but an ongoing debate remains as to the extent to which articulation may be activated during IS (see Section 3.4.2 for further discussion).

Over the past century, the phenomenon of IS has often been studied in the context of other mental processes. There is a significant literature on the role of IS in various aspects of cognition, including prominent work on the role of IS in working memory (A. Baddeley & Hitch, 1974) and language learning (Vygotsky, 1962). Since then, IS has been associated with a wide range of other mental tasks, including logical reasoning, reading, and executive functioning abilities such as planning, decision-making and task-switching, among many others (for recent reviews, see Perrone-Bertolotti et al., 2014 and Alderson-Day and Fernyhough, 2015). In this section, I will present several approaches to defining and studying IS as well as a summary of findings regarding the neural bases of the phenomenon.

1.7.1.1 Definitions of inner speech

Importantly, there are at least two ways define IS. The first is characterized as the ability to generate language silently in one’s head, or in other words to develop “an auditory-articulatory image of speech without uttering a sound” (Levine, Calvanio, & Popovics, 1982). This definition emphasizes the viewpoint of the person who is experiencing IS and possibly using IS in the service of a greater purpose, such as verbal rehearsal, self-reflection, or executive functions such as planning or decision-making. Within this definition, some researchers draw a line between IS that is deliberately generated and more general verbal mind-wandering (Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Løvenbruck, 2014). While there is agreement that the experience of IS is most readily accessible to the person who is experiencing it, a competing definition focuses on the more observable aspects of IS: “the objectively measurable ability to
appreciate the auditory-articulatory structure of speech irrespective of its meaning” (Levine et al., 1982). Theorists focused on this more objective approach are often interested in demonstrable language behaviors that can be performed based on IS, such as syllable counting or silent judgment of homophones or rhymes.

1.7.1.2 Approaches to the study of inner speech

In healthy language users, IS has been scientifically examined in the context of both definitions. First, to facilitate understanding of the experience of IS as verbal thinking or self-directed internal language, researchers have designed questionnaires to elicit self-reports about IS in healthy adults (Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016; Morin, Uttl, & Hamper, 2011). These studies have contributed subjective evidence to our knowledge of the wide variety of purposes for which IS is used, with one study finding the most commonly reported purpose of IS to be self-referential thought, i.e., thinking silently about any aspect of the self (Morin et al., 2011).

In studies taking a more objective approach, the properties of IS are investigated through performance on measurable tests. In studies of healthy language users as well as a variety of disordered populations, rhyme and homophone judgments have been frequently utilized as a proxy measure for IS (Geva & Warburton, 2018; Gustafson, Bess, & Lancaster, 2017; Perrone-Bertolotti et al., 2014). Another objective approach to the study of IS is to examine errors in silent recitation of tongue-twisters. Early comparisons of errors in inner and overt speech illustrated that errors in IS often parallel overt slips of the tongue (Dell & Repka, 1992; Postma & Noordanus, 1996). More recently, Oppenheim and Dell (2008) investigated two specific effects seen in overt speech errors: lexical bias, where the error is likely to be a real word, and phonemic similarity effect, where closely related phonemes (e.g., /d/ and /t/) are more likely to
be substituted for one another than phonemically distant sounds (e.g., /d/ and /m/) (Oppenheim & Dell, 2008). Healthy young adults performed silent repetition of four-word tongue twister phrases and provided self-report of imagined errors. Results showed that IS exhibits lexical bias but not a phonemic similarity effect, indicating that IS is not fully specified (Oppenheim & Dell, 2008). Such a finding is at odds with early models claiming fully specified articulatory plans within IS (Levelt, 1983; Postma & Noordanus, 1996) and instead is consistent with newer models suggesting that lexical phonology is impoverished relative to post-lexical phonology (Goldrick & Rapp, 2007; Indefrey & Levelt, 2004). Later, the same authors showed that IS can evoke articulatory processing under certain circumstances, asserting that the level of specificity within IS is flexible (Oppenheim & Dell, 2010).

1.7.1.3 Neural correlates of inner speech

There is some evidence regarding the brain structures and networks that underlie IS in healthy individuals. An early PET study (N=6) showed activation in left inferior frontal gyrus (IFG) when healthy participants were asked to recite sentences silently in their minds, “without speaking or making subvocal articulatory movements” (McGuire et al., 1996). Other neuroimaging findings in healthy speakers use tasks like covert rhyme judgments as a proxy for IS, in order to focus on the more objective definition of IS as the ability to perform a computation on internal language. Such studies use functional MRI show activation of language regions including supramarginal gyrus in addition to left IFG during silent rhyme judgment (Hoeft et al., 2007; Lurito, Kareken, Lowe, Chen, & Mathews, 2000). These suggest that IS utilizes neural substrates that overlap with those underlying overt speech. In studies that directly compare covert to overt speech to test this question, however, results typically indicate a greater
response in sensorimotor and premotor regions for overt speech (Owen, Borowsky, & Sarty, 2004; Shuster & Lemieux, 2005).

1.7.2 Inner speech in individuals with aphasia

1.7.2.1 Objective evidence for the preservation of inner speech in aphasia

There is previous literature suggesting that IS can be preserved in individuals with acquired language impairments. An early study examined IS specifically in the context of conduction aphasia, to test a theory that conduction aphasia represents a disturbance of IS (Feinberg, Rothi, & Heilman, 1986). They used picture-based rhyme and homophone judgments as objective measures of IS ability and results indicated that four of five individuals with conduction aphasia performed well on these tasks despite displaying a general impairment of spoken output (Feinberg et al., 1986). Another study focused on a different aphasia subtype, Broca’s aphasia, to examine the use of pre-articulatory monitoring for error detection and correction (Oomen, Postma, & Kolk, 2001). While this study was not framed as an investigation of IS, findings indicated that individuals with Broca’s aphasia are able to monitor speech errors successfully using only pre-articulatory information (due to masking of spoken output), suggesting some access to internal phonology in this population (Oomen et al., 2001).

More recently, a few studies have used behavioral proxies for IS (e.g., silent homophone judgments based on written words) to show that individuals with conduction aphasia and individuals with motor planning impairments have preserved IS relative to spoken naming or oral reading (Geva, Bennett, Warburton, & Patterson, 2011; Stark, Geva, & Warburton, 2017). Impairments in IS, defined here as “the ability to create an internal representation of the auditory word form, and to apply computations or manipulations to this representation,” have been localized via voxel-based lesion-symptom mapping (VLSM) to damage in left IFG (pars
opercularis) and the white matter adjacent to the supramarginal gyrus (Geva, Jones, et al., 2011). These VLSM findings are consistent with the functional neuroimaging studies in healthy speakers (section 1.7.1.3), which localize activation primarily to the left IFG during tasks relying on IS.

1.7.2.2 The subjective experience of inner speech in individuals with aphasia

1.7.2.2.1 Inner speech in the context of anomia

Our work focuses on individuals with aphasia, so our examination of IS is specific to a population in which a failure of overt naming, anomia, is common. For the most part, we restrict our study of IS to the context of confrontation naming, in which a person sees a picture and is asked to name the picture out loud. Given the multi-stage nature of naming (see section 1.3 and Figure 1.2), a person may experience some intact lexical knowledge during a naming task, despite being subsequently unable to say that word out loud. Accordingly, many individuals with aphasia do make comments suggesting that their internal knowledge exceeds their spoken output, such as, “I know it but I can’t say it” (Blanken et al., 1987; Martin & Dell, 2007). A comment like this likely reflects a general experience of word retrieval failure, in which a person recognizes a picture and accesses the relevant semantics/concept, but cannot recall its name. This contrasts with a more specific experience in which a person may report “I can say it in my head” or “I can hear it in my head,” two comments that seem to reflect intact knowledge of the sounds of the word, which can be described as successful IS. In the study that will be presented in Chapter 2, we label these experiences as “idea without word” (IwW) and “successful inner speech” (sIS), respectively. Throughout this dissertation, I will use the term successful IS, or sIS, to refer to a subjectively reported experience of being able to hear or say a word correctly in one’s head; these terms will not necessarily reflect objective accuracy of IS. In our model of
naming, we hypothesize that sIS in the context of anomia represents a failure at the level of post-lexical output processing, after successful lexical retrieval (through the level of the phonological representation). By contrast, we propose that IwW represents a failure at an earlier stage of word-finding, due either to degradation of the semantic representation, the abstract lemma, or the phonological representation, or to impaired access to any of these forms (see Figure 1.2).

These two experiences (sIS and IwW) do not encompass all possible experiences of anomia. Another subjective comment that is heard in clinical settings is, “I can see it in my head,” which in the context of a naming task suggests that the individual believes that he or she can envision some or all of the letters comprising the written target word. Additionally, a person with aphasia may also experience two common word retrieval failures that can be experienced by healthy language users as well: tip-of-the-tongue and feeling-of-knowing. Tip-of-the-tongue (ToT) was defined in one early experiment as “a failure to recall a word of which one has knowledge” and is typically associated with a feeling of closeness to accessing the target word (R. Brown & McNeill, 1966). In terms of processing models of naming, we support a view of ToT in which there is a failure to access the full phonological form of the word, despite successful access to the semantic representation and lemma, or abstract word-form (Burke, MacKay, Worthley, & Wade, 1991; Dell et al., 1997; Levelt et al., 1999). A second metacognitive state, feeling-of-knowing (FoK), has been characterized not purely as a language phenomenon but as a failure of memory, in which a person believes that a target word or answer is stored in their memory, despite being unable to access it at the present moment (Hart, 1965). The experience of FoK is often considered to be less intense than ToT and is not accompanied by the strong feeling of closeness that is typically associated with a state of ToT (Hanley, 2014; Hart, 1965). In Chapter 2, I will describe in detail the relationship between these two common
experiences associated with word-finding failures and the experiences that are specific to the context of anomia, including sIS and IwW.

1.7.2.2 Prior studies on the subjective experience of inner speech in aphasia

As described above in section 1.7.1, there are two primary ways to define the phenomenon of IS (Levine et al., 1982), which can be generally characterized as subjective vs. objective viewpoints. While the majority of work on IS in aphasia has relied on objective measures, the subjective approach to studying IS may also contribute to understanding the phenomenon of IS in the context of aphasia. Goodglass and colleagues examined the more subjective aspects of IS by studying ToT, a closely related phenomenon. During a confrontation naming task, they asked patients with a variety of subtypes whether they had an “idea of the word” for items that they were unable to name aloud and then compared these self-reports to general performance on measures of lexical knowledge (first letter identification, syllable counting, and multiple-choice selection of the target word) (Goodglass, Kaplan, Weintraub, & Ackerman, 1976). Results showed that all participants reported ToT experiences at a frequency greater than their average performance on the lexical knowledge tasks, but individuals with Broca’s and conduction aphasia showed the lowest discrepancy between these two measures, suggesting that their reports of ToT were meaningful (Goodglass et al., 1976). Individuals with anomic aphasia reported relatively low levels of ToT and exhibited poor performance on the objective tasks, both consistent with the relatively poor lexical access that is characteristic of this aphasia subtype. In Wernicke’s patients, there was a mismatch: they exhibited poor performance on lexical knowledge tasks despite frequent reports of ToT, which was ascribed to poor self-monitoring (Goodglass et al., 1976).
While the Goodglass study was an examination of ToT rather than IS specifically, it is an important example of how the subjective experience of word retrieval can be meaningfully studied in individuals with aphasia, providing preliminary evidence for an important connection between self-reports and objective measures (Goodglass et al., 1976). In our own work, we similarly aim to tie subjective reports to objective evidence, thus merging these two distinct ways in which IS is typically characterized. Importantly, our work builds on the prior study by Goodglass et al. by comparing subjective reports of IS to objective task performance at the level of specific, individual items, in addition to comparing overall proportions of task-level accuracy. Previous work in our lab has examined self-reported IS by asking individuals to report the success of IS during a silent picture-naming test, with IS defined as being able to say a word in one’s head, with all the right sounds in the right order. In two individuals with aphasia, item-level IS reports predicted the subsequent success of spoken naming and the likelihood that, if naming failed, the error would be phonologically related to the target (Hayward et al., 2016). In a larger group of six participants, findings showed that self-reported successful IS was related to evidence for (at least) partial access to phonology and suggested that it was not related to articulation (Hayward, 2016).

1.8 Feasibility of using self-report in aphasia

As described in the previous section, IS can be studied via objective measures, but it (along with ToT and FoK) is primarily a subjective experience. The studies presented in this dissertation are focused on validating and understanding the experience of IS, which necessitates that some of our data be based on self-report by the individuals participating in our studies. Although there is a precedent for asking healthy language users about their experience of inner
speech (Hurlburt et al., 2016; Morin et al., 2011), our participants are individuals with acquired aphasia, who by definition have some difficulty with language use, either in comprehension, expression, or both. Before presenting results from this subjective approach in the chapters that follow, I will first give consideration to the feasibility of relying on self-report by individuals with aphasia.

1.8.1 General comprehension issues in aphasia

Although communicating with individuals with aphasia can be challenging, there is a considerable precedent in the aphasia literature for using self-report and other subjective measures in this population, particularly for exploring issues related to social participation and quality of life, using both qualitative and quantitative approaches (K. Brown, Worrall, Davidson, & Howe, 2012; Cocchini, Gregg, Beschin, Dean, & Della Sala, 2010; Hilari, Byng, Lamping, & Smith, 2003; Howe, Worrall, & Hickson, 2008; Lomas et al., 1989; Worrall et al., 2011). In our own approach, we followed general recommendations for supported conversation as well as specific guidelines for successfully interviewing people with aphasia (Kagan, 1998; Luck & Rose, 2007). In the first study (chapter 2), we obtained self-reports from individuals with aphasia in-person in an interview style, using communication strategies to maximize comprehension and verbal output. These techniques included: repeating questions, using written key words and pictures, probing with yes/no questions in addition to open-ended questions, and allowing extended time for responses. In the second and third studies, we added a self-report measure of IS to be provided on an item-by-item basis during a silent naming task; for this, we gave clear instructions and offered examples and practice items prior to the task. In addition to these measures, we also required a predetermined, minimum degree of sentence-level comprehension ability for inclusion in the final analyses of each study.
1.8.2 Issues of self-monitoring related to self-reported IS

Beyond task comprehension, one might raise concerns about whether individuals with aphasia have adequate self-monitoring in order to describe the success of their internal word retrieval. Most models of self-monitoring (see section 1.4) do include a mechanism by which word retrieval can be monitored prior to spoken output; however, if self-monitoring were impaired, participants might inaccurately report their experience of successful IS. In response to this concern, we present two main counterpoints for consideration.

First, there is an extensive prior literature demonstrating that many individuals with aphasia can detect errors in their spoken output (R. C. Marshall, Neuburger, & Phillips, 1994; Nickels & Howard, 1995b; Oomen et al., 2001; M. F. Schwartz, Middleton, Brecher, Gagliardi, & Garvey, 2016). Furthermore, individuals with Broca’s aphasia have been shown to rely even more heavily on internal, pre-articulatory monitoring as opposed to post-articulatory monitoring (Oomen et al., 2001). This is particularly important to our work, as it provides evidence that at least some individuals with aphasia should be able to monitor their IS. Generally, our model of IS in the context of naming is compatible with the most common models of self-monitoring. In the first study (Chapter 2), we operated under the assumption of a comprehension-based monitor, the most widely accepted theory over the past several decades, in which a single mechanism performs both auditory comprehension and self-monitoring (Levelt, 1983, 1989; Levelt et al., 1999). We excluded participants with poor comprehension at the level of single words, in case their comprehension impairments may have led to unreliable self-monitoring at that level. For our second and third studies, we did not commit to a single model of self-monitoring, instead assuming only that monitoring can occur at a level prior to articulation, which is consistent with
both comprehension and production-based monitors (Levelt et al., 1999; Nickels & Howard, 1995b; Nozari et al., 2011).

Secondly, and perhaps most importantly, the crux of this dissertation work involves testing the hypothesis that subjective reports of IS are meaningful and reliable, at least to some degree. In each successive experiment, we compare self-reported IS to objective measures, including behavioral language tasks, structural imaging, and psycholinguistic features of word stimuli. These analyses will serve the purpose of answering this concern, i.e., are individuals with aphasia reliable reporters when it comes to the experience of IS? If the connections between our subjective and objective measures are significant in predictable and interpretable ways, we can conclude that there is some preservation of self-monitoring for IS in our participants with aphasia. There are, of course, limitations and potential biases inherent to subjective self-reports and I will discuss these in detail in the context of study findings throughout the dissertation and in Chapter 5, the overall discussion.

1.9 Overview of the dissertation

1.9.1 Research questions

The background material presented thus far leads to an interesting set of open questions in the field of aphasiology. First and foremost: is the subjective experience of IS meaningful in individuals with aphasia? What is the nature of individual variability in experiencing and accurately reporting IS? Finally, what can we learn about the mental and neural mechanisms underlying the experience of IS in individuals with aphasia? In this dissertation, I intend to present evidence that helps to fill these gaps in knowledge regarding the experience of IS in individuals with aphasia.
1.9.2 Aims and hypotheses

In the studies that follow, I aim to test two main hypotheses: (1) that the subjective experience of IS in aphasia reflects the status of lexical (phonological) retrieval and (2) that IS itself does not rely on post-lexical stages of naming, but a failure of naming following successful IS results from deficits at the level of post-lexical output processing.

1.9.3 Overview of the three studies

The first study (Chapter 2) is a pilot investigation into the subjective experience of IS in people with aphasia, specifically focused on IS in the context of anomia. I developed a questionnaire in order to provide a structured opportunity for the participants to describe their experience with successful IS during instances of anomia, i.e., the feeling of retrieving the word but not being able to say it aloud. Successful IS in the context of anomia was explained in contrast to other experiences of word-finding difficulty, i.e., where the individual does not perceive successful retrieval of the word form. Questionnaire responses were compared to objective language testing in a group of 37 adults with chronic aphasia to assess relationships with various language processes. Then, a multivariate lesion-symptom mapping approach was utilized to examine relationships between lesion location and the different experiences of anomia.

In the second study (Chapter 3), the subjective experience of IS was examined on an item-level basis, rather than in the more general context of daily word-finding difficulties, with the primary aim of comparing the subjective reports with objective task performance. Specifically, participants reported the presence or absence of IS on a set of 120 items during a silent naming test. The same stimuli were used in a set of objective language tasks in two categories: (1) picture-based tasks requiring lexical retrieval and (2) matched auditory tasks.
where the spoken word was presented to the participant. We used linear mixed effects models across 27 participants’ data to assess the extent to which IS report related to performance on these other tasks, all at the individual item level. We also examined the relationship of relevant word features (e.g., frequency, length, age of acquisition, and articulatory complexity) to IS report and to spoken naming.

In the final study (Chapter 4), we investigated individual differences in IS by examining relationships between IS report and performance on an extensive battery of language assessments. In a large group of participants (N=53), we used a factor analysis to identify the main components of the testing battery. Then, we correlated the resulting factor scores with spoken naming ability and with IS report (same silent naming test from the previous study). In this analysis, we also examined correlations between error patterns on the naming task and IS report as well as correlations between error patterns and overall spoken naming accuracy. This task-level approach allowed us to ask two main questions related to our hypotheses. First, are people who report more successful IS report better at lexical retrieval tasks in general? Second, is speech output processing important for IS or only for spoken naming?

In the Discussion chapter (Chapter 5), I will discuss the implications of the results from each of these three studies and will also describe a few other approaches that we attempted in our study of IS, each of which resulted in null findings. I will conclude with open questions and future directions for this line of work.
Figure 1.1. Aphasia subtypes.
Figure 1.2. A general model of naming.
CHAPTER 2

Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates


2.1 Introduction

People with aphasia universally struggle with anomia, an acquired deficit of naming and word finding. These individuals often report that their internal knowledge of words exceeds what they demonstrate through aloud speech, saying, for example, “I know it but I can’t say it.” At times, these reports include the specific feeling that one can hear or say the correct word in one’s head, an experience that we label here as “successful inner speech,” or sIS. No prior studies have examined whether these subjective feelings of sIS provide useful information about the cognitive processes underlying anomia in a large group of individuals with aphasia. In this exploratory study, we gathered information from individuals with aphasia about sIS and related experiences, to test how these experiences map onto specific language deficits, preserved language abilities, and lesion locations.

2.1.1 Anomia and our model of naming

One-third of all stroke survivors are diagnosed with aphasia, a language disorder that often results in chronic communication deficits (Berthier, 2005; Engelter et al., 2006). The specific language impairments associated with aphasia can vary widely from person to person, but a hallmark symptom of aphasia is anomia (Goodglass & Wingfield, 1997; Laine & Martin, 2006; Maher & Raymer, 2004). Anomia is easily observable: a person with aphasia is sometimes unable to produce certain words, either during spontaneous speech or during an attempt to name an object or picture (both labeled “overt anomia” here). The overt deficit is conspicuous, but the
cognitive mechanisms underlying anomia are best understood in the context of a theoretical model showing the stages of successful naming. **Figure 2.1** presents a simplified model of the processing stages that have been suggested by existing naming models: access to a word’s semantic representation (encompassing both semantic knowledge for the concept and the corresponding abstract word-form), access to the phonological representation, and the post-lexica output processes that are necessary to turn that phonological form into a spoken word (Dell & O’Séaghdha, 1992; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Goldrick & Rapp, 2007; Levelt, 2001; Levelt, Roelofs, & Meyer, 1999; Walker & Hickok, 2015).

Correct naming requires success at each step of the naming process; consequently, overt anomia may result from a breakdown at any point, either within a stage or in the processing phase between stages (Dell et al., 1997; Laine & Martin, 2006). The locus of word-finding impairment within the *access* stages may be at the level of the semantic representation, at the level of the phonological representation, or in the mapping stage between the two. The process of word retrieval is typically understood to be complete once access to the phonological representation has been achieved (i.e., the end of the white *access* stages, **Fig. 2.1**). Next, there are additional *output* processes that are required to achieve successful spoken output (shown in gray, **Fig. 2.1**), i.e., the “word-production” components of naming (Kohn & Goodglass, 1985). A deficit at the level of the sensorimotor interface or motor programs, which may include phonological, phonetic and articulatory processes, can also result in spoken errors that can be difficult to distinguish from a word retrieval deficit (Feinberg et al., 1986; Geva, Bennett, et al., 2011; Miceli, Amitrano, Capasso, & Caramazza, 1996).
2.1.2 The subjective experience of anomia

Anecdotally, many persons with aphasia endorse the idea that their out loud naming abilities do not match their inner speech (IS), stating, “I know it but I can’t say it” (Blanken et al., 1987; Martin & Dell, 2007) or more specifically, “I can say it in my head” or “I can hear it in my head.” These statements are often accompanied by a sense of frustration, but their exact meaning is unclear and, to date, they have not been systematically explored. We hypothesize that many individuals with anomia are aware of the level at which their inability to find a word arises, and that these statements relate to underlying word retrieval and production processes.

More precisely, we hypothesize that there are at least two discrete internal experiences of anomia, both of which result in a person being unable to produce a word correctly out loud:

A. Wanting to communicate a concept or idea but failing to find the right word in one’s head (idea without word, IwW)

B. Finding the right word in one’s head but failing to turn that lexical form into a spoken word (sIS)

Note that throughout this manuscript we use the terms IwW and sIS to refer to the subjective experience of anomia. Therefore, the term sIS does not imply that the IS is necessarily correct, only that the individual reporting the experience of sIS feels that it is; the accuracy of these reports may vary across individuals.

The difference between these two subjective experiences can be understood in terms of the possible loci of impairment in the naming model above (section 2.1.1): we suggest that IwW relates to a deficit at an early stage of processing involving access to either the semantic or phonological representation, whereas sIS reflects successful access to the word but a deficit in the post-lexical output processes (Figure 2.1). We hypothesize that the sense of successful sIS
arises after successful retrieval of both the semantic representation and the phonological form. In the model, self monitoring occurs via an inner loop that utilizes speech perception areas in the superior temporal gyrus (Indefrey, 2011), but the precise mechanism of monitoring is not critical for the current study (for further discussion of self-monitoring in IS, see Discussion section 2.4.4). IwW as described above is heterogeneous: an individual may feel that he/she cannot retrieve any word at all, or has retrieved a related word, or has retrieved a word that is close but not exactly right. Stated plainly, IwW encompasses all experiences of anomia that do not meet the criteria for sIS. In contrast, sIS is discrete: an individual reporting he/she feels able to say or hear the right word internally, despite being unable to say it out loud. We can make targeted predictions about who should experience sIS, both with respect to language processing abilities and to lesion location (see section 2.1.5).

2.1.3 Relationship between sIS, IwW and other failures of word retrieval

To further clarify our operational definitions of sIS and IwW, it is useful to distinguish them from two related concepts, the tip-of-the-tongue (ToT) phenomenon and “feeling-of-knowing” (FoK). The ToT experience is well known to all language users and has been well characterized in the psychology literature since the early work of Brown & McNeill (1966). The ToT state is “a failure to recall a word of which one has knowledge” (Brown & McNeill, 1966), where an individual is unable to access a word, but has a feeling of being very close to recalling it. Here, we endorse the view that ToT arises from successful lexical-semantic access and incomplete (but partial) phonological access (Burke, MacKay, Worthley, & Wade, 1991; Dell et al., 1997; Harley & MacAndrew, 2014; James & Burke, 2000; Levelt et al., 1999; Meyer & Bock, 1992).
By definition, individuals experiencing ToT do not have full access to the phonological form of the target word at the time of attempted production, so ToT is distinct from our definition of sIS. Furthermore, most language users have experienced the ToT sensation, but individuals without neurological speech impairments should rarely experience a feeling of sIS without successful spoken output. ToT is also distinct from IwW, although healthy language users may experience both types of word-finding failure. It is possible for IwW to reflect a feeling of partial phonological access, in which case it would be very closely related to ToT; however, as described in section 2.1.2, IwW also encompasses retrieval of a related word or an inability to retrieve any word at all. These word-finding failures do not necessarily include the ToT state’s sense of closeness to retrieving the target word, which likely results from the achievement of partial phonological access (A. S. Brown, 1991; Jersakova, Souchay, & Allen, 2015). Thus, we hypothesize that the ToT experience does not precisely map onto the anomic experiences we describe here, but lies in between IwW and sIS at the level of partial phonological access.

Similarly to ToT, a FoK experience is a metacognitive state in which a person can identify that a word is stored in memory, believes that he/she may be able to recall it at a later time, and would be able to recognize the target word when it is presented to him/her (Hart, 1965). Importantly, FoK does not necessarily include the sense of closeness that accompanies a ToT state (Hanley, 2014). FoK is easily distinguishable from our definition of sIS, as it does not include access to the phonological form. In contrast, FoK and IwW share many characteristics and may be indistinguishable in some cases. Specifically, in the context of naming, both FoK and IwW involve a sense of recognition of a certain object or person presented, without the ability to retrieve the name. FoK likely surpasses the level of knowledge included in IwW, however, as a
person reporting a FoK state is typically able to recognize the name when they subsequently see or hear it (Hart, 1965; Maril, Simons, Weaver, & Schacter, 2005). IwW, as we have defined it here, does not require an individual with anomia to demonstrate any recognition ability of the target concept or word. We do not include FoK in our interviews with individuals with aphasia, due to potential for confusion with IwW.

2.1.4 Inner speech in aphasia

There is some objective evidence, albeit in limited previous literature, that IS abilities can exceed overt speech output in some persons with aphasia (Feinberg et al., 1986; Geva, Jones, et al., 2011; Geva, Bennett, et al., 2011; Goodglass et al., 1976). Very few studies have systematically examined self-reports of this experience: first, Goodglass et al. found that individuals with conduction aphasia and Broca’s aphasia often reported having “the idea of a word,” labeled tip-of-the-tongue (ToT) in their study, and correspondingly demonstrated relatively intact phonological awareness of target words during anomic events (Goodglass et al., 1976). They used objective measures of lexical access (e.g. rhyme judgments, syllable counting) as a proxy for IS, finding that individuals with Broca’s and conduction aphasia performed better on these tasks than their aloud naming abilities would suggest. This study showed that some individuals with aphasia demonstrate some lexical knowledge of words that they cannot say out loud and, importantly, that certain aphasic subgroups show intact self-awareness with respect to this ability. Their approach, however, focused on ToT and did not require participants to endorse the specific experience of sIS as we define it. A recent study used silent homophone judgments as a proxy for IS and showed that individuals with either conduction aphasia or motor planning impairments performed better on these tests than their out loud naming would predict (Geva, Bennett, et al., 2011). There is only one study since Goodglass et al. to directly explore the
subjective experience of anomia. Hayward et al. (2016) asked two people with aphasia to report the accuracy of their IS for individual items on a covert naming task. Reports of sIS were associated with lexical frequency, accuracy of naming aloud on other days, and the likelihood of phonological errors when naming was incorrect; report of sIS also related to the rate of relearning of individual naming items during subsequent anomia therapy (Hayward et al., 2016). This recent study provides strong evidence that reports of sIS by people with aphasia can provide useful information regarding the mental processes of word-finding and production. Specifically, the results suggest that reports of sIS on individual naming attempts relate to lexical phonological access, and that subsequent errors relate to post-lexical output processes. Here, we assessed more general relationships between how often people with aphasia report experiencing sIS in daily life and their stroke locations and language deficits.

2.1.5 Expected relationships between self-reported sIS, language functions, and lesion locations

In the current study we aim to validate individuals’ subjective experiences of sIS by comparing their self-report to lesion location and performance on behavioral testing, to determine whether there are meaningful relationships. First, there should not necessarily be a relationship between reports of sIS and overall severity of aphasia or anomia. The integrity of sIS should depend on which specific language processes are affected as a result of the brain injury, not on the overall severity of impairment.

What specific language processes must be intact in order for an individual to achieve sIS? What language processes must be impaired for someone to experience overt anomia, despite sIS? And finally, what brain regions underlie those abilities? Returning to our simplified model of naming (Fig. 2.1), we suggested that sIS requires successful lexical access. As such, individuals who report frequent sIS should perform relatively well on tests of semantic processing as well as
input-level phonological processing. We suggest further that individuals who experience overt anomia despite frequently achieving sIS have deficits in post-lexical output processes. The post-lexical phase of word production encompasses several proposed processes and a disruption of any of them could lead to overt anomia after sIS, e.g., loss of the phonological trace or a failure of sensorimotor mapping, phonological assembly, phonetic encoding, or articulatory planning (Goldrick & Rapp, 2007; Indefrey, 2011; Indefrey & Levelt, 2004; Walker & Hickok, 2015). The last of these, a failure of articulatory motor planning, often falls under a diagnosis of verbal apraxia or acquired apraxia of speech (Duffy et al., 2015; Strand et al., 2014; Whiteside, Dyson, Cowell, & Varley, 2015). Individuals with anomia who report frequent sIS should perform poorly on tasks that rely heavily on these output processing skills, such as repetition or oral reading. Furthermore, there may be a *lexicality* effect, defined by poorer performance on production tasks using pseudowords (pronounceable non-words) as opposed to real words. Real words benefit from stable patterns of activation and a boost from semantic content, so can be more resilient to phonological processing impairments than pseudowords, which do not share these benefits (Coltheart, 1996; Crisp & Lambon Ralph, 2006; Patterson, Suzuki, & Wydell, 1996). The lexicality effect thus serves as a specific measure of post-lexical phonological output processing impairments, which we predict relate to reports of sIS.

Parallel predictions can be made regarding the integrity of relevant brain regions. Individuals who report frequent sIS are predicted to have relatively intact lexical semantic and phonological access, so are unlikely to have lesions in areas that underlie these processes, such as the middle temporal gyrus and the anterior/middle regions of superior temporal gyrus (STG) (Hickok, 2009; Indefrey & Levelt, 2004; Price, 2012). These individuals, who report frequent sIS but still experience anomia, are likely to have lesions in areas that affect post-lexical output
processes (Feinberg et al., 1986; Geva, Bennett, et al., 2011). These dorsal stream areas may include the posterior STG and supramarginal gyrus (sensorimotor interface for phonological assembly/encoding) and the ventral motor and premotor cortices (for articulatory coding and sequencing) (Hickok, 2009; Indefrey & Levelt, 2004; Price, 2012; Walker & Hickok, 2015).

2.1.6 Overview of study aims

Studies on IS in aphasia have been limited and only two prior studies have focused on the individual’s own subjective experience of anomia. In this study, we aimed to fill this gap by providing individuals with aphasia a structured opportunity to describe and report their experience of IS. Specifically, we aimed to establish preliminary evidence that (1) individuals with aphasia themselves can inform us about their IS abilities and (2) their self-reports have meaningful relationships with overt language abilities and with lesion locations. Overall, these findings may validate the common reports made by individuals with aphasia regarding their IS. Additionally, the study of sIS may contribute to existing models and, in turn, inform future approaches to anomia treatment.

2.2 Materials and Methods

2.2.1 Participants

Participants for this study were thirty-seven individuals drawn from the participant group of a larger project investigating the brain mechanisms of aphasia recovery. All participants gave informed consent under a protocol approved by the Georgetown University Institutional Review Board. For inclusion in this study, participants were required to be a native English speaker (spoken since age five or younger), to have adequate hearing and vision (with correction, if appropriate) and to have a diagnosis of aphasia from a left hemisphere ischemic or hemorrhagic
stroke that occurred at least six months prior to enrollment, without history of other significant neurological or psychiatric illness.

Participants included 26 men and 11 women, with an average age of 60.5 years old (SD = 10.1, range = 39-83 years) and an average of 16.8 years of formal education (SD = 2.8, range = 12-24 years). Using the Edinburgh Handedness inventory, 30 participants were identified as right-handed, 5 as left-handed, and 2 as ambidextrous (R. Oldfield, 1971). Post-stroke chronicity ranged from 11 months to 21 years (median = 50.6 months, mean = 64.4, standard deviation = 52.2). Five participants had aphasia resulting from a hemorrhagic stroke and thirty-two participants had aphasia resulting from an ischemic stroke, including one individual whose aphasia developed from a series of strokes that followed left hemisphere brain tumor resection.

The group presented with a wide range of aphasia severity and subtype, as established by the Western Aphasia Battery – Revised (WAB-R) (Kertesz, 2006). WAB-R Aphasia Quotient (AQ) scores ranged from 20 to 96.2, with an average of 68.6 (SD = 23.9). Three of the included participants scored above a 93.8 on the WAB-R AQ, so are technically considered to be within the normal range according to the assessment scoring system. They each reported ongoing communication difficulties and exhibited clinical evidence of mild anomia, so were included in this study. Participants were diagnosed using the WAB-R aphasia subtype classification criteria: 19 anomic (including the three “recovered” patients), 13 Broca’s, two Wernicke’s, two conduction, and one transcortical sensory. We excluded participants with significant impairments in sentence- and/or word-level comprehension. Cut-off scores were a minimum of 4/10 on the WAB-R Auditory Verbal Comprehension composite and a minimum score of 24/48 on an in-house auditory word-picture matching task (in a field of 6, 50% performance on this task is better than chance at P<1x10^{-7}). For diagnosis of verbal apraxia, we utilized the Apraxia of
Speech Rating Scale (ASRS) 2.0, a rating-based checklist of distinguishing features of apraxia of speech (Strand et al., 2014 and Duffy, 2016). Authors S.F.S. and M.E.F. individually observed videos of verbal subtests from the WAB-R, rated each participant, and then discussed discrepancies as needed to arrive at a consensus score for each participant.

2.2.2 Methods for probing the subjective experience of anomia

As stated in the Introduction, many individuals with aphasia make vague statements such as, “I know it but I can’t say it” – our goal was to generate more specific descriptions of the experience of anomia that correspond to our hypothesized experiences (described in section 4.1.2). We generated easily understandable phrasing to describe sIS vs. IwW by combining anecdotal experiences with individuals with aphasia with knowledge of theoretical models of anomia. To obtain self-report data from our participants, we used open-ended, yes/no, and rating scale-based questions, beginning with:

1. Do you ever know what you want to say but you can’t say it out loud?

2. How would you describe that feeling?

After these initial questions, we presented a set of written phrases and pictures to help explain the various concepts that follow (Figure 2.2). Note that the pictorial support accompanying IwW clarifies that the idea or concept is available, but the word is not. With the pictures still in view, participants were asked to reply yes/no to the following questions about each specific experience:

Do you ever feel...

- “I can’t say it out loud, but I can say the right word in my head and it sounds right” (sIS)
- “I can’t say it out loud, but I know what it is in my head” (IwW)
- “I can’t say it out loud, but it’s on the tip of my tongue” (ToT)
Participants were then asked to identify how frequently they experience each of the above scenarios, on the following scale: never, rarely, sometimes, often, and almost always. [For subjects who answered “no” on the prior yes/no question, the expected response for the rating scale was “never”]. All possible ratings were available in written form for participants to choose from and were presented in vertical orientation, to avoid possible negative impact of hemispatial neglect (Kleinman et al., 2007). When giving answers, participants could either point to the written word or provide a verbal response. In the Results and Discussion, participants’ subjective responses regarding the frequency of each experience will be referred to as “anomia ratings,” with specific reports of the frequency of sIS referred to as “sIS ratings.”

Interview sessions for obtaining anomia ratings took place at either Georgetown University Medical Center or MedStar National Rehabilitation Hospital, often in conjunction with sessions involving other language testing (described below in section 2.2.3) or structural imaging (2.2.4). All sessions took place in a quiet, private room and were videotaped. As described above, the spoken questions were accompanied by written/picture support in order to help participants understand the questions as well as remember and utilize the possible answers; some participants required additional time and explanation with these materials, so this portion of the session lasted from 10-30 minutes, depending on the participant. The first author of this paper (M.E.F.) obtained anomia ratings from 33 of the participants, with two other trained lab members completing the sessions with the remaining four participants. All transcribed live responses were confirmed for accuracy via later video review. Additionally, M.E.F. reviewed the videos for the four subjects who were seen by a different researcher, to confirm reliability of administration procedures. No amendments to the recorded responses were made during this process.
2.2.3 Language testing

This study’s participants were drawn from a larger cohort of study participants in the lab and were each administered an extensive language/cognitive assessment battery as a part of that larger study. In the current study, we analyzed only those behavioral assessments that either (1) provided a measure of overall severity or (2) fit our hypotheses about the relationship between self-reported IS and language processing, namely that frequency of self-reported sIS in individuals with anomia should correlate with poor performance on tasks of output processing and relatively intact performance on tests of lexical/semantic access. The language measures are given below; all unpublished tasks were developed in-house and normed in a healthy cohort of older adults.

- Overall language ability: WAB-R AQ (Kertesz, 2006)
- Overall naming ability: Philadelphia Naming Test (PNT) – 60 item short-form version (Roach, Schwartz, Martin, Grewal, & Brecher, 1996)
- Motor speech: ASRS 2.0
- Lexical/semantic composite score (serving as a surrogate measure of lexical/semantic access). This score is an average of:
  - The Pyramids & Palm Trees Test (Howard & Patterson, 1992)
  - Auditory word-picture matching task (48 items, field of 6)
- Lexicality effect (serving as a measure of post-lexical phonological output processing). This score is the difference between word and pseudoword performance, averaged across repetition and reading tasks:
  - Repetition
    - Pseudoword repetition (30 single words ranging from 1 to 5 syllables)
- Word repetition (composite of 30 single words from the *Apraxia Battery for Adults-2* subtests 2A: Increasing Word Length and 5: Repeated Trials, selected for presentation format and to match average word length of the pseudoword repetition task)
  - Single word oral reading
    - Pseudowords (20 single syllable pseudowords)
    - Words matched to pseudowords (20 real single syllable words)

In distinguishing between tasks of *access* and *output*, we acknowledge that we do not have tasks that directly assess lexical access in production; however, a similar word-to-picture matching task and the *Pyramids & Palm Trees* test have both been shown to be good predictors of output abilities, as measured by error patterns on a naming task (Martin, Schwartz, & Kohen, 2006). All language assessments were given to the participants by a certified speech-language pathologist or a postdoctoral researcher, each of whom was thoroughly trained in administration and scoring procedures relevant to the current study. All sessions were videotaped.

2.2.4 Structural imaging: Acquisition, lesion segmentation, and warping

Structural magnetic resonance imaging (MRI) was performed at the Center for Functional and Molecular Imaging at Georgetown University, which uses a 3T Siemens Magnetom Trio scanner. Participants underwent a high-resolution T₁-weighted structural scan (MPRAGE) for lesion localization (TR of 1900 ms, TA of 2.56 ms, FOV 250 x 250, 9° flip angle and 160 contiguous sagittal slices for voxel size of 1 x 1 x 1 mm. Trained lab members (blinded to the behavioral data) manually traced the lesions onto these T₁-weighted structural images by identifying differences in signal between healthy and lesioned brain areas in MRicron (http://www.mccauslandcenter.sc.edu/micro/mricron). All lesion segmentations were verified
and finalized by a board certified neurologist (P.E.T.). Lesion masks were warped into the Montreal Neurological Institute (MNI) space using the VBM8 toolbox in SPM8 (http://www.fil.ion.ucl.ac.uk/spm) running under Matlab R2014a.

2.2.5 Statistical analysis

Statistical analyses were performed in SPSS 24. Ordered logistic regression was used to determine relationships between language scores and subjective anomia ratings. Prior to the regression, a series of bivariate correlations (Spearman and Pearson as appropriate) were performed to examine relationships between the subjective anomia ratings and to explore the collinearity among the language scores. These exploratory tests were not corrected for multiple comparisons.

Imaging analyses were performed in N=36, excluding one participant who had a right ACA stroke at the same time as his left hemisphere stroke. We used a multivariate lesion-symptom mapping technique in which patterns of lesion status across voxels in the brain are considered simultaneously to predict a single behavioral score (Zhang, Kimberg, Coslett, Schwartz, & Wang, 2014). This was performed using the support vector regression-based lesion-symptom mapping (SVR-LSM) toolbox in MATLAB (https://cfn.upenn.edu/~zewang/), which is more resistant than voxel-based lesion-symptom mapping to theoretical bias in localization due to lesion covariance, especially when behaviors rely on multiple distinct brain regions (Herbet, Lafargue, & Duffau, 2015; Mah, Husain, Rees, & Nachev, 2014). Here, we used SVR-LSM to identify lesion locations that predict the three anomic experiences: sIS, IwW, and ToT. We did so by identifying voxels where the presence of a lesion was predictive of a higher self-reported frequency for each experience (as opposed to most LSM analyses, which look for voxels that predict lower scores). We excluded voxels that were lesioned in fewer than seven participants.
(minimum 20%). Significance testing was performed using 10,000 permutations of anomia ratings, and the significance threshold was set at a voxelwise P<.01 with a cluster size threshold of 200 mm$^3$. Direct total lesion volume control was applied to control for relationships between lesion size, stroke distribution, and our measures of interest. Because SVR-LSM considers all voxels simultaneously in a single regression model, correction for multiple comparisons is not required (Zhang et al., 2014).

2.3 Results

2.3.1 Subjective anomia ratings

All participants endorsed the general experience of anomia by answering yes to the question: “Do you ever know what you want to say but you can’t say it out loud?” Of those, 29 participants (78.4%) endorsed the specific anomic experience of sIS - “I can’t say it out loud, but I can say it in my head and it sounds right.” Figure 3A shows participants’ reported frequencies of sIS during anomia, from never to almost always. In Figure 3B, this distribution is compared to participant-reported frequencies of “I know what it is in my head” (IwW) and “It’s on the tip of my tongue” (ToT).

2.3.2 Behavioral comparisons

We first examined relationships between the participant-reported frequency of the three experiences of anomia (sIS, IwW, and ToT; on a scale of 0/”never” to 4/”almost always”) using nonparametric correlations. These a priori analyses tested the hypothesis that sIS and IwW are discrete perceptions but that ToT, which represents partial phonological access, bears similarities to both. In line with our hypothesis, no relationship was found between frequencies of sIS and IwW (P=.73), whereas ToT correlated with both sIS (P=.006) and IwW (P=.03; Table 1).
In preparation for the main regression analysis examining relationships between the participant-reported frequency of the anomia experiences and language scores, we first performed exploratory bivariate correlations to examine relationships among the variables (Table 1). These correlations revealed that sIS correlated with each of our language measures (all $P<.05$ uncorrected), except the lexical/semantic composite score ($P=.13$ uncorrected). In contrast, no relationships with language measures were observed for ToT or IwW (all $P>.10$ uncorrected). The exploratory correlations also revealed that WAB-R AQ, PNT, ASRS, and the lexical/semantic composite were highly inter-correlated (all $P<.001$, except ASRS with lexical/semantic composite at $P=.007$). Thus, relationships between sIS and these individual scores may not be distinguishable from its relationship with overall aphasia severity. It is also notable in this context that although the lexical/semantic composite score related to overall aphasia severity, it was not correlated with sIS, even at an uncorrected threshold. The lexicality effect did not correlate with any of the other language scores (all $P>.10$).

Next, relationships between sIS and behavioral scores were formally examined using ordered logistic regression with reported frequency of sIS as the dependent measure. Because of the observed collinearities discussed above, WAB-R AQ, PNT, ASRS, and the lexical/semantic composite were scaled from 0-1 and averaged to create a measure of overall aphasia severity. This score and the lexicality effect score were used as the independent variables in the regression. The overall regression model fit was significant ($\chi^2(2)=12.19$, $P=.002$). Both aphasia severity and lexicality effect had independent relationships with sIS (aphasia severity Wald $z(1)=7.53$, $P=.006$; lexicality effect Wald $z(1)=6.23$, $P=.013$). The direction of effects show that more frequent sIS was associated with poorer reading and repetition of pseudowords compared
to words (lexicality effect), and more severe aphasia. Regressions on IwW and ToT using the same independent variables yielded no significant results (P=.421 and P=.324, respectively).

2.3.3 Multivariate lesion-symptom mapping

Figure 2.4A shows a lesion overlap map demonstrating our participant group provided coverage across the entire perisylvian language network, including frontal, parietal, and temporal lobe regions. In nonparametric correlations, none of the three experiences of anomia were correlated with overall lesion size (all P>.10). SVR-LSM analyses were conducted to identify specific lesion locations associated with higher reports of each experience of anomia. sIS was associated with lesions in the inferior frontal gyrus pars opercularis (IFG pOp) and the ventral pre- and postcentral gyri, whereas IwW was associated with lesions in the mid-posterior middle temporal gyrus and the angular gyrus (Table 2.2, Fig. 2.4). ToT was associated with lesions in the dorsal portion of IFG pars triangularis and pOp, near the inferior frontal junction. The main analysis showed little to no spatial overlap between the lesion sites associated with the three subjective experiences of anomia. To ensure this dissociation was not related to the statistical threshold used, we examined the lesion-symptom maps at a reduced threshold (P<.10). At this lower threshold, there was some overlap between sIS and ToT in the ventral premotor cortex and POp, but still no spatial overlap between sIS and IwW. In this map, as predicted, lesions in broad dorsal stream language areas (frontal/parietal lobes) involved in post-lexical output processes relate to sIS and ToT, whereas lesions in superior and middle temporal areas involved in lexical access relate to IwW.
2.4 Discussion

By probing subjective experiences, we found that people with aphasia differentiated between distinct experiences of anomia: the sense of being able to say a word correctly in one’s head (sIS) and the sense of wanting to communicate an idea without finding the right word in one’s head (IwW). Both of these experiences related to the ToT state in terms of reported frequency. We found that reports of sIS, but not IwW or ToT, related to behavioral measures of language impairment, including overall severity of anomia and post-lexical output processes. Further, multivariate lesion-symptom mapping results demonstrated that damage in discrete brain regions was associated with each of the three subjective experiences of anomia.

2.4.1 Successful inner speech is subjectively dissociable from a failure to retrieve the word (IwW)

During our testing sessions, all subjects endorsed the general experience of anomia, in which they want to communicate a word or idea but are unable to do so aloud, consistent with our knowledge of the universality of anomia in aphasia (Goodglass & Wingfield, 1997). As predicted, our precise probing enabled them to go beyond this general endorsement to provide a more detailed characterization of how frequently they experience sIS, IwW, and ToT. Although there was a bias towards a “sometimes” response, participants gave a wide range of responses, which verifies that the wording of our questions did not consistently lead respondents to a particular response pattern. We demonstrated a lack of within-subject correlation between reports of sIS and IwW, supporting the conclusion that participants conceived of these as two different experiences. Reports of both of these, however, were positively correlated with reported frequency of ToT. This is consistent with our hypothesis that ToT, as a state of partial phonological access, falls somewhere in between the two experiences of anomia that we have defined here. Like IwW, ToT is a failure of word retrieval, in which a person cannot access the
target word. Like sIS, ToT includes a strong sense of closeness to being able to produce the
target word, generated by some amount of access to the phonological representation of the word
(Burke et al., 1991; Dell et al., 1997; James & Burke, 2000). ToT was reported to be nearly
universal in our participants (see Fig. 2.3B), which is expected due to its prevalence in healthy
populations (Brown, 1991).

2.4.2 sIS is common in aphasia and relates to aphasia severity and phonological output processes

The widespread reports of sIS suggest that it is common in aphasia – nearly 80% of
participants endorsed having some experience of sIS during overt anomia, with varied frequency
ranging from rarely to almost always. We predicted that participants who reported frequent sIS
during anomia would demonstrate deficits primarily in phonological output processes, such as
sensorimotor mapping and articulatory motor programs (see Fig. 2.1 for the model) (Walker &
Hickok, 2015). We found support for this hypothesis in the strong positive correlation between
sIS and the lexicality effect. There is no consensus as to the exact nature of the phonological
impairment that gives rise to the lexicality effect, but Friedman (1995) gives several possibilities:
a deficit in generating the phonological code, in maintaining the phonological code, or at the
level of motor output (Friedman, 1995). The relationship observed between sIS and the lexicality
effect is thus consistent with a deficit in phonological output processing. However, rather than
reflecting a post-retrieval output failure, the relationship with the lexicality effect could also be
consistent with the alternate interpretation that sIS arises from activation of a nearly-correct
phonological form that is strong enough to result in an internal sense of retrieval, but not strong
enough to be accurately translated to a correct motor speech act. By this account, however, the
only distinction between sIS and ToT would be a subtle difference in the strength of the retrieved
phonological form. The distinct relationships with behavior and lesion location observed here for sIS compared to ToT suggests otherwise.

The positive relationship with severity of apraxia of speech indicates that an articulatory impairment may relate to the experience of sIS, but additional research using more precise behavioral tasks would be necessary to delineate the exact relationship between sIS and the phonological, phonetic, and articulatory subprocesses of the post-lexical processing stage. Furthermore, apraxia severity was strongly correlated with overall naming/language impairment, so we cannot disentangle any specific relationship between sIS and apraxia from a more general relationship with overall severity. Indeed, we found a relationship between sIS ratings and overall naming or language impairment that was not predicted *a priori*. It is possible that this relationship relates to a sampling bias. Our cohort was comprised mostly of individuals with Broca’s or anomic aphasia, in whom severity of apraxia and aphasia were closely correlated. The relationship between sIS and aphasia severity might not exist in a population in which these deficits were uncorrelated. Using our methods in a sample that includes more subjects with fluent aphasia and severe anomia related to large temporal lobe lesions would help to clarify the relationship between sIS ratings and overall naming/language ability. The lack of a significant negative correlation between sIS and our lexical/semantic composite score, even at an uncorrected significance threshold, supports our hypothesis that sIS reflects intact lexical/semantic access.

In considering the significant relationships that exist between sIS and behavior, it is equally important to note the lack of relationship between IwW and these same measures, in either a positive or negative direction. Given our focus on sIS, we aimed primarily to differentiate between the specific experience in which lexical retrieval feels successful (sIS) and
all other experiences of anomia (IwW). The incorporation of diverse anomic experiences – including partial retrieval, retrieval of a related word, or failure to retrieve any word at all – into a single category may have led to the lack of strong relationships between IwW and any individual behavioral measure.

2.4.3 Lesion locations conform to predicted anatomical patterns

The multivariate lesion-symptom mapping results provide a second line of evidence for the validity of the subjective anomia ratings. Taken as a whole, the SVR-LSM results illustrate a striking dissociation between the two main experiences of anomia, with minimal overlap in lesion localization between sIS and IwW even at a very low statistical threshold. Strong endorsement of sIS is associated with lesions in inferior frontal regions and sensorimotor cortices, which are involved in articulatory coding/sequencing and motor speech output (Indefrey & Levelt, 2004; Price, 2012; Schwartz, Faseyitan, Kim, & Coslett, 2012). The experience of IwW is associated with lesions in the middle temporal gyrus and angular gyrus, areas that underlie lexical semantic retrieval (Binder, Desai, Graves, & Conant, 2009; Mechelli, Josephs, Lambon Ralph, McClelland, & Price, 2007; Price, 2012; Troiani et al., 2008). The ToT experience is associated with lesions to the dorsal pars opercularis of the inferior frontal gyrus, near the inferior frontal junction, which contributes to phonological assembly prior to articulatory planning (Ghosh, Tourville, & Guenther, 2008; Indefrey, 2011). This discrete localization supports that a ToT experience is distinct from either anomic experience. Also, this ToT finding is consistent with prior neuroimaging research on the neural correlates of ToT in healthy adults, which shows that a loss of structural integrity in a nearby region, the left insula, relates to increased experience of ToT during normal aging (Shafto, Burke, Stamatakis, Tam,
Tyler, 2007) and that young adults show stronger insula activation than older adults during ToT states (Shafto, Stamatakis, Tam, & Tyler, 2009).

Overall, these lesion findings support our hypothesis: an early failure of word retrieval results in a feeling of IwW, whereas successful retrieval followed by a failure of post-lexical output, primarily related to damage in dorsal stream language areas, results in a feeling of sIS. A feeling of ToT arises in between, consistent with the view of ToT as a state in which the lexical item is retrieved but there is only partial access to the phonological representation (Burke et al., 1991; Dell et al., 1997; James & Burke, 2000; Levelt et al., 1999; Meyer & Bock, 1992).

2.4.4 The role of self-monitoring in sIS

Despite the support for our hypotheses regarding language abilities and lesion locations related to sIS, we cannot indisputably establish the validity of our participants’ self-reports within the context of the current study. There is potential concern regarding overall comprehension ability for understanding our questions, as well as self-monitoring ability for making such fine-grained metacognitive judgments regarding internal experiences. With regard to sentence-level comprehension, we took care to use clear, simple language that would maximize participants’ ability to understand the concepts and questions being presented, while also using a slow pace and written/pictorial support. In addition, individuals with severe sentence-level comprehension deficits were not included in the study.

A second, more specific concern about using self-report as a measure of language function involves the integrity of self-monitoring in individuals with aphasia, an impairment of which would negatively impact the reliability of self-reports regarding the anomic experience, i.e., their perception of having retrieved the correct word may simply be wrong. Prior studies have shown that many individuals with aphasia do show preserved error detection for their aloud
speech (R. C. Marshall et al., 1994; Nickels & Howard, 1995b; Oomen et al., 2001). Models of self-monitoring in speech production have largely relied on a comprehension-based monitor, where a speaker monitors his/her own output through the same speech perception pathways used to comprehend others’ speech (Indefrey & Levelt, 2004; Levelt et al., 1999). Both the 2004 and a more recent model include a mechanism by which the accuracy of IS can be monitored prior to spoken output via an “internal loop” between phonological representations and speech perception areas in superior temporal gyrus (see Fig. 2.1, inner loop) (Indefrey, 2011). Given this model, an individual with relatively intact single word auditory comprehension should have similarly spared self-monitoring ability for both aloud and IS. To maximize the likelihood of adequate self-monitoring ability in our participants, we therefore excluded individuals with severely impaired single word auditory comprehension.

There are notable alternative proposals to this comprehension-based model of self-monitoring, however, as criticisms have been raised regarding its ability to account for all available evidence regarding self-monitoring (Nickels & Howard, 1995b; Nozari et al., 2011). In place of a comprehension-based monitor, Nickels & Howard (1995) suggested a pre-articulatory production-based monitor, but this type of model was not fully characterized for many years. Recently, Nozari et al. (2011) expanded upon these prior suggestions by proposing a domain-general, conflict-based monitor that is accomplished within the production system itself, through relaying of information to an executive center (such as the anterior cingulate cortex, ACC) about the level of response conflict at the time of word retrieval. This conflict-based model is also consistent with our model of sIS, since the conflict signal in Nozari et al.’s proposed model is generated within the access stages of word-finding based on the relative activation strength of retrieved forms. As such, these signals would arise prior to the post-lexical output processing
that we predicted (and found) to be impaired in our individuals who reported frequent sIS during anomia.

Our results however, are less consistent with ideas that IS is generated and monitored based on efference copies generated by motor speech regions (Tian & Poeppel, 2013). In this case, severe verbal apraxia and lesions in the ventral motor and premotor areas would be expected to result in less sIS, not more, as we found in our analyses. Another recent functional imaging study on unimpaired participants found monitoring of both inner and aloud speech was associated with increased activation in a domain-general error detection area, the ACC, as well as the posterior IFG (Gauvin, De Baene, Brass, & Hartsuiker, 2015). These results suggest that lesions in the posterior IFG might cause unreliable self-monitoring, not a relationship with frequency of sIS as we found. These fMRI studies do not specifically establish a necessary role of IFG in monitoring of IS, so this seeming conflict may simply reflect a difference in the type of evidence provided by different methods. Clearly, additional research on mechanisms of IS self-monitoring will be needed to fully understand the psychological and neural bases of subjective experiences of anomia.

2.4.5 Clinical implications of self-reported IS

The findings of this study demonstrate that many individuals with aphasia can provide reliable, nuanced reports of their experience of anomia, when given a structured opportunity to do so. These findings complement our recent report that self-reported IS on individual naming attempts corresponds to success of naming aloud, error types, and rate of relearning during subsequent anomia treatment (Hayward et al., 2016). Although not expected to be universally reliable, evidence for meaningful self-monitoring of IS raises interesting questions about the potential for self-cueing in aphasia, an approach that has received mixed support from the
treatment literature (Tompkins, Scharp, & Marshall, 2006). It is possible that a deeper understanding of sIS could inform candidacy for such treatment paradigms.

2.4.6 Limitations and future directions

Our method for probing the subjective experience of anomia was newly designed for the current study and certain limitations were revealed during data collection. As mentioned in sections 2.1.2. and 2.4.2., we defined sIS to refer to one very specific experience of anomia, whereas IwW includes various different types of experiences. Additionally, participants were asked to use a categorical scale when identifying how frequently they experienced the three scenarios – never, rarely, sometimes, often, and almost always – it is possible that individual respondents interpreted the scale differently. We did ask participants to use these ratings with regard only to experiences of anomia rather than all communication attempts in general, but we cannot be certain that all participants did so. Although such inconsistencies did not impact our ability to identify meaningful relationships between subjective anomia ratings, behavioral performance, and lesion locations, future research should address these concerns by including a more specific set of experiences under the heterogeneous IwW scenario and using a continuous frequency scale rather than a categorical scale.

We were also limited in this initial study by the use of language and cognitive measures collected for other purposes, and thus not prospectively selected to test our hypotheses on self-perceived IS. For example, this pilot study used a relatively short naming task that did not provide enough error responses for meaningful analysis. Future studies will address these limitations by using behavioral tasks designed specifically to address these issues.
2.5 Conclusions

Although preliminary in nature, the relationships discovered between subjective and objective measures suggest that participants with aphasia may have the metacognitive ability to distinguish between different subjective experiences of anomia. The findings suggest that self-reported sIS in individuals with anomia is associated with poor post-lexical phonological output processing as well as with lesions in the dorsal stream brain regions that support those language processes. More generally, our results suggest that many people with aphasia can often serve as reliable sources of information about their experience of anomia, which may enable researchers and clinicians to better diagnose and treat anomia on an individual basis.

2.6 Acknowledgements

The authors would like to extend their sincere gratitude to the study participants for their time, commitment, and willingness to speak openly about their experience with aphasia. We would also like to thank Dr. Laura Skipper-Kallal for data collection and for providing hands-on training in SVR-LSM to the first author. Finally, we are grateful to the funding that allows us to do this work. M.E.F. is supported by National Institutes of Health (NIH) Grant F31DC014875. W.H. is supported by NIH Grant F30DC014198. P.E.T. is supported by NIH/NIDCD R03DC014310, NIH/NCATS via the Georgetown-Howard Universities Center for Clinical and Translational Science (KL2TR000102), the Doris Duke Charitable Foundation Grant 2012062, and the Vernon Family Trust.
Figure 2.1. Simplified model of naming. Simplified model of naming, demonstrating a distinction between access, which is necessary and sufficient for sIS, and output, which is then required for successful aloud naming.
Figure 2.2. Picture support for the concepts of sIS and IwW. No pictorial support was given for ToT.
Figure 2.3. sIS and other anomia ratings.
Figure 2.4. Lesion-symptom mapping results. A. Lesion overlay map (N=36). B. SVR-LSM results showing lesion locations associated with higher reports of sIS, IwW, and ToT (p < .01, see figure legend for color representations). C. SVR-LSM results at p < .10 to illustrate the minimal overlap between lesions associated with sIS and IwW. Total lesion volume was controlled in both maps, using the direct lesion volume control method.
Table 2.1. Relationship between anomia ratings and behavioral language measures. Correlation coefficients are shown. Correlations with subjective anomia ratings (first three columns) are Spearman’s rho; correlations among language scores (last four columns) are Pearson’s r. Statistical significance denoted as follows: *significant at P<.05; **significant at P<.01; ***significant at P<.001. The three correlations between the subjective anomia ratings are a priori tests and do not require correction for multiple comparisons. All other tests are considered exploratory, so uncorrected p-values are shown.

<table>
<thead>
<tr>
<th>Correlation coefficients</th>
<th>sIS</th>
<th>IwW Correlation</th>
<th>ToT Correlation</th>
<th>WAB-R AQ Correlation</th>
<th>PNT Correlation</th>
<th>ASRS Correlation</th>
<th>Lexicality effect Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IwW</td>
<td>-0.059</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToT</td>
<td>0.447**</td>
<td>0.359*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAB-R AQ</td>
<td>-0.393*</td>
<td>0.175</td>
<td>-0.130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNT</td>
<td>-0.344*</td>
<td>0.014</td>
<td>-0.164</td>
<td>0.863***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASRS</td>
<td>0.384*</td>
<td>-0.106</td>
<td>0.265</td>
<td>-0.655***</td>
<td>-0.547***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexicality effect (output phonology)</td>
<td>0.343*</td>
<td>-0.199</td>
<td>0.116</td>
<td>0.153</td>
<td>0.192</td>
<td>-0.221</td>
<td>-0.211</td>
</tr>
<tr>
<td>Lexical/semantic composite</td>
<td>-0.252</td>
<td>-0.023</td>
<td>-0.199</td>
<td>0.715***</td>
<td>0.668***</td>
<td>-0.438**</td>
<td>0.151</td>
</tr>
</tbody>
</table>
Table 2.2. SVR-LSM results. SVR-LSM results showing significant clusters (threshold = 200 mm$^3$) at a threshold of P<.01.

<table>
<thead>
<tr>
<th>Anatomical description of cluster location</th>
<th>Peak voxel</th>
<th>Number of voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sIS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior frontal gyrus (IFG), pars opercularis</td>
<td>(-58, 6, 4)</td>
<td>93</td>
</tr>
<tr>
<td>Ventral precentral gyrus</td>
<td>(-57, -4, 9)</td>
<td>1097</td>
</tr>
<tr>
<td><strong>IwW</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angular gyrus</td>
<td>(-46, -54, 24)</td>
<td>266</td>
</tr>
<tr>
<td>Posterior middle temporal gyrus (MTG)</td>
<td>(-54, -34, -4)</td>
<td>1414</td>
</tr>
<tr>
<td><strong>ToT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG, dorsal portion of pars triangularis</td>
<td>(-56, 18, 21)</td>
<td>89</td>
</tr>
<tr>
<td>IFG, medial portion of pars opercularis (near inferior frontal junction)</td>
<td>(-38, 9, 26)</td>
<td>86</td>
</tr>
</tbody>
</table>
CHAPTER 3

The subjective experience of inner speech in aphasia is a meaningful reflection of lexical retrieval

3.1 Introduction

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 suffering from 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 et al., 2017; Feinberg et al., 1986; Geva, Jones, et al., 2011; Geva, Bennett, et al., 2011; Goodglass et al., 1976; Hayward et al., 2016; Stark et al., 2017). In this study, we will examine the relationship of self-reported IS to objective measures of the mental process of naming in order to better understand the validity and meaning of the experience of IS in aphasia.

Inner speech (IS) has been intermittently studied in aphasia, beginning with a study that examined a related phenomenon, the tip-of-the-tongue experience (ToT) (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, syllable counting, and finally target word identification from multiple choices. Results showed that individuals with Broca’s and conduction aphasia both demonstrated relatively intact phonological knowledge along with their
frequent ToT experiences (Goodglass et al., 1976). These findings were confirmed by a later study, which used objective measures like picture-based rhyme and homophone judgments as a proxy for IS and showed that performance was preserved in 4 of 5 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).

While these studies have consistently demonstrated relatively preserved IS alongside speech production deficits, several open questions remain. Prior studies have drawn conclusions about discrepancies between inner and overt speech without matching the assessments utilized, with different task design 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, Jones, et al., 2011; Geva, Bennett, et al., 2011; Stark et al., 2017). An alternative approach to the study of IS is to examine it in the context of a person’s own experience of it. In fact, the study of IS in healthy language users has often made use of subjective approaches such as questionnaires to elicit self-reports of the experience (Alderson-Day & Fernyhough, 2015; Hurlburt et al., 2016; Morin et al., 2011). Many individuals with aphasia spontaneously provide evidence for some level of metalinguistic awareness, making comments like “I know it but I can’t say it” (Blanken et al., 1987; Martin & Dell, 2007), suggesting that they may be an informative source regarding their internal experience in the context of anomia.
In our work, therefore, 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 (Chapter 2, Fama et al., 2017; Hayward et al., 2016). Here, we define the subjective experience of IS as the feeling 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 demonstrated (in two individuals with aphasia) that self-report of successful IS 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, where self-reported successful IS was again related to evidence of phonological knowledge (Hayward, 2016). In a larger participant group, we have shown that the general experience of successful IS followed by overt anomia is common in aphasia, is distinct from other anomic experiences (e.g., a vaguer sense of “knowing it”), and is associated with lesions primarily in ventral sensorimotor cortex, a brain region that supports speech output processes (Chapter 2, Fama et al., 2017).

In this study, we aim to further examine the validity of self-reported successful IS in people with aphasia by comparing subjective reports of the experience to 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 and for post-lexical output processing (Dell & O’Seaghdha, 1992b; Dell et al., 1997; Indefrey & Levelt, 2004; Levelt et al., 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.

To test the relationship between self-reported IS and lexical access, we compare self reports of IS on a silent picture naming task to 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 claims to be able 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, or 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.

3.2 Method
3.2.1 Participants

Participants for this study included adults in the chronic (>6 months) stage of recovery from left-sided stroke. We had several participants with 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 48/60 on the “Yes/No Questions” subtest of the Western Aphasia Battery – Revised (Kertesz, 2006), since individuals who perform in this range 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 cut-off. One additional participant was excluded from analysis due to having scored near floor performance on all tasks in the battery (despite meeting the comprehension cut-off). Thus, the final participant group was comprised of 53 participants, 22 women and 31 men, with an average age of 60.2 years (SD=9.8, range 40-80), average education of 16 years (SD=2.8, range 12-24), and average time since stroke of 5 years (SD=4.8, range 0.5-22.9), with handedness as follows: 46 right-handed, 6 left-handed, and one ambidextrous. All participants underwent an informed consent process that was approved by the Georgetown University Institutional Review Board.

3.2.2 Session format

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

3.2.3 Language testing battery

3.2.3.1 Task stimuli and norming procedures

The primary stimuli were 60 words selected to vary in length, frequency, age of acquisition (AoA), and articulatory complexity (Table 3.1). The list included 20 each of 1-, 2- and 3-syllable words. Length was measured in phonemes. Frequency was calculated as the log10 of the frequency identified in the SUBTLEX-US database, which is based on spoken English (Brysbaert & New, 2009). AoA was drawn from a database of 30,000 English words with self-
reported age of acquisition obtained via Amazon Mechanical Turk (Kuperman et al., 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, non-initial 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 and Syllable Counting (see below), the 60 item list was split into two matched sets (Sets A and B, Table 3.1) 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, for a total of 120 items on each (Roach et al., 1996) (see Table 3.1 for word features of this 120-item stimulus list).

All norming was performed in healthy older adult controls who were native English speakers and had no history of developmental learning disability, neurological disorder, or major psychiatric illness. For the picture stimuli, we utilized black-and-white line drawings that were previously normed in-house in a set of 24 healthy older adult controls (average age 54 years old (SD=9.1), average education 15.4 years (SD=2.3)), with at least 70% name agreement during confrontation naming. Novel tasks in the language battery using these picture stimuli were normed in a separate set of 20 healthy older adult controls (average age 65.7 years old (SD=8.4), average education 17.2 years (SD=2.6)). The primary purpose of this norming was to determine whether there were any problematic items on the tests, as judged by incorrect performance by at
least 5/20 healthy controls. No individual items met this criterion and all original stimuli were maintained for patient testing.

3.2.3.2 IS Report

Items were presented one at a time 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. The test items advanced automatically upon key press. The 120 stimulus pictures for the IS Report task were split into two sets of 60 items (Sets 1 and 2), each including 30 in-house items and 30 PNT items. These sets were matched on all four word features: frequency, AoA, length, and articulatory complexity. Each set was administered on a separate day of testing, counterbalanced for order across participants. This task was administered using PsychoPy presentation software (Peirce, 2009), as were all other tasks in the battery with the exception of repetition.

3.2.3.3 Spoken Naming

For the spoken naming task, stimulus pictures were presented one at a time on a laptop screen and participants were given up to 20 seconds to name the picture, with instructions asking participants to “please use only one word.” Participants advanced the test items by pressing the space bar, either once they were satisfied with their response or after the 20 seconds had ended and the item disappeared from the screen. Participants were not given any explicit instructions around self-monitoring or self-correction. All sessions were videotaped. Two independent raters (authors MEF and MPH) coded naming accuracy and in the case of coding discrepancies, 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. Finally, all errors were scored for the presence of spontaneous detection and correction. Detection was identified if the participant verbally rejected a response (e.g., “apple – no”) or produced a second naming attempt that differed from the initial response. Correction was scored if any subsequent attempts at naming resulted in production of the correct target.

The same two 60-item sets of stimuli used for IS Report were also used for Spoken Naming, for a total of 120 items. If set 1 was used for IS Report on day one of testing, set 2 would be used for Spoken Naming, with the opposite sets then being used on day two of testing. This structure ensured that after two 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.

3.2.3.4 Repetition

The repetition task was performed using pre-recorded stimuli in a natural speaker’s voice (author MEF), 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-second inter-trial interval. When participants demonstrated a need for additional time, the audio recording was paused manually. All sessions were videotaped and items were scored as correct/incorrect, using video review as needed.

3.2.3.5 First letter identification 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 and the alphabet (in order, in lowercase letters in Arial, a sans-serif font) and the participant was allowed to point to the correct answer rather than giving a verbal response. If participants self-corrected spontaneously, their final answer was accepted as correct.

3.2.3.6 First letter identification 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 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. As with the picture-based version, spontaneous self-correction was allowed.

3.2.3.7 Picture description

Participants were shown the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination (Goodglass et al., 2001) and were given unlimited time and the following instructions: “Tell me everything you see going on in this picture.” Their narratives were recorded with a video camera and then transcribed into a text file. The work of the primary transcriber (author MPH) was checked in its entirety by a second rater (author MEF). 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 two seconds indicates an utterance break). Each plain text file was analyzed by an automated script that was written to extract the average number of words per utterance and average number of words
spoken per minute, along with a count of the total real words, unique real words, total non-words, and unique non-words produced.

3.2.3.8 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 twenty items including questions that are biographical, environmental, and non-contextual/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 et al. (in press), adapted originally from the Philadelphia Comprehension Battery for Aphasia (Martin, Minkina, Kohen, & Kalinyak-Fliszar, n.d.; Saffran, Schwartz, Linebarger, Martin, & Bochetto, 1988). Participants hear a word and point to the target item in a field of four (all semantic foils).

3.2.4 Statistical analyses

Since IS reports were provided on two different days (at least ten days apart), we first assessed how consistent participants’ subjective judgments were across the two sessions by comparing IS Report scores from day one and day two of testing. We also examined differences in day one and day two scores on Spoken Naming, in order to compare IS Report consistency to this more objective measure.

Our goal for the primary analysis was then to examine whether individual participants had more information about the phonology of words they claimed to be 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 could not be included in the analyses. We set an arbitrary cutoff at 5 words, which excluded 26 participants from subsequent analyses. Because
such a large number of participants were excluded, 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: 13 women and 14 men, with an average age of 61.9 years (SD=9.9, range 40-80), average education of 15.4 years (SD=2.4, range 12-20), and average time since stroke of 4.5 years (SD=4.7, range 0.5-22.9), with handedness as follows: 22 right-handed, 4 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 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 Identification (IS-based and auditory), and Syllable Counting (IS-based and auditory). The use of generalized linear mixed effects models was chosen in order include random effects of item and participant as well as the following fixed effects: item-level IS report, participant features (age, education, and chronicity), and word features (frequency, AoA, length, and articulatory complexity). For all analyses, 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 (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 vs. unsuccessful on the IS Report task; hereafter these categories will be denoted as successful IS (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 vs. 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 in a lack of better performance for sIS vs. uIS words.

3.3 Results

3.3.1 Overall performance on IS Report and the language tasks

In comparing IS Report scores for day one and day two for each of the 53 participants, we found that across the group, participants’ scores changed very little out of the total possible 60 points per day, with a median change of 1 point across days ($IQR=0-5$) (Figure 3.1). This is comparable to the day-to-day performance on Spoken Naming, which had a median change of 3 items across days ($IQR=2-5$). Notably, there are three outliers with a difference of $>30$ items across the two days on the IS Report task, which suggests unreliable reporting on at least one of the two days for those three participants.

As noted above, 26 participants were excluded from the planned analyses because they reported successful IS on nearly every trial ($>115/120$). 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 over-reported their IS compared to the group included in the analysis. In the first case, 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, we found several important differences between them (Table 3.2). First, the excluded group performed significantly better on the Spoken Naming task by an average of 35 words, a difference that is strikingly similar to the difference in self-reported IS, where the excluded group claimed to have sIS for an average of 32 more words. 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 performed significantly better on the two picture-based tasks measuring word retrieval, but not on the matched auditory versions. The two groups did not significantly differ in performance on general language measures of auditory comprehension, fluency, or spontaneous error detection/correction (Table 3.2).

3.3.2 Relationship between word features and task performance

In the final group of included participants (N=27), we were first interested in identifying which 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 age of acquisition,
frequency, and length, but not articulatory complexity (prior to removing the articulatory complexity from the model as a non-significant predictor, its Z-score value was Z=.842, P=.399).

In contrast, Spoken Naming accuracy related to all four of the word features we examined, including articulatory complexity (Table 3.3).

3.3.3 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 to 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’ accuracy was more than 25% higher for sIS vs. uIS items on the three picture-based tasks, but this raw difference was less than 4% for the matched auditory tasks (Figure 3.2, with means and SDs provided in Table 3.4). Average performance on sIS items on the first letter identification (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, age of acquisition, 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 (Table 3.4). 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 (Table 3.3). A Chi-squared 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), $\chi^2(1)=60.573$, $P<.001$.

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 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 vs. 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 Methods, 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 (Figure 3.1) and the three participants who did not consistently show better performance for sIS vs. uIS words. When examining performance of these participants compared to the rest of the group, there were a few important differences (Table 3.5). 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 identification (both picture-based and auditory) and syllable counting tasks (picture-based only). There were no significant differences in overall IS Report, sentence-level auditory comprehension, repetition, or fluency measures.
3.4 Discussion

The primary aim of this study was to determine whether self-reported IS is meaningful on an item-by-item basis 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.

3.4.1 The subjective experience of inner speech 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 identification and syllable counting tasks, participants performed nearly as well on words that they reported as successful IS as when the words were presented to them on the matched auditory tasks, which suggests that the experience of successful IS 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 (D. 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; thus, since first letter identification and 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 post-lexical 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 post-lexical phonology. An impairment at that level would spare repetition ability, since repetition can occur in a non-lexical route based on activation of post-lexical 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 ToT 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 age of acquisition (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 et al., 1984; Kittredge et al., 2008). Similar patterns are observed for AoA (Brysbaert & Ellis, 2016; Jescheniak & Levelt, 1994; Morrison et al., 1992; Snodgrass & Yuditsky, 1996) and the effects of AoA have actually been suggested to supersede frequency effects, e.g., when both variables are included simultaneously in a regression model predicting naming accuracy (Nickels & Howard, 1995a). 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 (Oppenheimer & Dell, 2010) suggesting that IS arises from lexical (phonological) retrieval.
3.4.2 The subjective experience of inner speech 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 post-lexical 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 did find 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 found a significant relationship between IS Report and word length, measured in phonemes. Although we did not predict it, this relationship is not incompatible with our hypothesis. While most studies describe the word length effect as post-lexical 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) (M. F. Schwartz et al., 2004). Since remote errors commonly arise during lexical retrieval, the authors conclude that length effects can arise during lexical retrieval in addition to post-lexical processing (the more common origin of proximate errors) (M. F. 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 that 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 post-lexical 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 left ventral sensorimotor cortex, which supports speech output processes (Chapter 2, Fama et al., 2017).

Importantly, the role of articulation in IS is not relevant only 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 pre-articulatory 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, 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). Even more recently, neuroimaging work in healthy adults has reexamined 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 inner speech prior to overt articulation (Tian & Poeppel, 2013, 2015).

Our findings would be 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.

3.4.3 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. Self-monitoring 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 (R. C. Marshall et al., 1994; Nickels & Howard, 1995a; Schuchard, Middleton, & Schwartz, 2017; M. F. Schwartz et al., 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, since our measure of reliability requires an adequate number of both successful and unsuccessful IS 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 (three of 27) may have been unreliable in their IS reporting, as evidenced by a lack of better performance on items reported as successful.
IS vs. those reported as unsuccessful IS. Three additional participants showed significant variability across the two days of IS Report testing, which could also be evidence of unreliability. This group of six participants with some potential 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.

The group of participants who were 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. In 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 (1) how processing models account for the monitoring of IS in general, and (2) the potential impacts of aphasia on self-monitoring ability.

There is current debate as to the processing mechanisms and neural bases through which self-monitoring of spoken output is achieved. A longstanding 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 we 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 vs. 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 et al., 2011). This model was developed based on the two-step interactive processing model of word retrieval (Dell & O’Seaghdha, 1992b), which does not account for post-retrieval speech production, so the conflict-based monitor is acting within the stages of word retrieval (Nozari et al., 2011).

It is critical to note that 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-based and production-based monitors. If a feeling of sIS arises in conjunction with successful lexical (i.e., phonological) retrieval, then this phonological form would be available prior to spoken output to either 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 why 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 pre-articulatory monitoring systems, as observed in a prior study directly comparing pre- and post-articulatory monitoring in individuals with Broca’s aphasia (Oomen et al., 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 pre-
articulatory, 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.

3.4.4 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 (Chapter 2, 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 et al., 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 will be learned faster, which might inform stimulus
selection during treatment. For instance, a clinician could focus on stimuli that were reported as
sIS in order to make treatment more efficient overall, or he/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; A. M. Raymer,
Thompson, Jacobs, & Le Grand, 1993), since 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; Julie L. 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/her retrieval is
relatively intact might benefit from the use of 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 letterboard, 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.

3.4.5 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 successful IS 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.
Figure 3.1. Day to day variability of scores on IS Report and Spoken Naming. IS Report and Spoken Naming tasks were administered in two 60-item sets across two testing sessions at least ten days apart. The horizontal axis represents the difference in scores across these two days.
Figure 3.2. Item-level performance for words reported as successful or unsuccessful 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.
Table 3.1. Word features for the 60-item and 120-item stimulus lists. 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 IS Report.

<table>
<thead>
<tr>
<th></th>
<th>Length (phonemes)</th>
<th>Frequency</th>
<th>AoA</th>
<th>Articulatory complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>60-item list</td>
<td>mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.47 (1.65)</td>
<td>2.69 (.61)</td>
<td>5.75 (1.26)</td>
<td>3.55 (1.41)</td>
</tr>
<tr>
<td></td>
<td>range</td>
<td>2 – 9</td>
<td>1.2 – 4.1</td>
<td>3.3 – 8.7</td>
</tr>
<tr>
<td>Set A (30-item subset)</td>
<td>mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.33 (1.52)</td>
<td>2.78 (.68)</td>
<td>5.54 (1.21)</td>
<td>3.37 (1.38)</td>
</tr>
<tr>
<td></td>
<td>range</td>
<td>2 – 9</td>
<td>1.2 – 4.1</td>
<td>3.9 – 8.4</td>
</tr>
<tr>
<td>Set B (30-item subset)</td>
<td>mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.6 (1.79)</td>
<td>2.6 (.53)</td>
<td>5.97 (1.28)</td>
<td>3.73 (1.47)</td>
</tr>
<tr>
<td></td>
<td>range</td>
<td>3 – 8</td>
<td>1.5 – 3.6</td>
<td>3.3 – 8.7</td>
</tr>
<tr>
<td>120-item list (Spoken Naming and IS Report)</td>
<td>mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.01 (1.75)</td>
<td>2.86 (.61)</td>
<td>5.22 (1.31)</td>
<td>3.41 (1.55)</td>
</tr>
<tr>
<td></td>
<td>range</td>
<td>2 – 10</td>
<td>1.2 – 4.4</td>
<td>2.5 – 8.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 – 8</td>
</tr>
</tbody>
</table>
Table 3.2. Task scores for general language measures. Overall task results for the group of participants who were included in further analyses (N=27) vs. the group who was excluded due to a score >115 on the IS Report task. Independent samples t-tests were performed to compare scores for the two main participant groups.

<table>
<thead>
<tr>
<th>Task</th>
<th>Possible score</th>
<th>Participants included in the analyses (N=27)</th>
<th>Participants excluded due to sIS &gt;115 (N=26)</th>
<th>Group comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS Report</td>
<td>120</td>
<td>86.33 (26.9)</td>
<td>118.85 (1.38)</td>
<td>t(26.141)=6.266, p&lt;.001</td>
</tr>
<tr>
<td>Spoken Naming</td>
<td>120</td>
<td>49.19 (34.33)</td>
<td>84.35 (33.48)</td>
<td>t(51)=3.773, p&lt;.001</td>
</tr>
<tr>
<td>Error detection (proportion out of total errors made)</td>
<td>1</td>
<td>.28 (.24)</td>
<td>.31 (.25)</td>
<td>t(51)=.565, p=.575</td>
</tr>
<tr>
<td>Error correction (proportion out of total errors detected)</td>
<td>1</td>
<td>.35 (.31)</td>
<td>.52 (.33)</td>
<td>t(48)=1.826, p=.074</td>
</tr>
<tr>
<td>WAB Yes/No Questions</td>
<td>60</td>
<td>55.78 (3.56)</td>
<td>56.19 (3.56)</td>
<td>t(51)=.424, p=.673</td>
</tr>
<tr>
<td>Lexical Comprehension</td>
<td>48</td>
<td>43.52 (5.42)</td>
<td>45.92 (4.45)</td>
<td>t(49.777)=1.760, p=.083</td>
</tr>
<tr>
<td>Repetition (single words)</td>
<td>60</td>
<td>41.41 (15.0)</td>
<td>47.42 (16.02)</td>
<td>t(51)=1.412, p=.164</td>
</tr>
<tr>
<td><strong>Fluency measures from picture description task</strong></td>
<td></td>
<td>Average words per minute unlimited</td>
<td>39.55 (27.4)</td>
<td>t(51)=1.534, p=.131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average words per utterance unlimited</td>
<td>4.12 (2.27)</td>
<td>t(51)=1.158, p=.252</td>
</tr>
<tr>
<td><strong>Primary IS tasks</strong></td>
<td></td>
<td>First letter identification: <strong>picture-based</strong></td>
<td>30</td>
<td>17.3 (9.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>First letter identification: auditory</td>
<td>30</td>
<td>21.85 (9.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Syllable counting: <strong>picture-based</strong></td>
<td>30</td>
<td>17.15 (7.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Syllable counting: auditory</td>
<td>30</td>
<td>22.30 (7.76)</td>
</tr>
</tbody>
</table>
Table 3.3. **Word feature comparison across tasks.** Means (SDs) for sIS vs. uIS items on the IS Report task and correct vs. 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 analysis here due to a score of 0 on this task). Z-scores are provided with significance reported as follows: ***P<.001, **P<.01, and *P<.05.

| Word feature          | Inner Speech Self-Report |                      |  | Spoken Naming |                      |
|-----------------------|---------------------------|----------------------|  |              |----------------------|
|                       | sIS Items | uIS Items | LMEM Z-score | Correct Items | Incorrect Items | LMEM Z-score |
| Frequency              | 2.93 (.12) | 2.64 (.15) | 3.661***     | 3.08 (.16)     | 2.69 (.12)     | 5.40***       |
| Age of acquisition     | 5.06 (.12) | 5.72 (.43) | -4.32***     | 4.78 (.33)     | 5.54 (.28)     | -6.14***      |
| Length                 | 4.85 (.17) | 5.49 (.64) | -2.33*       | 4.49 (.40)     | 5.35 (.33)     | -3.47***      |
| Articulatory complexity| 3.36 (.12) | 3.59 (.40) | n.s.         | 3.04 (.35)     | 3.62 (.18)     | -2.45*        |
Table 3.4. Results of generalized linear mixed effects models predicting task performance.
For each task, means and SDs are shown for the average proportion of correct performance for sIS items and 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 (AIC) and Bayesian Information Criterion (BIC). Significance is reported as follows: ***P<.001, **P<.01, and *P<.05.

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean Accuracy (SD)</th>
<th>LMEM Z-scores for Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sIS Items</td>
<td>uIS Items</td>
</tr>
<tr>
<td>Picture-based tests of lexical retrieval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spoken naming</td>
<td>.47 (.28)</td>
<td>.23 (.25)</td>
</tr>
<tr>
<td>First letter ID</td>
<td>.67 (.32)</td>
<td>.35 (.30)</td>
</tr>
<tr>
<td>Syllable counting</td>
<td>.64 (.27)</td>
<td>.38 (.27)</td>
</tr>
<tr>
<td>Matched auditory control tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repetition</td>
<td>.73 (.25)</td>
<td>.67 (.31)</td>
</tr>
<tr>
<td>First letter ID</td>
<td>.76 (.31)</td>
<td>.70 (.38)</td>
</tr>
<tr>
<td>Syllable counting</td>
<td>.74 (.28)</td>
<td>.75 (.30)</td>
</tr>
</tbody>
</table>
Table 3.5. Task scores for general language measures: reliability analysis. Overall task results for the group of participants who revealed some evidence of unreliable IS reporting (N=6) compared to 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.

<table>
<thead>
<tr>
<th>Task</th>
<th>Possible score</th>
<th>Participants with some evidence of unreliable IS reporting (N=6)</th>
<th>Remaining participants in the included group (N=21)</th>
<th>Group comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS Report mean (SD)</td>
<td>120</td>
<td>74.5 (26.28)</td>
<td>89.71 (26.76)</td>
<td>t(25)=−1.233, p=.229</td>
</tr>
<tr>
<td>Spoken Naming</td>
<td>120</td>
<td>10.5 (5.75)</td>
<td>60.24 (30.77)</td>
<td>t(23.77)=−6.993, p&lt;.001</td>
</tr>
<tr>
<td>Error detection (proportion out of total errors made)</td>
<td>1</td>
<td>.09 (.07)</td>
<td>.33 (.25)</td>
<td>t(25)=−2.208, p=.037</td>
</tr>
<tr>
<td>Error correction (proportion out of total errors detected)</td>
<td>1</td>
<td>.03 (.05)</td>
<td>.44 (.29)</td>
<td>t(23.19)=−6.269, p&lt;.001</td>
</tr>
<tr>
<td>WAB Yes/No Questions</td>
<td>60</td>
<td>55 (3.63)</td>
<td>56 (3.59)</td>
<td>t(25)=−.600, p=.554</td>
</tr>
<tr>
<td>Lexical Comprehension</td>
<td>48</td>
<td>38.33 (6.28)</td>
<td>45 (4.24)</td>
<td>t(25)=−3.050, p=.005</td>
</tr>
<tr>
<td>Repetition (single words)</td>
<td>60</td>
<td>39.67 (15.58)</td>
<td>41.9 (15.18)</td>
<td>t(25)=−.317, p=.754</td>
</tr>
<tr>
<td>Fluency measures from picture description task</td>
<td>unlimited</td>
<td>25.45 (21.26)</td>
<td>43.58 (28.04)</td>
<td>t(25)=−1.460, p=.157</td>
</tr>
<tr>
<td>Average words per minute unlimited</td>
<td>3.31 (2.43)</td>
<td>4.36 (2.22)</td>
<td></td>
<td>t(25)=−1.008, p=.323</td>
</tr>
<tr>
<td>Primary IS tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First letter identification: picture-based</td>
<td>30</td>
<td>5.5 (4.51)</td>
<td>20.67 (6.84)</td>
<td>t(25)=−5.089, p&lt;.001</td>
</tr>
<tr>
<td>First letter identification: auditory</td>
<td>30</td>
<td>9.33 (9.22)</td>
<td>25.43 (6.87)</td>
<td>t(25)=−4.7, p&lt;.001</td>
</tr>
<tr>
<td>Syllable counting: auditory</td>
<td>30</td>
<td>17.33 (10.39)</td>
<td>23.71 (6.46)</td>
<td>t(25)=−1.860, p=.075</td>
</tr>
</tbody>
</table>
CHAPTER 4

Self-reported inner speech relates to individual differences in phonological retrieval in people with aphasia

4.1 Introduction

In the previous chapter, we examined the subjective experience of inner speech (IS) at the level of individual items and found that self-reported IS closely related to lexical retrieval but did not relate to articulation. The item-level approach required us to limit the analysis to only those participants (N=27) who reported enough items as successful IS (sIS) and unsuccessful IS (uIS) in order for statistical comparisons to be made. A large group was excluded due to near-ceiling levels of sIS, so we compared their performance on other language tasks to the group that reported lower, more varied levels of sIS. We were able to show that the group reporting high sIS performed better on naming and other picture-based tasks requiring word retrieval, suggesting that it is likely that their IS reports accurately reflected relatively intact word retrieval. These preliminary findings raise the general question of what types of individuals with aphasia report more or less sIS, in terms of their objective performance on other language tasks.

In this chapter, we approach the study of self-reported IS at the level of overall task performance, which allows us to address questions across the entire participant group and identify what objective language behaviors correspond to more frequent experience of sIS during silent picture naming.

Our hypotheses, as previously outlined, are that the subjective experience of sIS is associated with lexical retrieval and that output processing (for example, motor articulatory planning) is not required for sIS. In this study, we will compare overall reports of sIS on a silent-picture naming task to performance on a battery of language tests, which were designed to emphasize various aspects of the mental process of naming. In addition, we will compare reports
of sIS to the types of errors made on a spoken naming task. We predict that participants who report greater levels of sIS on the silent picture-naming task should do well on tasks requiring retrieval and should be more likely to make phonologically related errors (vs. semantic errors) during a naming task. We do not expect there to be a relationship between sIS and performance on tasks that emphasize output processing, such as oral reading or repetition. In contrast, we expect participants who demonstrate good spoken naming ability to perform well on both tasks requiring lexical retrieval and tasks that are focused on output, since successful spoken naming requires both sets of processes.

4.2 Method
4.2.1 Overview

A significant portion of the methods for this study overlaps directly with the methods from the previous chapter. Participants were the same 53 adults with chronic post-stroke aphasia as described in Chapter 3 (Section 3.2.1.1), who completed a language assessment battery across two sessions, at least ten days apart. The task stimuli and norming procedures were as described in Chapter 3. Although the language tasks also overlap in part with those presented previously, there are some important differences, which will be described below.

4.2.2 Language testing battery
4.2.2.1 IS Report and Spoken Naming

These tasks were administered as described in Chapter 3 (Sections 3.2.3.2 and 3.2.3.3).

4.2.2.2 Error coding for the Spoken Naming task

Scoring for the spoken naming test occurred in the following manner: author MEF (a certified speech-language pathologist) scored all responses online during the testing session, then
later reviewed responses on the video recording and assigned error codes (see below). Author MPH independently reviewed all participant videos, transcribed responses, and assigned error codes as appropriate. In the case of discrepancies in transcription, the two raters reviewed the participant video either together or separately to reach a consensus. In the case of discrepancies in error code assignment, the two raters met to discuss, referring to the scoring guidelines of the PNT (Roach et al., 1996). In the event where the two primary raters continued to disagree on either a transcription or an error code, a third listener was consulted.

In all cases, accuracy and error type were assessed based on the first complete attempt at spoken naming. Incorrect responses were coded based on their relationship to the target word, following the scoring guidelines of the PNT (Roach et al., 1996), which include two primary error categories that are of primary interest to our analysis:

- Semantic error: a real word (usually a noun) that is related to the target as a synonym, category coordinate, superordinate, subordinate, diminutive, or associated concept
- Phonological error: a real word or non-word with at least 50% overlap with the phonemes in the target word

Some errors did not fall into one of these two categories and were identified as other error types, including mixed errors (semantic/phonological), unrelated real or non-word errors, descriptions, and no response. These were not examined in the context of this study.

4.2.2.3 Sentence-level auditory comprehension

This task was administered as described in Chapter 3, for the purpose of participant exclusion only.
4.2.2.4 Measures of lexical retrieval

To assess lexical retrieval, we utilized three sets of matched tasks that each included a picture-based version and an auditory version. Each picture-based task required the participant to name the picture in his/her head in order to perform the task, whereas in the auditory tasks, the word was presented aloud to the participant.

- Rhyme judgment – picture-based

Participants completed a silent rhyme judgment task in which two pictures appeared on the screen and they were asked to answer (yes/no) whether the names of the pictures rhymed (40 items total). All items were 1-syllable words. Orthography was considered in addition to phonology, in that stimuli included congruent pairs (i.e., rhyming pairs with identical spelling of the rimes; non-rhyming pairs with different spellings for each rime) as well as incongruent pairs (i.e., non-rhyming pairs with identical spelling of the rimes; rhyming pairs with different spellings for each rime). There were 22 congruent word pairs and 18 incongruent word pairs. The stimuli for this task were not drawn from the main 60-item list due to constraints for single syllable, rhyming word pairs.

During control norming, we identified two problematic word pairs, each of which was subsequently removed from analysis. First, the pair hoof/roof was eliminated due to dialect differences in pronunciation, leading 11/20 control participants to answer “yes” while the other 9/20 control participants answered “no.” Also, for the pair fly/tie, 7/20 control participants incorrectly answered “no.” The final task analysis thus included 38 word pairs.
• Rhyme judgment – matched auditory task

    Participants heard two pre-recorded words and were asked to answer (yes/no) whether the two words rhymed. Although the control norming for this auditory task revealed no problematic word pairs, we eliminated the word pairs hoof/roof and fly/tie, which were removed from analysis in the picture-based version, so that these two rhyme judgment tasks would be fully matched on stimuli (each with 38 pairs of items). Auditory stimuli for this task (and all other auditory tasks in the battery) were presented through high-fidelity headphones.

• First letter identification and syllable counting – picture-based (as described in Chapter 3, Section 3.2.3.5)

• First letter identification and syllable counting – matched auditory tasks (as described in Chapter 3, Section 3.2.3.6)

For the picture-based tasks, performance relies both on the ability to retrieve the name of the picture and to make the judgments needed to accurately perform the task. To isolate the portion of task performance specifically related to phonological retrieval, we subtracted scores on the auditory version of each task, in which the word was provided to the participant and they needed only to make the judgments necessary to perform the task. A higher picture-based – auditory difference score (in most cases, less negative) represents better phonological retrieval ability.

4.2.2.5 Measures of speech output processing

    To assess post-lexical output processing, we included several tasks that rely on speech output processes without requiring lexical retrieval. We used real words in the context of a general motor speech evaluation, but utilized pseudowords during separate repetition and oral reading tasks in order to limit the support of lexical forms during production.
• Pseudoword repetition

The repetition task was performed using pre-recorded stimuli in a natural speaker’s voice, played through Quicktime software on a laptop computer. For this task, pronounceable non-words were generated from the 60 real word stimuli (Chapter 3) by changing at least one phoneme for 1- and 2-syllable words and at least two phonemes for 3-syllable words. Each pseudoword was identical to its matched real word for number of syllables and articulatory complexity, measured using the Word Complexity Measure (Stoel-Gammon, 2010). Each item was presented a single time with a 5-second inter-trial interval. When participants demonstrated a need for additional time, the audio recording was paused manually. All sessions were videotaped and items were scored as correct/incorrect, using video review as needed. Accuracy was scored based on participants’ first complete attempt at repetition.

During control norming, three problematic items were identified on this task. For the pseudoword /gib/, we decided to allow the acceptable alternative /dib/, still a non-word, after 9/20 control subjects produced this same response. The pseudowords /lɛmp/ and /θæsn/ were removed from analysis due to >5 errors out of 20 control subjects, all of which were real word productions (e.g., “limp” and “fasten”), resulting in a total of 58 items for final analysis.

• Pseudoword oral reading

Participants completed a 20-item pseudoword reading task. A single pseudoword appeared on the laptop screen at a time and participants had 10 seconds to read the pseudoword aloud. All items were single-syllable pseudowords composed of 3-4
letters/phonemes. Accuracy was scored based on participants’ first complete attempt at reading the word.

- **Motor speech evaluation**

  To assess motor speech ability, we used a protocol that elicited production of the following: alternating motion rates (AMRs; repeated production of a single syllable tested for three separate syllables: /pʌ/, /tʌ/, and /kʌ/), sequential motion rates (SMRs; repeated production of a syllable triad /pʌtʌkʌ/), multisyllabic words, symmetric consonant–vowel–consonant (CVC) words, words of increasing length (1-3 syllables), and sentences (Haley et al., 2012). The evaluation was videotaped and later reviewed independently by two raters (authors SFS and MEF, both certified speech-language pathologists). As with the assignment of error codes in the naming task, discrepancies were resolved through video review and discussion between the two raters. In the event that the two primary raters were unable to reach a consensus, a third listener who was familiar with the participants was consulted.

  For each utterance produced during the evaluation, the raters identified the presence or absence of the following features: (1) phonemically relevant segment errors, (2) ambiguous/distorted production of consonant voicing and/or place of articulation, (3) slow rate, and (4) segment and/or pause prolongation (Jacks & Haley, 2015). These measures were calculated separately for multisyllabic words and for sentences. Of these features, distorted consonants and prolongations are particularly characteristic of apraxia of speech, whereas the other two often present with language-based deficits and dysarthria (Jacks & Haley, 2015). In our analysis, we focused on four scores (note that we focused on multisyllabic word production rather than sentence production, due to the
large proportion of participants whose sentence repetition attempts were too limited to allow rating of relevant features):

- AMR rate: the mean number of syllables produced per second, averaged across separate attempts for pa, ta, and ka
- Adequate SMR performance: a binary measure reflecting whether or not the participant produced at least three accurate syllable triads (pataka x 3)
- Segmental errors in multisyllabic word repetition: substitution, omission, or addition of one or more phonemes
- Ambiguous/distorted consonant productions in multisyllabic word repetition

4.2.3 Statistical analyses

We first performed a factor analysis to identify the main components of our testing battery to confirm our predictions about the task demands of lexical retrieval vs. output processing. The tests entered into the analysis were as follows: pseudoword repetition, pseudoword reading, the four measures from the motor speech evaluation, and the three difference scores from the matched picture-based/auditory tasks. We implemented a principal components analysis of the correlation matrices, using a standard eigenvalue > 1 cutoff to extract components and Varimax rotation with Kaiser normalization to generate orthogonal factors. The rotated component matrix provided the loadings of each test onto the resulting factors. Factor scores for each subject were determined using the regression method.

The self-reported IS scores (hereafter, IS Report) were left-skewed, so we applied a logit transform (ln(p/(1-p))) to these scores. Using bivariate Pearson’s product-moment correlations, we then examined relationships between the logit-transformed IS Report scores and spoken naming accuracy (hereafter, Spoken Naming) and two sets of measures: the factor scores
resulting from the principal components analysis and the relative proportions of phonological and semantic errors on the spoken naming test. We set our significance threshold at a Bonferroni-corrected p-value of $P<.05$ (corrected for nine bivariate correlations performed). In a final analysis, we used a hierarchical regression approach to examine the relative contributions of the two factor scores and the IS Report score in predicting performance on Spoken Naming, to determine whether IS Report added significant predictive value beyond the objective measures of retrieval and production. All statistical analyses were performed in SPSS 24.

4.3 Results

Examination of average test scores shows that participants as a group reported successful IS that exceeded Spoken Naming accuracy (Table 4.1). On the naming task, phonological errors were more prevalent than semantic errors. Performance on the picture-based retrieval tasks (rhyme judgment, first letter identification, and syllable counting) was lower than performance on each of the auditory matched tasks. On the motor speech evaluation, segmental errors were more prevalent than distorted/ambiguous productions.

The principal components factor analysis yielded two factors, together accounting for a total of 60.4% of the variance in patient performance (Figure 4.1). Rotation converged in three iterations. The tasks requiring speech output (reading, repetition, and motor speech production) loaded primarily onto Factor 1. The retrieval measures, represented by the three Picture-based – Auditory task difference scores, loaded primarily onto Factor 2. These factor scores (Factor 1 and Factor 2) were then utilized in a set of bivariate correlations to assess the relationship of IS Report and Spoken Naming to retrieval and output (Table 4.2). We predicted that IS Report would relate to behavioral measures of lexical retrieval but not output processing, and that
Spoken Naming would relate to both retrieval and output. As predicted, we found that IS Report related only to Factor 2 (retrieval); the relationship between IS Report and Factor 1 (output) is essentially null (R=.094). In contrast, Spoken Naming related to both Factor 1 and Factor 2. Additionally, we found a positive correlation between IS Report and Spoken Naming.

We then examined correlations between both IS Report and Spoken Naming and naming error types, to test the prediction that individuals with high self-reported scores on IS Report should have relatively intact word retrieval. If their IS judgments are an accurate reflection of partial or complete phonological retrieval, they should be more likely to make phonological errors than semantic errors during a confrontation naming test, since phonological errors typically result from failure at the level of phonological representations or in post-lexical output processing (Nickels & Howard, 1995b; Romani, Olson, Semenza, & Granà, 2002; M. F. Schwartz et al., 2004; Wilshire, 2002). As predicted, we found that IS Report related to phonological errors only (Table 4.3). In contrast, we found a significant, positive relationship between overall naming accuracy and semantic errors and a trending relationship between naming accuracy and phonological errors (R=.357, P=.009, Bonferroni corrected P=.08).

In our final analysis, we performed a hierarchical regression analysis with Spoken Naming as the dependent variable to determine if IS Report provides additional explanatory power for naming ability beyond that provided by objective tests of retrieval and output processing. First, the two factor scores (retrieval and production) were entered as independent variables. The overall model was significant (F(2,50) = 26.93, adjusted $R^2 = .499$, $P<.001$), and Factor 1 and Factor 2 were both significant predictors (Factor 1 standardized beta = .602, $P<.001$; Factor 2 standardized beta = .394, $P<.001$). When IS Report was added as an additional independent variable, the model fit improved significantly (Overall model $F(3,49) = 29.45$, $P<.001$).
adjusted $R^2 = .621$, $P<.001$; Model Change $F(1,49) = 17.12$, $R^2$ change = .125, $P<.001$), and only IS Report and Factor 1 (production) were significant predictors (IS report standardized beta = .425, $P<.001$; Factor 1 standardized beta = .563, $P<.001$; Factor 2 standardized beta = .160, $P=.124$).

4.4 Discussion

The goal of the current study was to clarify individual differences in the experience of successful inner speech (sIS) in aphasia. Building on the findings from Chapter 3, we examined the relationship between the subjective experience of IS and task-level performance on objective measures of different aspects of the mental process of naming. We found that the frequency of successful IS reported by an individual relates to their phonological retrieval ability and not to their speech production ability. In contrast, an individual’s success at naming aloud relates to both their phonological retrieval and speech production abilities. These findings strongly support our hypothesis that the subjective experience of IS relates to word retrieval and not to speech output processing.

Acknowledging the validity of self-reported sIS in aphasia has important implications for clinical practice. In Chapter 3, we described ways in which a speech-language pathologist planning anomia treatment could use self-reported sIS to select treatment items or to match treatment approaches to specific items, based on the reported status of IS as successful or unsuccessful. The current findings reinforce the notion that overall levels of self-reported sIS may be informative as to the main cause of anomia in a given patient, since our subjective IS measure related to overall measures of word retrieval. These data, combined with the findings from Chapter 3, suggest that word retrieval is relatively intact in individuals who report frequent
sIS. Of great practical importance, including self-reports of IS along with more traditional objective speech and language measures improved predictions of spoken naming ability, which demonstrates the potential value of using IS judgments to improve anomia diagnosis. In a clinical setting, a speech-language pathologist could utilize a silent-picture naming test, in addition to a spoken naming test, to gather more information about the status of word retrieval. Overall levels of self-reported sIS could be combined with information gathered via other common approaches to anomia diagnosis such as error patterns (Laine & Martin, 2006), response to cueing (Howard & Gatehouse, 2006; Julie L. Wambaugh et al., 2001), and overall pattern of language ability (Goodglass & Wingfield, 1997; Howard & Gatehouse, 2006; Laine & Martin, 2006), in order to provide a more comprehensive and accurate assessment of the locus of word-finding impairment. The current findings show that sIS may provide diagnostic information that is complementary to error patterns and objective tests of speech and language ability. Future research could examine relationships between self-reported sIS and cueing response.

Our studies on the subjective experience of IS in aphasia have focused on spoken naming, but there are other language processes that require stages of both retrieval and production and it is possible that self-reports about internal processing may be useful in these contexts as well. Written naming is a clear analog to spoken naming; individuals with aphasia may be able to report a sense of being able to “see” a word in their heads, which may correspond to successful retrieval of the orthographic representation, regardless of subsequent ability to write that word correctly. Another comparable process is oral reading; it is plausible that some individuals may experience a feeling of being able to read a word in their heads, despite being unable to say those words out loud. Obtaining self-reports about internal experiences during these other tasks may be informative as to the level of processing breakdown occurring in the
contexts of agraphia (i.e., writing impairments) or alexia (i.e., reading impairments). There is also a broader scientific question about the ways in which the mental processes underlying the subjective experience of IS may vary across task contexts in healthy language users (Oppenheim & Dell, 2010).

Importantly, the significant relationship identified between the subjective experience of sIS and word retrieval suggests more generally that these self-reports are often valid and meaningful. Our measure of sIS is entirely subjective in nature so is potentially subject to individual inaccuracy or bias. We found that participants, on average, reported successful IS for a greater number of items than they were able to produce accurately during spoken naming. We expected this to be the case, since it is likely that some people with aphasia have relatively intact word retrieval alongside production deficits, which would cause errors during spoken naming despite a judgment of successful IS. Although this explanation aligns with our hypotheses, we also cannot rule out the possibility that some (or all) participants may be over-reporting the success of their IS to some degree. Given the group-level analysis performed here, we cannot comment on individual variability in self-monitoring for IS in the context of these particular findings; however, in Chapter 2 (Fama et al., 2017) and Chapter 3, the potential impacts of self-monitoring on the accuracy of self-reported sIS were considered in detail. Overall, we found significant statistical relationships between our subjective measure of IS and several predicted objective measures, so it is unlikely that a large proportion of our participants were erroneously reporting their IS (see Figure 3.1 for a measure of day-to-day reliability in reporting).

Consistent with our prior findings showing that self-reported sIS at the item level is a meaningful reflection of phonological retrieval (Chapter 3) (Hayward, 2016), the current findings clearly link the experience of IS to phonological retrieval, and not post-lexical speech
production processes. Our strong hypothesis was that the feeling of successful IS arises from access to a lexical phonological form that matches the target, but this may not be possible to prove definitively. A more conservative conclusion is that successful IS arises from retrieval of a phonological form close enough to the target to be perceived as matching it. Although we isolated retrieval processing by subtracting task performance when the word was provided (auditory tasks), the three picture-based tasks still only require limited phonological knowledge of the word for accurate performance. For example one could perform the first letter identification task by only retrieving the initial phoneme/grapheme, or could perform the rhyme judgment task by only retrieving the coda of each one-syllable word in the pair. Taken as a set, accurate performance of these three tasks requires knowledge of the beginning, end, and syllable number for each stimulus word, but even together they still do not require the full lexical phonological form and some elements of phonology, such as stress patterns, were not examined. The relationship between sIS and the word retrieval factor score thus suggests that the report of successful sIS reflected preserved phonological knowledge in our participants, but it does not necessarily show that successful IS reflects retrieval of the complete phonological form.

The results of the analyses relating IS to error types are also consistent with a relationship of successful IS to retrieval of either complete phonological forms or partially specified forms. Given the hypothesized relationship between the subjective experience of IS and word retrieval, we predicted that IS Report would relate to errors that are phonologically related to the target word (here, defined by at least 50% phonological overlap with the target word). In previous literature, there is widespread agreement that phonological errors occur either during lexical phonological retrieval or during post-lexical output processing (Nickels & Howard, 1995b; Romani et al., 2002; M. F. Schwartz et al., 2004; Wilshire, 2002). The positive relationship we
found between IS Report and phonological errors, therefore, supports a theory in which self-reported sIS represents at least partial retrieval of the lexical phonological form. When naming fails in association with successful IS, it is due either to retrieval of an incompletely specified phonological form or due to errors in post-retrieval output processes after successful phonological retrieval.

Given that tip-of-the-tongue (ToT) sensations are associated with partial phonological access to words (Burke et al., 1991; Dell et al., 1997; James & Burke, 2000; Levelt et al., 1999), the differences between sIS and ToT may help to clarify the question of how much phonological information is needed to give rise to a sense of sIS. The two experiences are categorically different in that with ToT, the individual by definition has not retrieved the word, whereas in describing sIS, individuals report hearing or saying the correct word internally. In Chapter 2, we showed that sIS during anomia was related to a feeling of ToT but that each experience was associated with damage to different frontal brain regions and only sIS was related to measures of phonological output processing (Fama et al., 2017). This suggests that, at a minimum, sIS reflects access to a substantially more complete form than is associated with ToT. In future studies, more information may be provided by allowing participants with aphasia to report either sIS or TOT during silent picture naming. Also, one could implement additional picture-based tasks, such as prosodic tasks requiring participants to identify the stressed syllable of multisyllabic words or recognition tasks where participants are asked to identify whether certain sounds are present in a target word or not, to further clarify the aspects of phonology most relevant to a judgment of successful IS.

In general, there are several open questions about self-monitoring of IS that are also relevant to the broader literature on the experience of IS in healthy language users regarding the
manner in which IS is monitored. Additional research is needed to determine what the threshold is for a judgment of successful IS (or ToT), i.e., how close to the target word the retrieved form must be, and whether this varies across individuals and/or across tasks. Also, future studies are necessary in order to determine the precise nature of the information being monitored, i.e., whether IS judgments are based on auditory images, abstract amodal phonology, orthography, or abstract word-forms (lemmas) – and again, whether the basis of IS judgments varies across individuals and/or across tasks (see Section 5.3 of the Discussion chapter for alternative interpretations and Section 5.6 for additional discussion of these open questions).

A few prior studies have investigated individual differences in IS among individuals with aphasia (Geva, Bennett, et al., 2011; Stark et al., 2017), but they use a definition of IS that differs significantly from our own: “the ability to create an internal representation of the auditory word form, and to apply computations or manipulations to this representation” (p. 1, Stark, Geva, & Warburton, 2017). In such studies, silent rhyme and homophone judgments on written words are used as the measure of IS ability, emphasizing the role of IS in performing mental operations in one’s head. The mental processes that are being measured are thus not directly comparable to the form of IS we have examined in our own study, which is a naturalistic sense of IS that involves the retrieval of words from semantic representations. As the only prior studies specifically examining individual differences in IS among individuals with aphasia, however, their findings warrant consideration here.

When IS is defined as the ability to perform mental manipulations on internal representations, it can be preserved relative to overt speech (measured by oral reading) in individuals with motor speech impairments or deficits in translating from the phonological to the articulatory code (Geva, Bennett, et al., 2011). Such findings are consistent with the data from
Chapter 2, which showed that sIS in the context of anomia relates to phonological output processing and to damage to sensorimotor brain regions supporting speech output (Fama et al., 2017). A more recent study built on those 2011 findings by examining relationships between IS and overt language abilities in groups of individuals with relatively preserved IS, defined as >50% task accuracy on silent rhyme and homophone judgments on written words, alongside either good or poor overt speech (again measured by oral reading). In the group with poor overt speech, there were significant relationships between (1) silent rhyme judgment and spoken naming, which they interpret as an effect of the role of working memory and possibly a reflection of the contribution of IS to pre-articulatory processes and (2) silent homophone judgment and written mean length of utterance, which they ascribe to the use of IS as a buffer for error monitoring or mediating overall task difficulty (Stark et al., 2017). Although the results cannot be directly compared, their tentative conclusion regarding the relationship between IS and articulatory processes may seem at odds with our own findings; however, as mentioned above, they have examined a process of mental manipulation that is fundamentally different from the sense of IS measured in our work. Previous studies have suggested that IS is flexible and can evoke articulatory processing during certain task contexts (Oppenheim & Dell, 2010; Sokolov, 1972), so their conclusions are not incompatible with our own.

One incidental finding of note was a positive relationship between overall Spoken Naming accuracy and the prevalence of semantic errors. This finding, although not specifically relevant to our primary research questions, is of interest because it contrasts with prior studies suggesting that there is no relationship between severity of anomia and semantic errors or that semantic errors are made with equal frequency across aphasia subtypes (Dell et al., 1997; Goodglass & Wingfield, 1997; M. F. Schwartz & Brecher, 2000). There are several reasons why
this relationship, which may seem counterintuitive, may be reasonable in the context of our particular analysis. We examined errors as a proportion of the total number of errors made by each individual participant, so this finding does not indicate that individuals who are good at naming produce a higher number of semantic errors, but rather that a higher proportion of the errors that they do make are semantically related to the target. As illustrated in Chapter 3, there was a large group of participants (among the 53 total in the study) whose language impairments are relatively mild overall. These participants, who would likely be categorized as having anomic aphasia, may have tended toward semantic errors while overtly using circumlocution in search of the target word (Dell et al., 1997) or may have offered a close semantic substitute in cases where they were unable to retrieve the primary target (e.g., “bed” for target crib or “cup” for target glass). Importantly, the negative relationship between severity and semantic errors does not indicate that all participants with mild naming deficits have semantic impairments, as previous literature suggests that semantic errors in production (e.g., picture naming or oral reading) can arise during post-semantic stages of processing (Caramazza & Hillis, 1990; Dell et al., 1997; Rapp & Goldrick, 2000).

4.5 Conclusions

This study provides robust evidence that the experience of sIS is meaningful and is most common among individuals who have relatively spared lexical/phonological retrieval. Participants’ speech output processing abilities support their spoken naming ability, but do not relate to the self-report of sIS. Along with the error patterns associated with sIS, these findings demonstrate that the subjective experience of sIS relates to phonological retrieval. Further, adding IS reports to objective measures improves predictions of spoken naming ability,
demonstrating their potential clinical value. We suggest that self-reported sIS can be an informative addition to the diagnosis of anomia in the context of processing models of naming, which in turn may inform clinical decision-making regarding anomia treatment.
<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Output</td>
<td>Retrieval</td>
</tr>
<tr>
<td>Variance accounted for by factor</td>
<td>39.9%</td>
<td>20.5%</td>
</tr>
<tr>
<td>Pseudoword reading</td>
<td>0.816</td>
<td>0.033</td>
</tr>
<tr>
<td>Pseudoword repetition</td>
<td>0.831</td>
<td>-0.058</td>
</tr>
<tr>
<td>Adequate SMR production (yes/no)</td>
<td>0.791</td>
<td>-0.075</td>
</tr>
<tr>
<td>Average AMR rate (syll/sec)</td>
<td>0.498</td>
<td>-0.178</td>
</tr>
<tr>
<td>Multisyllable word repetition (motor speech evaluation)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmental errors</td>
<td>-0.812</td>
<td>-0.109</td>
</tr>
<tr>
<td>Ambiguous/distorted productions</td>
<td>-0.744</td>
<td>-0.012</td>
</tr>
<tr>
<td>Rhyme judgment: difference score</td>
<td>0.291</td>
<td>0.608</td>
</tr>
<tr>
<td>Syllable count: difference score</td>
<td>-0.126</td>
<td>0.891</td>
</tr>
<tr>
<td>First letter identification: difference score</td>
<td>-0.227</td>
<td>0.794</td>
</tr>
</tbody>
</table>

**Figure 4.1. Results of factor analysis.** Principal components analysis using Varimax rotation and Kaiser normalization to generate orthogonal factors. The two emerging factors relate to our measures of output (factor 1, in blue) and lexical retrieval (factor 2, in purple). Cell color is graded according to the strength of the factor loading with the strongest loadings in dark shades and weakest loadings in white. Each of the difference scores in the bottom three rows was intended to isolate phonological retrieval from other task demands and was calculated by subtracting the auditory task score from the picture-based task score.
Table 4.1. Average performance on the language battery (N=53). For most tasks, mean accuracy scores and standard deviations are represented as a proportion of the total number of items per test. “AMR rate” represents the average syllables per second and “adequate SMR production” illustrates the number of participants who did or did not reach criterion performance (at least three repetitions of the syllable triad /pʌtʌkʌ/). Proportions of phonological errors and semantic errors were calculated relative to the total number of errors produced by each individual participant.

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean score (SD) proportion accuracy, unless otherwise noted</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS Report</td>
<td>.85 (.21)</td>
</tr>
<tr>
<td>Spoken Naming</td>
<td>.57 (.32)</td>
</tr>
<tr>
<td>Error types during Spoken Naming</td>
<td></td>
</tr>
<tr>
<td>(proportion of total errors)</td>
<td></td>
</tr>
<tr>
<td>Phonological errors</td>
<td>.36 (.26)</td>
</tr>
<tr>
<td>Semantic errors</td>
<td>.15 (.17)</td>
</tr>
<tr>
<td>Rhyme judgment</td>
<td></td>
</tr>
<tr>
<td>Picture-based</td>
<td>.69 (.15)</td>
</tr>
<tr>
<td>Auditory</td>
<td>.84 (.14)</td>
</tr>
<tr>
<td>First letter identification</td>
<td></td>
</tr>
<tr>
<td>Picture-based</td>
<td>.70 (.29)</td>
</tr>
<tr>
<td>Auditory</td>
<td>.80 (.30)</td>
</tr>
<tr>
<td>Syllable count</td>
<td></td>
</tr>
<tr>
<td>Picture-based</td>
<td>.68 (.25)</td>
</tr>
<tr>
<td>Auditory</td>
<td>.80 (.23)</td>
</tr>
<tr>
<td>Pseudoword repetition</td>
<td>.49 (.31)</td>
</tr>
<tr>
<td>Pseudoword oral reading</td>
<td>.35 (.36)</td>
</tr>
<tr>
<td>Motor speech evaluation AMR rate</td>
<td>3.42 (2.45)</td>
</tr>
<tr>
<td>(syllables/second)</td>
<td></td>
</tr>
<tr>
<td>Adequate SMR production (Y:N)</td>
<td>25:28</td>
</tr>
<tr>
<td>(at least three syllable triads)</td>
<td></td>
</tr>
<tr>
<td>Segmental errors</td>
<td>.41 (.29)</td>
</tr>
<tr>
<td>Ambiguous/distorted consonants</td>
<td>.09 (.14)</td>
</tr>
</tbody>
</table>
Table 4.2. Correlational analysis comparing logit-transformed IS Report, Spoken Naming, and factor scores. Factor scores were generated in the context of the principal components analysis described above and presented in Figure 4.1. All values represent Pearson’s $R$. * denotes statistical significance at a Bonferroni-corrected $p<.05$.

<table>
<thead>
<tr>
<th></th>
<th>Spoken Naming</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IS Report (logit transform)</strong></td>
<td>.565*</td>
<td>.094</td>
<td>.550*</td>
</tr>
<tr>
<td><strong>Spoken Naming</strong></td>
<td>---</td>
<td>.602*</td>
<td>.394*</td>
</tr>
</tbody>
</table>
Table 4.3. Correlational analysis comparing IS Report and Spoken Naming to naming error types. All values represent Pearson’s $R$; * denotes statistical significance at a Bonferroni-corrected threshold of $p<.05$.

<table>
<thead>
<tr>
<th></th>
<th>Phonological Errors</th>
<th>Semantic Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IS Report (logit transform)</strong></td>
<td>.438*</td>
<td>.249</td>
</tr>
<tr>
<td><strong>Spoken Naming</strong></td>
<td>.357</td>
<td>.508*</td>
</tr>
</tbody>
</table>
CHAPTER 5

Discussion

In these studies, I have taken two main approaches to the study of the subjective experience of inner speech (IS) in aphasia, both of which involve directly asking people with aphasia about their experience of the phenomenon. In the pilot study (Chapter 2), I used an interview-based approach to probe the experience of successful IS in the context of anomia in general, day-to-day communication. In the second and third studies (Chapter 3 and Chapter 4), I used a more specific approach, asking participants to report the success of their IS on an item-by-item basis during a silent picture-naming task. In this chapter, I will summarize the findings in the context of our hypotheses, describe alternative interpretations, identify relevant limitations of the work, and present a set of open questions and possible future directions for this line of work.

5.1 Summary of main findings

5.1.1 Successful inner speech relates to lexical retrieval

The primary hypothesis tested in this dissertation is that the subjective experience of IS arises in conjunction with successful lexical retrieval, the final stage of which involves accessing the lexical phonological representation (Figure 1.2). This hypothesis has some support from prior work in my lab (Hayward, 2016; Hayward et al., 2016). It is also consistent with previous literature elucidating the mechanisms of inner speech in healthy individuals (Oppenheim & Dell, 2008) and demonstrating preserved ability of individuals with aphasia to generate phonological representations via silent picture-naming (Feinberg et al., 1986). Our pilot study (Chapter 2) was not designed to show a direct relationship between IS and word retrieval, since it focused on successful IS solely in the context of a subsequent naming failure, but the results did show that
this phenomenon is meaningful and dissociable from other experiences of anomia in which the process fails prior to successful word retrieval. Building on that result, the findings from our main studies (Chapter 3 and Chapter 4) show explicitly that self-reported successful IS relates to lexical phonological retrieval, as evidenced by relationships with performance on objective language measures, at the level of individual items as well as overall task accuracy. IS reports were so successful in capturing phonological retrieval ability that in a regression model predicting spoken naming accuracy, they supplanted the objective measures of retrieval and improved the overall model fit. In this sense, IS reports could be said to be a better measure of phonological retrieval than the objective measures used.

5.1.2 A failure of naming after successful inner speech relates to poor output processing

Our secondary hypothesis is that successful IS does not require articulatory processing and that a failure of naming after successful IS must occur due to impairment in the stages of post-lexical output processing (Figure 1.2). In Chapter 2, behavioral analyses revealed a tentative relationship between sIS during anomia and an objective measure of phonological output processing. Correspondingly, a multivariate lesion-symptom mapping analysis demonstrated that the experience of successful IS followed by anomia is associated with damage to brain regions supporting speech output. Findings from Chapter 3 and Chapter 4 show that articulation and other output processes relate to spoken naming but not to IS judgments, supporting the hypothesis that output processing is not required for a subjective experience of successful IS, although it is necessary for spoken naming. Additionally, as stated above, we found that self-reported IS was a better predictor of spoken naming than a set of objective tests of lexical retrieval, generating a model in which spoken naming comprises self-reported IS and
speech production. This finding contributes direct support to the hypothesis that a failure of naming after successful IS arises due to deficits in output processing.

5.2 Implications for self-monitoring of inner speech

In addition to interpreting our findings in the context of processing models of word-finding, we can also consider how our findings relate to speakers’ ability to monitor their IS. As described in Section 1.4, there are a few competing theories of self-monitoring, including comprehension-, production-, and conflict-based monitoring systems, most of which include an account of how speech can be monitored internally, prior to spoken output (Levelt et al., 1999; Nozari et al., 2011). We have demonstrated that people with aphasia are generally able to monitor the success of their IS despite difficulties with various language processes. Furthermore, we have shown that monitoring can be performed based on lexical phonology, independent of one’s ability to produce a word aloud. These findings are consistent with most theories of self-monitoring, with the exception of a theory in which motor-sensory modeling pathways are the primary substrate for internal self-monitoring (Tian & Poeppel, 2012, 2015). Our findings do not yet reveal the nature of the representations on which monitoring is based, nor do they indicate whether the basis of IS judgments can differ across tasks contexts, so future studies would be required to answer these questions. For example, an experiment using auditory masking or interference during the silent picture naming task could help to clarify whether participants are relying on auditory imagery or a more abstract, amodal form when judging the success of IS. Comparison of picture-based IS to other task contexts (e.g., silent reading) would be informative as to whether the basis of IS judgments differs based on task demands.
5.3 Alternative interpretations of the findings

5.3.1 The level of word-finding at which a feeling of successful inner speech arises

Although we have produced several pieces of converging evidence for a relationship between self-reported IS and lexical retrieval, our findings cannot definitely prove that self-reported IS reflects complete phonological retrieval in all cases. As described in Chapter 3, some of the objective language tasks relating to IS could be performed via partial phonological access only. Furthermore, the relationship between self-reported IS and the proportion of phonological errors during spoken naming (Chapter 4) is consistent with either partial or complete phonological retrieval as the basis for IS, as phonological errors can arise either from failure to retrieve the full phonology or from a failure to successfully complete output processing stages that follow complete phonological retrieval (Nickels & Howard, 1995b; Romani et al., 2002; M. F. Schwartz et al., 2004; Wilshire, 2002). Importantly, a feeling of IS may arise based on different degrees of success at the level of phonological retrieval across participants, or even within participants in different instances. We have found robust support for the conclusion that the subjective experience of IS related to lexical retrieval, although we cannot yet draw conclusions about the exact degree of successful retrieval that is required for a judgment of successful IS.

An alternative possibility is that a sense of IS arises from successful access to the lemma, a stage that precedes lexical phonological retrieval (Figure 1.2). The objective measures that relate to self-reported IS in our analyses rely on specific phonological knowledge and this level of knowledge is not associated with the lemma, which is an abstract word form (Dell & O’Seaghdha, 1992b; Dell et al., 1997; Goldrick & Rapp, 2007; Levelt et al., 1999). In many cases, however, intact access to the lemma will lead to a degree of phonological access (either
partially or completely correct), so participants’ demonstration of relatively intact phonological knowledge of words reported as successful IS does not necessarily rule out the possibility that the judgment of IS is performed on the basis of lemma access.

In the context of the current analyses, we cannot definitely distinguish whether IS judgments were based on access to the lemma vs. lexical phonology, but some aspects of the data make it unlikely that lemma access was the basis for a sense of successful IS. In Chapter 3, we found that self-reported IS related to word length, which should not be the case if IS is based on the abstract word-form. In that same chapter, we found that words reported as unsuccessful IS were named, on average, with 23% accuracy (compared to an average of 47% accuracy for successful IS words). Given the low variability in naming scores across testing days, it seems unlikely that individuals would retrieve the phonological form and accurately produce nearly one-quarter of items for which they failed to access the lemma on a different day. One caveat to this interpretation is that we do not have data regarding day-to-day variability at the individual item level in this study, so it is possible that there may be considerable item-level variation despite stability in overall naming accuracy across days.

There are several methods that could be implemented to explicitly examine the role of access to lemmas vs. lexical phonology in the experience of successful IS. First, one could utilize the computational model underlying the two-step interactive processing model of word retrieval (Dell, 1986; Dell & O’Searghda, 1992a; Dell et al., 1997). There, the strength of the connection between semantic representations and lemmas (s-weights) and between lemmas and lexical phonological representations (p-weights) can be modeled based on speakers’ error patterns. If IS reflects access to lemmas, it should relate only to s-weights, whereas if it reflects access to lexical phonology it should relate to p-weights (or perhaps to both s- and p-weights). In a
different approach, one could potentially make use of homophones, words that share a lexical phonological representation but have different lemmas (Biedermann, Blanken, & Nickels, 2002).

For example, a homophone judgment task could be added to the IS task battery, requiring participants to identify two homophones from a field of four (or more) pictures, with close phonological neighbors as the foils. If participants are able to select the homophones accurately (for words reported separately as successful IS), this would suggest that their IS is unlikely to be based on lemma access, since this abstract word form should not be adequate for judging the full phonological form. Future studies like those described above would help to clarify whether the success of IS can be judged based on access to the lemma, rather than the lexical phonological form.

5.3.2 Access to phonological vs. orthographic representations

A second question regarding the relationship between self-reported IS and phonological retrieval relates to orthographic representations. The lexical word-form, sometimes termed a *lexeme*, can be phonological or orthographic in nature, depending on the task (Caramazza, 1997; Goldrick & Rapp, 2007). In describing anecdotal evidence regarding the subjective experiences of anomia (Section 1.7.2.2.1), we noted that some individuals with aphasia make comments such as, “I can see the word in my head,” suggesting that they are generating mental imagery of the orthography of the word, either partially or in its entirety. In defining IS to the participants prior to administration of the IS report task (silent picture naming), we were careful to describe the experience as a feeling of having “all the right sounds, in the right order,” but it remains possible that some participants could have judged the accuracy of their IS based on retrieval of the letters, rather than the sounds, comprising the word-forms corresponding to each picture.
Although this was not our primary research question, some aspects of the experimental design specifically allow us to comment on the likelihood of IS judgments arising entirely from orthographic retrieval. On the picture-based rhyme judgment task, we selected word pairs such that the orthography of the word endings (rimes) was not adequate for judging whether the pair rhymed or not, with roughly half of the pairs being incongruent in terms of rhyming and orthography (e.g., rhyming pairs with differently spelled word endings or non-rhyming pairs with identically spelled word endings; see Section 4.2.2.4 for more details). In a preliminary analysis comparing performance on congruent vs. incongruent stimulus pairs (not reported in this dissertation), it appears possible that some participants may have been relying more heavily on orthography on the picture-based rhyme judgment task as compared to the matched auditory rhyme judgment task. More precise future analyses will examine individual differences in how strongly participants rely on orthography and will also compare this phenomenon in patient vs. control participants, likely by using reaction times to assess variability in control performance.

In our primary word stimuli (60 items, see Table 3.1), we did not include a large enough number of items with irregular spelling-sound correspondence to assess contributions of orthographic knowledge to the first letter identification or syllable counting tasks. It is possible that orthographic knowledge supported performance on these tasks. This would not necessarily be inconsistent with the conclusion that a subjective experience of IS most often arises in conjunction with lexical phonological retrieval, but would suggest that it is also possible for some individuals to report successful IS given successful retrieval of the orthographic word-form. An additional study using stimuli designed to vary systematically in phonology and orthography would contribute to our knowledge of relative contributions of phonological vs. orthographic retrieval to IS judgments across individuals and across tasks.
5.4 Practical considerations regarding the use of subjective reports in individuals with aphasia

5.4.1 The relative benefit of objective vs. subjective measures of inner speech

In the Introduction (Section 1.7.1.2), I described two different approaches to the study of IS in healthy language users: one that is focused on the subjective experience, utilizing measures such as structured questionnaires to elicit information about IS (Hurlburt et al., 2016; Morin et al., 2011), and another in which the properties of IS are measured through objective measures such as rhyme or homophone judgments (Gustafson, Bess, & Lancaster, 2017; Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Løeivenbruck, 2014). In studies of IS in aphasia, the latter approach has been more common, with several studies utilizing objective measures as a proxy for IS, and in fact defining IS as the ability to perform such tasks (Feinberg et al., 1986; Geva, Bennett, et al., 2011; Stark et al., 2017). Such approaches have the benefit of maintaining a level of objectivity that reduces the impacts of individual variability in self-monitoring and overall reliability, and also helps to avoid the effects of potential biases in subjective IS reporting (see next section for more on this).

By relying on a definition of IS that is objectively measureable in nature (Levine et al., 1982), however, these previous studies of IS in aphasia omit a potential source of information: the insight of the individuals who are experiencing IS. In this dissertation, I have utilized methods that include both subjective and objective approaches to studying IS, in an attempt to capitalize on the subjective experience for the purpose of better understanding the more objective, mechanistic nature of IS. There is a precedent for this in studies of inner speech in healthy language users, such as a study in which participants’ self-report of errors during silent recitation of tongue-twisters are used to examine patterns of errors in IS vs. overt production (Oppenheim & Dell, 2008). Prior work in my lab has shown the feasibility of this combined
approach in aphasia (Hayward, 2016; Hayward et al., 2016). In our work, we define IS as the experience of being able to say a word in one’s head with all the right sounds in the right order, which allows us to address research questions that are fundamentally different from most previous studies on IS in individuals with aphasia. The regression results obtained in Chapter 4 suggest that participants’ insights about their IS may be more informative about their phonological retrieval ability than common objective tests of this ability. Therefore, although our self-report measures are subject to some biases, they afford us direct insight into the experience of IS and allow us to illustrate the ways in which a common experience among individuals with aphasia is meaningful in the context of processing models of naming and word-finding.

5.4.2 Potential biases in IS reporting

Although our findings clearly demonstrate that the subjective reports of participants were on the whole reliable and meaningful, thorough consideration of concerns regarding a subjective measure of IS must also be considered. In addition to the mechanistic alternative interpretations presented above, there are also subjective biases that are pertinent to interpretation of self-reported IS. In Chapter 3 and Chapter 4, participants reported the success of IS during a silent picture-naming task. As described previously (5.3.1), some participants may have been reporting successful IS based on partial rather than complete lexical retrieval, which may reflect some level of tolerance for slight errors in word retrieval. A certain amount of leniency in error judgment has been previously observed in healthy language users, whose perception systems may not accurately identify errors in inner speech when similar phonemes are involved (Oppenheim & Dell, 2008). We did not have participants make explicit judgments of the accuracy of their spoken naming, but an additional experiment that includes such a measure would provide information about individuals’ willingness to allow minor production errors in the
context of accuracy judgments. Any tolerance exhibited during spoken naming may extend to judgments made on successful vs. unsuccessful IS as well, as was shown in previous work from our lab in a small number of participants (Hayward, 2016). A future analysis could also examine spontaneous error detection in relation to the phonological proximity of naming attempts to the targets, which may help to elucidate individual tolerance for slight errors in phonological retrieval and/or production.

More generally, there may be an effect of self-confidence or overall insight in reporting the accuracy of IS in general (Chapter 2) or at the item level (Chapter 3 and 4). Although reduced insight into deficits is more typical in right-hemisphere stroke as opposed to left-hemisphere stroke, it is widely acknowledged that some individuals with aphasia demonstrate poor error awareness (Helm-Estabrooks & Albert, 2004; R. C. Marshall et al., 1994). There is also evidence for general variability in confidence around communication ability among individuals with aphasia and a rating scale was recently developed to help clinicians and researchers gain quantitative information about communicative confidence (the Communication Confidence Rating Scale for Aphasia, CCRSA) (Babbitt, Heinemann, Semik, & Cherney, 2011). In future studies, it would be interesting to include the CCRSA and/or a measure of general awareness of language difficulties such as the VATA-L (Visual Analogue Test for Anosognosia for Language impairment) (Cocchini et al., 2010). The results of such assessments could contribute to our understanding of potential impacts of overall communicative confidence, error awareness, and insight on the subjective experience of IS.

A more specific concern about biases in IS reporting is that some participants may have been reporting the success of IS based on their expectations regarding their ability to name the picture aloud, which is a more common assessment task used with individuals with aphasia (vs.
silent picture-naming, which is relatively uncommon). It is plausible that, in a sequential task where IS is followed by spoken naming or another objective task, the subjective judgment of IS accuracy could be influenced by perceptions around ability to perform the subsequent task. When initially designing these experiments, we aimed to avoid this possibility by obtaining IS reports in isolation, rather than using a concatenated task where an objective task (e.g., spoken naming) immediately followed IS report. It seems unlikely for IS judgments to reflect expected performance on other tasks in this particular task battery design, although we cannot definitely rule out this possibility.

In general, we acknowledge that there may have been some under- or over-reporting of IS in some individual participants. Despite this possibility, we found very strong, predictable relationships between IS and objective language measure, both across individuals (Chapters 2 and 4) within individuals (Chapter 3). These associations make it clear that self-reported IS is, overall, a reliable reflection of the mental processes required for naming.

5.4.3 Potential impacts of impaired self-monitoring

Since our work is based on subjective reports that rely on self-monitoring of inner speech, we must carefully consider the potential impacts of impaired self-monitoring on our findings. If self-monitoring were impaired in a large proportion of our participants, self-reports would not emerge as a reliable measure of IS. As previously described (Section 1.8.2), we excluded participants with severe comprehension impairments at the level of single words (Chapter 2) and sentences (all studies), which may help to alleviate concerns about poor self-monitoring in the participant groups. The word-level comprehension criterion included in Chapter 2 stemmed from our assumption at that time of a comprehension-based monitor, wherein individuals with poor auditory processing of single words should also exhibit poor self-
monitoring, which relies on auditory comprehension pathways in such models. For the main studies (Chapters 3 and 4) we did not assume a comprehension-based monitor; we looked for a relationship between overall IS judgments (as measured in Chapter 3) and measures of comprehension ability, but did not find any notable relationships. In all chapters, we excluded individuals with poor sentence-level comprehension, primarily to help ensure that participants would be able to understand the IS-related interview questions and all task instructions. This criterion excluded any participants with severe global or Wernicke’s aphasia, who are known to have poorer error monitoring of spoken output (and likely inner speech) (R. C. Marshall et al., 1994).

These comprehension-based exclusion criteria do not eliminate all concern of poor reliability at the level of individual participants, however, and our Chapter 3 analyses reveal that a small number of participants did in fact show some evidence of unreliable reporting. For those individuals, it is possible that they were judging their IS based on expectation of performance on other tasks, based on retrieval of very limited information about the target word, or completely erroneously (as described in the previous section). Compared to other study participants, those participants showed lower accuracy on word-level auditory comprehension, as well as less frequent spontaneous error detection/correction on the spoken naming task, both of which may suggest general impairments of self-monitoring. In future studies or clinical practices that consider self-reported IS as a useful source of information about word retrieval, these possibilities should be taken into account; further work is needed to understand more about what types of individuals with aphasia are most likely to be unreliable in this way.
5.4.4 Individual differences in the overall reliability of self-reported inner speech

By finding meaningful, predictable relationships between self-reported IS and objective language measures at the group level, we have provided evidence for at least some level of reliability in IS judgments at the level of daily communication (Chapter 2) as well as individual items on a silent picture naming task (Chapters 3 and 4). As described above (Section 5.3.1), we are unable to draw any definitive conclusions regarding whether self-reported IS reflects fully intact lexical retrieval but can conclude that in most participants, IS reflects at least partial phonological retrieval. We are unable to rule out the possibility of some participants having over- or under-reported the success of their IS, and in fact did find some evidence for a small portion of our participants having been unreliable reporters to some extent.

It would be helpful to identify patterns that help to clarify what types of individuals with aphasia are more or less reliable with respect to IS reporting. In Chapter 3, we examined the reliability of IS reporting, measured by (1) day-to-day reliability and (2) the difference in performance for successful vs. unsuccessful IS items across the objective language measures using those same items. As described in the previous section, we found that participants who showed some evidence of unreliable IS reporting had poorer performance on a word-level auditory comprehension task and less frequent spontaneous error detection/correction during spoken naming, compared to other participants. These findings suggest that IS reliability is likely related to overall self-monitoring ability, but there are other ways to identify predictors of the reliability of self-reported IS as well. These could include measuring potential bias in leniency of error monitoring for spoken output (as described in Section 5.4.2) or administering structured assessments of self-awareness regarding language ability (e.g., the VATA-L or CCRSA) (Babbitt, Heinemann, Semik, Cherney, et al., 2011; Cocchini et al., 2010). Future research of this
type would help to predict generalizability of our findings by revealing more about the overall reliability of subjective IS judgments across other groups of individuals with aphasia.

5.5 Limitations and unsuccessful elements of the studies

5.5.1 Limitations of the pilot study

In the first study (Chapter 2), I presented a set of preliminary analyses examining the subjective experience of IS in aphasia. In the course of those analyses, several limitations were revealed (some of which are already mentioned in the discussion section of Chapter 2) (Fama et al., 2017). The frequency response scale (never, rarely, often, sometimes, almost always) was not precisely defined, so each participant may have interpreted the scale somewhat differently. Also, the objective language measures that were used alongside the questionnaire responses were chosen retrospectively from a language battery that was compiled in the context of a different primary research question. As such, we were not able to isolate specific levels of processing related to word-finding, which would have been essential in order for us to draw definitive conclusions about how the experience of sIS in anomia relates to processing models of naming and word-finding. Finally, we found an unexpected relationship between sIS during anomia and overall language severity, a finding that appears to undermine our hypotheses about relationships between sIS and specific processing levels. However, that study involved a group of participants who mostly exhibited either mild anomic aphasia or more severe Broca’s aphasia, and overall aphasia severity was positively correlated with apraxia as well as negatively correlated with fluency. Had our cohort included more participants with more severe, fluent aphasia and/or milder, non-fluent aphasia, we would not have found such a strong relationship between severity
and speech output deficits and would have been better able to disentangle relationships between sIS during anomia, specific processing deficits, and overall language impairments.

5.5.2 Unsuccessful elements of the main studies

In the experiments following this pilot work, we specifically addressed the concerns described above, but we encountered a few unsuccessful approaches in the context of our planned analyses for those studies (Chapters 3 and 4).

5.5.2.1 Revised inner speech interview yielded unreliable results

To address potential concerns regarding subjectivity in how the participants may have interpreted the questions and/or the response scale in the pilot study, I redesigned the interview for administration in the context of the main study. Here, I will describe the interview structure and administration procedures, as well as a few preliminary analyses conducted on data from a subset of participants. In general, no clear patterns were observed in these preliminary analyses, so we opted to focus the final analyses only on the item-level IS judgments as our measure of self-reported IS.

The interview began with a general introduction to anomia (the idea of wanting to convey an idea but not being able to say the word), with the first question being whether (and how often) the participant experiences anomia generally. Then, the idea of successful inner speech was introduced, and participants were asked whether (and how often) they’re able to say a word in their heads, despite being unable to say the word out loud; this experience of sIS was contrasted with a feeling of not being able to think of the right word. The frequency responses for both sets of questions (general anomia and sIS) were marked by the participant on paper, using an 11-point scale from “never” to “every time,” with the midpoint labeled “half of the time” (Figure 5.1). This structured scale was intended as an improvement to the subjective frequency rating
scale used in Chapter 2. Practice with the written frequency scale was provided prior to beginning the sIS interview, using a series of questions about the participant’s daily experience was driving a car and other forms of transportation (e.g., “when you go somewhere with your ______ (wife/husband), how often are you the driver?”). The two primary questions were:

1. Of all the times you want to say a word out loud, how often can you say the right word out loud? (frequency response was marked by the participant on written scale from “never” to “every time,” with responses converted to 0-10)

2. Of all the times you can’t say the right word out loud, how often can you say the right word in your head? (frequency response was marked by the participant on written scale from “never” to “every time,” with responses converted to 0-10)

All questions throughout the interview were read verbally to the participants and comprehension was supported through written words and pictures (see Figure 5.2 and Figure 5.3 for support pages corresponding to anomia generally and sIS vs. word-finding failure, respectively). Questions were repeated as needed.

The interview was administered twice, once at the beginning of each the two testing sessions (at least ten days apart). During preliminary analyses in 39 participants, we examined day-to-day (test-retest) reliability for the two questions listed above and observed a good deal of variability between day one and day two responses within participants. We also compared interview responses to other behavioral measures, including the level of successful IS reported during the silent picture-naming task and found no clear relationships.

There are several possible reasons for the absence of reliable effects in the preliminary analyses of the interview responses. The interview asks about the person’s experience with successful naming and IS in the context of daily communication and it is possible that the
experience of IS does vary day-to-day in individuals with aphasia, as does picture naming (D. B. Freed, Marshall, & Chuhlantseff, 1996; Howard et al., 1984), although our particular participant group showed relatively stable naming performance across days. It is also likely that these everyday experiences do not perfectly align with structured language tasks like confrontation naming, fluency, etc.; however, the pilot study revealed meaningful relationships between interview responses and objective measures of language ability. Another interpretation may be that despite our attempts to clarify questions through written and gestural support, our interview questions may simply have been too complex for some participants to understand. Given these various barriers, we decided to focus the main analyses (Chapters 3 and 4) on the item-level IS reports. In the future, I plan to revisit the interview data set in order to more closely examine individual differences in the meaningfulness of these interview responses (see Section 5.6 for additional discussion of future directions for obtaining IS reports via interview).

5.5.2.2 Lesion-symptom mapping revealed null effects

A second planned aspect of the main studies (Chapters 3 and 4) that did not turn out as expected was a set of lesion analyses using support vector regression lesion-symptom mapping (SVR-LSM) (Zhang et al., 2014), the same approach that was used in Chapter 2 to localize the questionnaire responses to certain lesion sites. The goal for the lesion analysis in the main study was to investigate relationships between self-reported IS on an item level (during the silent picture-naming task) and specific regions of stroke damage. We used this approach with a few different behavioral measures related to IS in a group of 45 subjects who underwent structural imaging in addition to completing the two-day language battery. First, we tested overall IS report on its own as well as with spoken naming accuracy as a covariate, seeking complementary evidence to the SVR-LSM findings presented in Chapter 2 showing that damage in sensorimotor
regions is associated with more frequent experience of sIS during anomia. In addition, we used a summary score of the three picture-based tasks (rhyme judgment, first letter identification, and syllable counting) and subtraction scores of each of these tasks paired with its auditory version (picture-based – auditory, tested both individually and as a summary score across the three tasks). Finally, we tested the frequency of IS as reported during the IS interview (Section 5.5.2.1).

None of these LSM analyses produced significant results, once voxel-wise and cluster-level family-wise p-value corrections and an appropriate lesion-volume correction method were applied. In Chapter 2, our SVR-LSM findings were reported as preliminary in nature, so we did not correct for multiple corrections, nor did we apply the most appropriate lesion-volume correction method. Had we applied the same stringent corrections as were used in the analyses performed on the imaging data from the main study, those preliminary findings (Chapter 2) would not have achieved significance. Importantly, we performed LSM analyses on some objective behavioral measures from the main study that are not essential to the research questions of this dissertation and did find significant results in expected brain areas when using the appropriate correction methods. It may be the case that our subjective measures of IS are not robust enough to produce strong LSM results; however, both sets of analyses involve relatively low numbers of participants and it is possible that data from a larger participant group would add the statistical power needed to reach significance. Our LSM findings from Chapter 2 do suggest that there may be meaningful relationships between stroke location and the subjective experience of IS in individuals with aphasia; however, it is clear that additional studies will be needed to clarify the brain basis of this phenomenon.
5.6 Open questions and future directions

Throughout this Discussion chapter, I have mentioned possible future directions for this line of work in the context of the specific topics discussed so far. Here, I will describe some open questions regarding the wider applicability of these findings and identify future directions for this work, both in terms of additional analyses of these data as well as potential studies that would serve as important next steps in this line of research.

5.6.1 Continued analysis of the existing data

In the preceding chapters, I have presented results that address the primary research questions and hypotheses of this dissertation. In the main study (described in Chapters 3 and 4), I have obtained a rich data set that offers opportunities for analyses beyond these main approaches. One immediate next step will be to take a closer look at the full set of responses from the interview that was administered as part of the main study (Section 5.5.2.1). There may be additional analyses that would help to reveal important predictors of day-to-day reliability and the ways in which the interview-based IS reports may be meaningful with respect to objective language ability. Further analysis of the interview responses would build on our Chapter 2 findings and further elucidate whether IS reports obtained in this manner can help to reveal the underlying mechanisms of the experience of IS. Such findings would be important because they would help to determine whether it may be feasible in clinical settings to use an interview or questionnaire-based approach to understanding the experience of IS in aphasia.

As described above (5.5.2.2), the multivariate lesion-symptom mapping (LSM) analyses performed on the data from the main study (Chapters 3 and 4) have not yet revealed significant relationships between lesion sites and experiences of IS. So far, we have focused only on LSM analyses related to the primary research questions, but there are additional questions to be asked
with this data set. For example, we used the same stimuli across many of the tasks in our battery, which offers the opportunity to ask nuanced questions about the brain regions underlying task performance. For instance, a future LSM analysis could compare spoken naming to repetition ability for the same target stimuli. We carefully designed our the primary task stimuli (see Table 3.1) to vary along several psycholinguistic variables: frequency, age of acquisition, word length, and articulatory complexity. To capitalize on this design, a future LSM analysis could compare the effects of variables related to word retrieval (e.g., frequency) vs. production (e.g., articulatory complexity) on naming, to determine areas of the language network that are sensitive to these different psycholinguistic features.

There are also additional analyses that could be performed on the spoken naming data that were obtained for the main study (Chapters 3 and 4). For error analyses performed in these studies, we used one criterion for phonological errors, which combined all real word and non-word errors that included at least 50% overlap in phonemes with the target word. In a future analysis, we could code phonological errors in a more precise manner, differentiating between errors that appear phonetic (e.g., an articulatory error such as ambiguous voicing of the initial sound in zipper) vs. phonemic (e.g., a phoneme substitution; “dipper” for zipper) in nature. This could be examined across the group of participants in the context of IS reports, which may help to shed light on whether participants’ errors result from deficits in phonological retrieval or post-retrieval output processing (Nickels & Howard, 1995b; Romani et al., 2002; M. F. Schwartz et al., 2004; Wilshire, 2002), which could further clarify the subjective experience of successful IS in the context of processing models of naming.

Finally, all analyses using the naming scores or error types were performed on the initial naming attempts, but we also coded final attempts so could look at differences between these two
stages of naming response, both generally and in relation to IS reports. While coding error types for naming responses, we also coded error detection and correction (see Section 3.2.3.3). In our analysis of individual differences in reliability, we identified a relationship between evidence for unreliable IS reporting and rates of spontaneous error detection and correction, suggesting preliminarily that self-monitoring mechanisms may be parallel across inner and overt speech. A more detailed analysis of these data in the future could shed additional light on these participants’ overt self-monitoring ability and its relationship to their experience of IS. For instance, we could look specifically at the likelihood of error detection/correction for items reported as successful vs. unsuccessful IS on the IS report task.

5.6.2 Ideas for future studies on the clinical applications of inner speech in aphasia

The possibility of clinical benefit from this line of research leads naturally to a set of potential treatment studies. In general, knowing how often a person experiences sIS during anomia might be informative as to the main cause of their anomia, which could in turn help with the selection of a treatment approach. For instance, someone who reports high levels of successful IS during anomia is likely to have relatively preserved word retrieval and is likely to benefit from a different therapeutic approach than someone who has difficulty at earlier levels of abstract word-forms or even semantic representations, or in the connections between these levels (see Figure 1.2). Future research could test overall levels of IS through item-level reports or interview-based approaches and compare these reports to response to various treatment approaches designed to primarily target either word-finding or word production.

Beyond the general diagnostic information that could be gained from knowledge of self-reported IS, there is an open question as to whether individual words that are reported as successful vs. unsuccessful IS respond differently to naming treatment. A study from our lab has
shown previously in two individuals with aphasia that words reported as successful IS were learned more effectively (and more quickly) than words reported as unsuccessful IS, during a subsequent period of paired-associate anomia treatment (Hayward et al., 2016); a future study replicating these findings in a large participant group would be beneficial. Furthermore, a prospectively-designed treatment study could compare treatment approaches to determine what types of anomia treatments are best suited for treating words that are reported successful vs. unsuccessful IS (with both sets being unsuccessful in spoken naming at baseline). I would predict that successful IS words would be best learned through treatments focused on output (e.g., treatments that are typically utilized to treat acquired apraxia of speech) (Farias, Davis, & Wilson, 2014; J L Wambaugh, West, & Doyle, 1998), whereas words that are unsuccessful in IS may require treatment approaches focused on earlier stages of word-finding (e.g., semantic feature analysis) (Boyle & Coelho, 1995).

Studies such as the ones described above may help clinicians make decisions about which words to treat during therapy and/or which treatment approaches to use to improve the production of those words; however, an important consideration in any treatment study is the likelihood for generalization. If future studies are able to show that self-reported IS can effectively inform clinical treatment planning, an essential next step would be to understand whether there is any level of generalization that can be achieved. One possible approach to producing treatment effects beyond the individual items utilized during therapy would be to investigate whether IS can serve as a mechanism for self-cueing in individuals with output deficits. At least one prior case study has shown that it is possible to train someone with aphasia to self-generate phonological cues, with improvements seen on trained words alongside some suggestion of a generalization effect (DeDe et al., 2003). While that study focused on written and
tactile cues, a future study utilizing IS as a mechanism for self-cueing could involve instructions about mental imagery related to IS, focused either on auditory imagery (imagine “hearing” the word in your head), motor imagery (imagine “saying” the word in your head), or both.

5.6.3 A broader set of open questions regarding the nature of inner speech

5.6.3.1 Applicability of these findings beyond the context of naming concrete nouns

These studies have focused on the subjective experience of IS primarily in the context of naming. The pilot questionnaire (Chapter 2) was designed to elicit experiences in general, daily communication. It is plausible that those findings may be applicable to a range of word-finding scenarios beyond naming, such as during spontaneous speech, picture description, or in natural conversational settings. In Chapter 3 and Chapter 4, our methods were focused on the subjective experience of IS specifically in the context of picture naming, so there is an important open question about the relevance of our findings beyond this limited context.

First, we might consider whether these findings would extend beyond naming of concrete nouns to single-word production of other word categories, such as abstract nouns, verbs, or adjectives. The mental process of word-finding should be similar across these categories, so the basic premise and hypotheses of our work would be the same for these other word types: that a subjective feeling of being able to say a word in one’s head should be associated with evidence of phonological retrieval for that word, concrete noun or otherwise. One interesting open question lies in the case of morphological aspects of words, such as plural markers on nouns or tense markers on verbs. Although not all processing models of word-finding specify the level of morphological processing, morphology has been positioned within stages of lexical retrieval, prior to assembly of the complete lexical phonological form (Levelt et al., 1999). The potential biases described above (Section 5.4.2) could also apply here, however, as it is possible that some
individuals might apply some level of leniency in reporting successful IS despite incomplete morphological processing. This is an empirical question that cannot be answered given our focus on concrete object naming, so additional studies would be required in order to understand how morphology affects the subjective experience of successful IS.

Many of the processing models on which we base our hypotheses and methods (see Section 1.3) focus on naming and word-finding at the level of single words. It is not obvious, therefore, how the subjective criteria for reporting successful IS of single words relates to the mechanisms by which someone would report successful IS for words embedded within sentences, or for evaluating the success of IS at level of whole phrases or sentences. The addition of a syntactic level of processing (in addition to morphological processing, as described in the previous paragraph) may affect the way in which IS is internally generated and monitored, and to date there are no studies examining IS at this level in aphasia. Many individuals with aphasia who are able to produce speech beyond the single word level are likely to experience IS that is more fluent as well, so studies in this area are needed to more fully understand the phenomenon of IS in aphasia. For example, one could examine whether there are individuals who report successful IS for phrases or sentences that they cannot produce successfully aloud and determine what levels of processing are responsible for the failure of spoken output, which could in turn help inform models of the mental processing stages required for the production of fluent, grammatical speech.

Prior literature suggests that the nature of IS in healthy language users depends on the task context (Oppenheim & Dell, 2010; Sokolov, 1972), so it would be interesting to determine whether this flexibility in IS applies to individuals with aphasia as well. In the studies presented here, we have found that articulatory processing is not essential to a judgment of successful IS in
the context of naming, but it is unclear whether these findings would differ in other task contexts. For example, we could design alternative versions of our tasks using silent reading (instead of picture-based generation of IS) to determine whether articulatory complexity might be relevant to task performance in that particular context. We could also design a future study in individuals with aphasia to replicate a previous experiment investigating the types of errors that healthy speakers produce during silent vs. overt recitation of tongue twisters (Oppenheim & Dell, 2008), to shed additional light on the role of post-lexical output processes on IS in various task contexts.

5.6.3.2 Additional questions regarding self-monitoring of IS

As previously described (Section 5.2), the results of these studies suggest that IS can be successfully monitored by individuals with aphasia on the basis of lexical phonology. There are several open questions regarding how IS judgments reflect the nature of self-monitoring more generally. In Chapter 3, we found that a small number of participants with some evidence for unreliability in their IS judgments showed less frequent error detection and correction during spoken naming. In our task design, we did not provide explicit instructions regarding self-monitoring of spoken naming responses, but instead relied on a measure of spontaneous error detection and correction. An individual may recognize an error without spontaneously exhibiting concrete evidence of this error awareness, so while this approach provides some insight into self-monitoring, it is limited in terms of the conclusions that can be drawn. In a future study, participants could be asked to provide explicit judgments of their spoken naming responses, to assess whether such expectations affect self-monitoring behavior and, in turn, how this monitoring behavior in a more explicit task paradigm relates to the reliability of IS judgments.

A second open question relates to the relationship between self-monitoring and domain-general cognitive processing. A recent model proposes that self-monitoring relies on domain-
general executive functioning ability (Nozari et al., 2011). Prior studies suggest that general executive function skills can affect language processing in aphasia (Zinn, Bosworth, Hoenig, & Swartzwelder, 2007), so future studies comparing these domain-cognitive skills to the reliability of IS monitoring would clarify whether this model of self-monitoring is viable with respect to IS. Such studies would also generally align with a current trend in aphasia research, which is to consider the role of domain-general cognitive processing in the understanding of deficits and spared abilities of individuals with aphasia.

5.6.3.3 The ability of people with aphasia to use inner speech to accomplish a broader set of mental tasks

An additional topic of interest related to the experience of IS beyond those with aphasia is the notion that IS is often used in the service of other cognitive functions. In the context of relatively abstract definitions of IS, where it represents verbal thinking or general self-directed inner language, IS has a role in many cognitive tasks including, but not limited to, working memory, language learning, reading, and executive functioning abilities such as planning, decision-making and task-switching (Alderson-Day & Fernyhough, 2015; A. D. Baddeley & Hitch, 1974; Hurlburt et al., 2016; Morin et al., 2011; Perrone-Bertolotti et al., 2014; Vygotsky, 1962). Our examination of IS in aphasia has focused on a more specific definition, being able to say a word in one’s head, for the purpose of understanding how the subjective experience aligns with processing models of naming. A new direction for this line of research, however, would be to examine the more abstract forms of IS in individuals with aphasia. Such approaches would require a different operational definition of IS, in order to provide individuals with aphasia to report about the experience more generally; recent questionnaires about IS developed for healthy language users may be an appropriate starting point (Hurlburt et al., 2016; Morin et al., 2011). If these methods could be adapted for use in individuals with aphasia, two areas of research
questions could be addressed: (1) whether individuals with aphasia are able to use IS to perform everyday, cognitive tasks which benefit from the ability to think verbally, and more generally (2) to what extent individuals with aphasia retain the experience of having “a little voice in one’s head,” a phenomenon that is well-known to all language users and seemingly integral to the conscious, human experience.

5.7 Conclusions

In this dissertation, I have provided robust evidence that individuals with aphasia are able to make meaningful judgments regarding their subjective experience of IS, defined here as a feeling of being able to say the right word in one’s head. Taken together, the findings show that the general experience of successful IS is most common in individuals with relatively spared lexical knowledge. Moreover, successful IS followed by overt anomia is most commonly experienced by individuals with damage to sensorimotor brain regions that support spoken output. The findings from the three studies thus support a model in which the subjective experience of successful IS arises together with lexical retrieval. Post-lexical output processing is not required for a person with aphasia to experience successful IS, which strongly suggests that pre-articulatory motor speech processes are not an essential component of IS. Given these findings, I propose that self-reported successful IS should be considered as an additional source of information when diagnosing a person’s anomia in the context of processing models of naming. Additional investigations as described above will further clarify the mechanisms underlying the experience, which will help to inform clinical decision-making around naming treatment and may also contribute to a greater understanding of the experience of IS in all language users.
Figure 5.1. Response scales used during the interview. Each of the two scales seen above was presented in isolation, as described in Section 5.3.2.1.
You have an idea in your head

OR

“cat”

You can say the right word out loud

✓

You can’t say the right word out loud

✗

Figure 5.2. Interview picture support: page one (anomia vs. successful naming).
Figure 5.3. Interview picture support: page two (sIS vs. word retrieval failure).


Evidence. *Journal of Medical Speech-Language Pathology, 14*(2), xvi–xxxiii.


http://doi.org/10.1080/02643294.2016.1192998


new challenge for computational neuropsychology. Cortex, 64, 417–419.


http://doi.org/10.1371/journal.pone.0147932


Disorders, 54, 113.


http://doi.org/10.1080/02687030601065470


http://doi.org/10.1093/brain/awu164


Nickels, L. A., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllable complexity on word production in aphasia: It’s the number of


inner speech. Memory & Cognition, 38(8), 1147–1160.
http://doi.org/10.3758/MC.38.8.1147

http://doi.org/10.1016/S0093-934X(03)00418-8

http://doi.org/10.1093/cercor/bhn239


