Talking to Myself: How the Phonological Network Supports Inner Speech During Computer Code Comprehension

Malayka Mottarella

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Chantel Prat, Chair

Scott Murray

Ariel Starr

Andrea Stocco

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Talking to Myself: How the Phonological Network Supports Inner Speech During Computer Code Comprehension

Malayka Mottarella

Chair of the Supervisory Committee:

Chantel Prat

Department of Psychology

Prior work has highlighted a connection between natural language skills and programming ability. In this dissertation, I investigate the role of a specific language-related process—phonological coding—and its potential role in aiding computer code comprehension. Using a neuroscientific individual differences approach I examine (1) whether the phonological system is actively engaged during Python code comprehension, (2) the mechanisms that might explain this relationship, and (3) how individual differences in behavioral factors indexing skill, capacity, and strategy modulate the involvement of the phonological system during code comprehension. Specifically, I investigate whether phonology's role in code comprehension is merely an

epiphenomenon of accessing English word meanings, or if it serves a functional role in supporting comprehension. My results suggest that the phonological system is involved in code comprehension, and that this involvement is modulated by individual differences in cognitive capacity and strategy. Together, this work suggests that phonological codes can be a functional building block for constructing an internal problem representation during computer programming tasks.

Acknowledgements

My doctoral journey has been shaped and supported by an incredible community of mentors, peers, friends, and family – thank you all for your unwavering support.

This work would not have been possible without the wisdom and guidance of my supervisory committee – thank you to Chantel Prat, Scott Murray, Andrea Stocco, Ariel Starr, and Naja Ferjan-Ramirez for your thoughtful questions and great suggestions as this project took shape. I would also like to thank the University of Washington Center for Human Neuroscience and its fantastic staff for helping to collect the neuroimaging data for the project and my undergraduate research assistants for the many hours you put into to helping collect the behavioral data for this project. This work was funded by an award from the Office of Naval Research (GR00970) to Chantel Prat. I would also like to thank the University of Washington Auditory Neuroscience Training Grant for funding me for the last two years so I could focus my attention on this project.

My doctoral journey has been filled with twists, turns, and bumps along the road. I would like to thank my partner in life Sam Coren for his endless love and support throughout this process. Thank you for listening to me talk about all the small details of this project, buying me ice cream when it all got to be a bit too much, and always reminding me what I was capable of. I would also like to thank my family, in particular my grandfather Vic Mottarella Sr, who always stressed the value of education and would have been so proud to see me reach this goal, and my parents Jean Potts and Victor Mottarella, who always encouraged me to follow my passions and trust that it would all work out. I am proud to share this accomplishment with you.

Table of Contents

Talking to Myself: How the Phonological Network Supports Inner Speech During Computer Code Comprehension

Computer programming skills are valuable in modern society, with many high-paying occupations requiring or preferring that applicants know how to program (e.g., Li, 2022; Sun et al., 2021). In 2022, the U.S. Bureau of Labor Statistics reported over 2.6 million computer-science-related jobs with a median wage nearly a third higher than the national average (U.S. Bureau of Labor Statistics, 2022). Despite the demand for programming skills and the monetary benefits associated with STEM careers, much is still unknown about the factors that drive successful programming. However, it is well established that learning to program is notoriously difficult (see Jenkins, 2002 for a discussion), with attrition rates from introductory programming courses estimated to approach 50% worldwide (Bennedsen & Caspersen, 2019). Both the written generation of new code and the comprehension and debugging of preexisting code require a programmer to access semantic memory stores (e.g., keywords, functions, etc.), apply rigid syntactic rules, and bind together smaller chunks of information hierarchically to understand the overarching goal of the program. While each of these processes alone can be difficult, programmers must often deploy them in tandem. The interwoven nature of programming skills has been cited as a key reason underlying the high dropout rates from programming courses - such that failure to understand one concept poses challenges for acquiring and mastering other related concepts (Robins, 2010).

Understanding the neurocognitive factors that support skilled computer programming is critical for developing a rich understanding of when, and for whom, programming is difficult. While empirical work dedicated to understanding the factors that drive programming skill acquisition and its variance has been on the rise (e.g., Floyd et al., 2017; Ivanova et al., 2020; Kuo et al., 2022; Prat et al., 2020), much is still unknown about the neurocognitive basis of

programming. In comparison, a rich body of literature has been devoted to understanding the factors that drive the successful acquisition of a second *natural* language (for reviews see Chalmers et al., 2021; Turker et al., 2021; Wen, Biedron, & Skehan, 2017). A growing body of work has suggested that second language aptitude may function as a useful framework from which to scaffold an initial understanding of programming aptitude (e.g., Hishikawa et al., 2023; Kuo & Prat, 2023; Prat et al., 2020). Several studies from our lab have adopted this framework and found that second language aptitude predicted both the speed and accuracy with which individuals learned to program (Kuo et al., 2022; Prat et al., 2020). Second language aptitude is an umbrella term used to categorize several language-related subskills that assess one's potential for acquiring a new natural language in the future (Carroll, 1981). My dissertation explores the role that one of these second language aptitude subskills, phonological coding, plays in programming using a neurocognitive individual differences approach.

In the sections that follow, I will summarize what is known about the neurocognitive factors that support programming skills and argue that phonological processing may be an overlooked factor contributing to programming success. First, I will provide a brief review of the neurocognitive factors that have been previously related to programming skills. Second, I will explore the idea that natural language, generally, and explicitly learned second language, in particular, can serve as a useful framework for understanding computer programming. Third, I will review how the construct of second language aptitude and its subcomponents has been used to study both natural and programming languages. Fourth, I will explore the idea that phonological coding, a subcomponent of second language aptitude, may contribute to programming, and I will introduce several hypotheses of *how* the phonological system might operate within the context of code comprehension. Lastly, I will explain how a neuroscience

approach using an individual differences framework can help to delineate between the hypothesized roles of the phonological system in code comprehension.

Advances in Developing a Cognitive Understanding of Computer Programming

Prior work on the cognitive basis of computer programming has investigated a wide range of general cognitive, mathematical, and language-related skills as potential drivers of individual differences in programming skills. Understanding the relative weight that these cognitive abilities play in supporting programming is important for both informing theories of computer programming aptitude and for developing a model of the underlying cognitive components that support comprehending code in real-time. In the sections that follow I will provide a non-exhaustive review of some of the advances made towards this ambitious goal.

Higher working memory capacity, or the ability to temporarily maintain and manipulate more information in mind, has been shown to be a robust predictor of programming skills across a number of studies (e.g., Pena & Tirre, 1992; Prat et al., 2020; Shute & Kyllonen, 1990; Shute, 1991). Cognitive control has also been implicated in programming, such that individuals who employed better cognitive control strategies performed better on programming tasks (Wang et al., 2022), and were less likely to drop-out of their introductory programming courses (Margulieux et al., 2020). Prior work has suggested that working memory is strongly linked to cognitive control, such that better cognitive control leads to more efficient usage of working memory resources (e.g., Engle & Kane, 2003; McNab & Klingberg, 2008; Vogel et al., 2005), and that cognitive control goals can be maintained and instantiated using working memory (e.g., Braver, 2012). Thus, the contributions of working memory and cognitive control to programming are likely interrelated.

Spatial reasoning has also been implicated to some degree in programming. In a four-year longitudinal study, spatial skills were assessed prior to learning to program and then correlated with programming performance after each year of learning (Parkinson & Cutts, 2022). Results indicated that spatial skills were not significantly related to programming skills after the first two years of learning, but that the relationship between spatial skills and programming abilities increased over time to become significant after years 3 and 4 of learning. One interpretation of this finding is that spatial skills become more important at advanced levels of programming. However, an alternate possibility is that the types of content used in more advanced programming courses is aided by spatial skills but not the act of programming per se. The findings of a recent neuroimaging study also failed to find a relationship between spatial skills and early programming success. In this study, novice programmers who shared more similarity in brain activity between programming and spatial rotation tasks had *lower* coding proficiency after 11 weeks of subsequent learning (Endres et al., 2021). This finding is consistent with the idea that spatial skills may not aid early programming success.

A number of studies have suggested that better mathematical skills can lead to benefits in programming (e.g., Bendersen & Caspersen, 2006; Quille & Bergin, 2018; Shute, 1991). However, in a study from our lab the predictive power of numeracy for explaining individual differences in programming acquisition was only 2% when other general cognitive and language factors competed to explain the same variance (Prat et al., 2020). This result is consistent with several other studies that found that while mathematical skills do predict programming, they do not do so at a magnitude stronger than other cognitive measures (e.g., Austin, 1987; Leeper & Silver, 1982). In contrast, a recent study using structural equation modeling pitted language and math skills against one another to compete for variance in programming ability and found that

the math, but not the language, factor predicted programming performance (Graafsma et al., 2023). However, one of the strongest loadings on the mathematical factor was a measure of logical reasoning.

Neuroimaging work has shown that the neural network involved in programming shares significantly greater overlap with the network involved in formal logic than it does with math (Liu et al., 2020). This result may suggest that some of the shared variance between math and programming is underpinned by logical reasoning or related general cognitive abilities to some degree. This interpretation may also shed light on the inconsistent results across the Prat et al. (2020) and Graafsma et al. (2023) studies surrounding the role of math skills in programming. Prat and colleagues (2020), included a large battery of general cognitive measures including working memory and fluid intelligence measures which competed with numeracy for variance in their predictive models of programming. These general cognitive measures predicted upward of 40% of the variance in learning to program. Considering that both fluid intelligence and working memory have been associated with individual differences in logical reasoning (e.g., Robinson & Unsworth, 2017), and that logical reasoning was included in the mathematical factor in the Graafsma et al. (2023) study, it is possible that some of the predictive power ascribed to the mathematical factor was not mathematically-specific but was related to other general cognitive factors like fluid intelligence and working memory.

Taken together, prior research suggests that general cognitive measures like fluid reasoning and working memory robustly predict programming success. More domain specific skills, like mathematical and spatial reasoning are frequently discussed as being important for programming. However, the empirical support for a unique role of these processes in programming has been mixed. In the section that follows I will discuss what is known about the

role of another frequently discussed domain-specific process – natural language – and its connection to programming.

Natural Language as a Framework for Understanding Programming Languages

The idea that programming languages share similarities to natural languages is not new. Rather, it has long been understood that the communication structure between humans and computers is aided by sharing similarities to the way humans communicate with one another. Grace Hooper, one of the early developers of COBOL, first pioneered this idea saying: "It is much easier for most people to write an English statement than it is to use symbols…data processors ought to be able to write their programs in English, and the computers would translate them into machine code" (Gilbert & Moore, 1981). This idea, once controversial, has now become so commonplace as to appear obvious. Modern programming languages have evolved to share increasing similarities with natural language. For example, Python – ranked the most popular programming language in 2024 (TIOBE Software, 2024) – is known for using indentation in a way that mimics the paragraph structure seen in natural language and for utilizing keywords and functions that are descriptive English words. The language-like composition of Python is frequently cited as one of the key reasons for Python's increased popularity over the last decade (e.g., Scarlett, 2023; Van Deusen, 2023).

In addition to the language-like features that are hardcoded into modern-day programming languages, programming education emphasizes natural language as a tool for clarity in the user-defined aspects of code. Modern-day best practices for writing code emphasize using meaningful and easily pronounceable variable names, writing informative comments, using indentations even when not explicitly required by the programming language, and leaving white space between sections to make the code more readable (e.g., Mckee, 2023; Sedgewick &

Wayne, 2016). The reliance on natural language components in computer programming may reflect an intuitive understanding that there are similarities between natural and programming languages. Empirical behavioral work supports this idea, with psychometric studies demonstrating that better language abilities predict higher programming scores (e.g., Kuo et al., 2022; Leeper & Silver, 1982; Sauter, 1986; Shute, 1991).

Theoretical work has gone a step beyond merely noting the shared features between programming and natural languages to argue that there may be computational similarities at the level of the underlying cognitive architecture. A recent paper from Fedorenko and colleagues (2019) outlines how this shared computational architecture might operate during the generation (e.g., writing) and comprehension (e.g., reading) of programming and natural language. During the comprehension of both natural and programming languages, readers progress through a similar cascade of embedded cognitive processes beginning with viewing a sequence of symbols, segmenting that input into smaller units, recognizing units of importance, considering the relationships between semantic and syntactic information, inferring complex dependency structures, and finally constructing a complex meaning representation (Fedorenko et al., 2019). A similar computational cascade is outlined for generation, such that writers of both natural language and computer code begin by identifying an overarching goal to convey, then breaking the goal down into smaller steps and operationalizing those steps within the semantic and syntactic constraints of the language being used, retrieving the individual semantic units from memory, and engaging in motor planning and execution (e.g., typing, writing, speaking).

The theorized computational similarities between programming and natural language have led to the related suggestion that neural areas supporting programming may overlap with the natural language network. This perspective is in line with evolutionary theories arguing that

‘newer’ cognitive processes (e.g., computer programming) will need to co-opt portions of the brain already utilized in related ‘older’ cognitive skills (e.g., natural language), and combine these regions into a network to support the novel skill (see Castelhano et al., 2021 and Liu et al., 2020 for discussions). Neuroimaging work investigating the extent to which the neurobiological systems supporting programming and natural language are shared has been mixed.

Some prior neuroimaging work has suggested that a left-lateralized network of regions commonly elicited during natural language tasks is also active during programming tasks. Several studies have found that portions of the left inferior frontal gyrus (IFG) - spanning Brodmann’s areas 44, 45, 47, and 6 – support code comprehension (e.g., Hishikawa et al., 2023; Ikutani et al., 2020; Liu et al. 2020; Peitek et al., 2020; Peitek et al., 2021; Siegmund et al., 2014; Siegmund et al., 2017; Xu et al., 2021). The left IFG has also been implicated as an important region for identifying bugs in preexisting code (Castelhano et al., 2019) and as a region that demonstrates plasticity and a shift towards right hemisphere lateralization with increased programming experience (Hishikawa et al., 2023).

In the natural language literature, the left IFG – commonly referred to as “Broca's area” – has been suggested to broadly serve the purpose of “binding” or “unifying” linguistic information (e.g., Hagoort, 2005). The subdivisions of IFG are thought to roughly map onto different aspects of linguistic processing in a functionally specific gradient such that semantic binding is associated with the most anterior-ventral portions of IFG in BA 47 and 45, syntactic binding is associated with BA 45 and 44, and phonological processing is associated with BA 44 and portions of BA 6 (e.g., Hagoort, 2005). Causal neurostimulation work further supports the theory that there is functional specificity within the subdivisions of left IFG (Klaus & Hartwigsen, 2019). Neuroimaging studies of computer programming have also implicated other

brain areas commonly activated in language tasks including semantic retrieval areas like the left middle temporal gyrus and angular gyrus, and subcortical linguistic control areas like the striatum (e.g., Endres et al., 2021; Hishikawa et al., 2023, Ikutani et al., 2021; Liu et al., 2020; Pietek et al., 2021; Siegmund et al., 2014; Siegmund et al., 2017; Xu et al., 2021). Together, these studies highlight that the neural network supporting programming seems to be at least partially overlapping with regions commonly ascribed to have linguistic functions.

Other neuroimaging studies have reached different conclusions, suggesting that the neural network supporting programming is more akin to other domain-general cognitive processes and does not overlap substantially with natural language function. In one study directly investigating this question participants completed both a language localizer (i.e., reading sentences > non-words) and multiple demand system localizer task (i.e., difficult > easy spatial working memory) and the contrast maps from these localizer tasks were compared against maps derived from a code comprehension task (i.e., calculating program outputs > sentence word problems) in the scanner (Ivanova et al., 2020). The results indicated greater activation for the programming problems in the multiple-demand system than in the language system, which was taken as evidence that the language system was not consistently recruited during programming. Several other studies have demonstrated that machine learning classifiers can accurately distinguish between tasks measuring natural language prose and computer programming generation (Endres et al., 2021; Krueger et al., 2020) and comprehension (Floyd et al., 2017) further supporting a distinction in the underlying neural networks.

An alternate explanation is that the extent to which the programming network overlaps with natural language-related areas may vary as a function of individual differences. For example, while Floyd and colleagues (2017) found that at the group-level, their machine-learning

classifier was highly accurate at distinguishing between neural representations obtained during code vs prose comprehension, with increased programming skill of the participant the classifier became significantly worse at distinguishing between the two tasks. This result suggests that as participants' programming skill increases, the neural network supporting code comprehension shares greater similarities with the neural network supporting natural language comprehension. Ikutani and colleagues (2021) also investigated how individual differences in programming skills modulated the neural network-supporting programming. In their study, a multivoxel pattern classification approach was used to decode different types of source code categories. The individual differences results demonstrated that greater programming skill was associated with better decoding accuracies in a network of regions including the bilateral inferior frontal gyrus, left inferior parietal lobule, and left middle temporal gyrus. This network of regions is predominantly composed of areas that are implicated in language-related processing (see Turker et al., 2023 for a review) and have been previously reported in neuroimaging studies of programming at the group-level (e.g., Catselhano et al., 2019; Hishikawa et al., 2023; Liu et al. 2020; Peitek et al., 2020; Peitek et al., 2021; Siegmund et al., 2014; Siegmund et al., 2017). This finding is also consistent with the idea that while programming and natural languages may look distinct in the brains of novice programmers, with increased skill these networks may converge to share more similarities. A similar phenomenon has been demonstrated with bilinguals reading in their first and second languages such that at higher levels of second-language proficiency there are fewer differences in the regions that support first- and second-language processing (Sebastian, Laird, & Kiran, 2011). These findings converge to support the viewpoint that second-language learning may be a more comparable framework to compare programming to than native language processing.

Programming languages are typically taught through explicit instruction in a classroom or in an online learning environment. This declarative instruction is quite different from learning one's first language which happens in an immersive way via procedural learning systems. A fairer comparison may be to say that learning to program shares similarities to instructed second-language learning (for further discussion see Kuo & Prat, 2024). Earlier work from our lab adopted this framework to see if neurocognitive factors known to be important for second natural language learning would also have predictive utility for explaining variation in learning to code in Python (Prat et al., 2020). In this study, we recruited participants with no prior programming experience to learn Python using an online learning software. Before learning, we administered a battery of language-related, general cognitive, and neurometric predictors derived based on a prior study of learning French (Prat et al., 2016; 2019). We entered these predictors into stepwise regression models individually for different programming outcomes of interest including learning rate, performance on a Python declarative knowledge test, and generation of a Python program. Our regression models indicated that together predictors of second natural language learning had strong predictive utility for explaining variation in programming with our best model predicting 72% of the variability in Python learning (Prat et al., 2020). Some of the predictors in these models – such as fluid intelligence and working memory – were well known to relate to learning generally and had been previously implicated in programming (e.g., Shute, 1991; Shute & Kyllonen, 1990). However, a particularly novel finding from this study was that a standardized measure of *second language aptitude* predicted learning to program in Python even after accounting for variance in general cognitive skills (Prat et al., 2020). In the section that follows I will review what is known about second language aptitude as a cognitive construct and how it has been implicated in learning programming languages.

Second Language Aptitude: What Is It, and How Has It Been Used to Study Programming Languages?

Second language aptitude is an umbrella term that refers to a set of cognitive abilities that predict individual differences in the successful acquisition of a second natural language (for reviews see Chalmers et al., 2021; Wen, Biedrón, & Skehan, 2017). The 'founding father' of language aptitude research, John Carroll, defined second language aptitude as "an individual's initial state of readiness and capacity for learning a foreign language, and probable facility in doing so [given the presence of motivation and opportunity]" (Carroll, 1981). Central to this definition are the notions that: 1) second language aptitude assesses skills in the present to make inferences about predicted success in the future, and 2) aptitude does not exist in a vacuum, but interacts dynamically with motivational and environmental factors. Second language aptitude is believed to be a relatively stable trait of an individual that is statistically separable from general intelligence (Sasaki, 1996).

Carroll's four-factor aptitude model proposes that the cognitive abilities that makeup second language aptitude can be separated into four distinct subcomponents. Specifically, the theoretical model suggests the subcomponent processes include: *phonemic coding ability, grammatical sensitivity, associative memory,* and *inductive language learning ability* (Carroll, 1962; 1981; 1990; 1993; Dörnyei & Skehan, 2003). Phonemic coding ability measures one's skill in learning and being able to recall and differentiate between new phonemic sounds. Grammatical sensitivity measures one's ability to parse the functional role of an informational unit within a larger structural context. Associative memory refers to one's capacity to form and recall new associations in memory. Inductive language learning measures one's ability to

extrapolate patterns from an existing corpus and apply them to novel contexts. Together these subcomponent processes make up the composite construct of second language aptitude.

While a variety of measures are used in the literature to assess second language aptitude (see Wen et al., 2017 for a discussion), the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959) has been reliably used as the 'gold standard' behavioral assessment for over 60 years (Chalmers et al., 2021; Saski, 2012). The MLAT is a standardized paper and pencil measure comprising five subtests (Carroll & Sapon, 1959). These MLAT subtests differentially draw on theorized underlying subcomponent processes that make up language aptitude, though they do not neatly map in a 1-1 manner. Typically, a composite score is calculated by summing the scores on the five subtests; this composite score has been well-validated as a predictor of second language learning. A meta-analysis with over 3,000 participants found a significant positive correlation between MLAT score and ultimate second language attainment (r = 0.34; Li, 2015). Additionally, several recent studies from our lab have demonstrated that the MLAT composite score is positively correlated with a range of programming skills at similar or stronger magnitudes (rs = 0.32 - 0.54; Kuo et al., 2022; Prat et al., 2020).

Despite the subcomponents embedded in Carroll's four-factor model, Carroll originally conceptualized the construct of second language aptitude within a unitary framework such that individuals were either high- or low-aptitude (Li, 2019). More recent work in the second language aptitude literature has theorized that different subcomponents of second language aptitude may be more or less relevant in particular contexts. For example, it has been proposed that phonemic coding and associative memory ability may be more strongly related to success in the early stages of learning relative to other subcomponents like inductive language learning and grammatical sensitivity (Skehan, 2002; 2012). This perspective aligns well with models

suggesting that vocabulary or the ‘mental lexicon’ of a new language is learned through fast declarative memory systems, whereas syntax or structural form rules are learned through slower procedural learning mechanisms (e.g., Ullman, 2001). Allowing the subcomponents of second language aptitude, theoretically, and the subtests of the MLAT, operationally, to be considered individually can help elucidate a more nuanced understanding of how second language aptitude functions at different skill levels and across individuals.

Examining the subcomponents of second language aptitude individually is particularly important for understanding how language skills contribute to programming. While it has been well established that *all* of the second language aptitude subcomponents contribute to second natural language acquisition, a particular subcomponent(s) may drive the observed relation between MLAT composite score and programming outcomes (Kuo et al., 2022; Prat et al., 2020). For example, grammatical sensitivity likely plays a clear role in programming skills considering that programming languages rely on syntax structures that are even more rigid than those seen in natural language (see Mottarella, Mortimore, & Prat, 2024 for a discussion). Likewise, learning a programming language requires amassing a new vocabulary of keywords, functions, and rules, and the successful retention and retrieval of this information likely varies as a function of associative memory ability.

Two recent studies from our lab provide converging support for the idea that programming acquisition and comprehension are related to grammatical sensitivity and recalling associated meanings of informational units. In one of these studies, we directly examined how the MLAT subtests most closely tied to grammatical sensitivity (MLAT IV) and associative memory (MLAT V) explained variation in programming learning across both Python and Java (Mottarella et al., 2024). Our results indicated that both MLAT subtests significantly predicted

variation in learning outcomes in both programming languages, however, the magnitude of these correlations were more robust for grammatical sensitivity. These results are consistent with another recent study from our lab showing that programmers show neural indices of semantic and syntactic processing during real-time code comprehension (Kuo & Prat, 2024). In this event-related potential (ERP) study, programmers read code that was either well-formed, had a syntax violation, had a semantic violation – such that the code would execute but the variables were named nonsensically, or had a doubly anomalous violation. Consistent with what has been observed in natural language and other related domains, programmers showed canonical N400 (Kutas & Federmeier, 2011) and P600 (Osterhout & Holcomb, 1992) effects when reading code with meaning or form violations, respectively. Additionally, individual differences in programming skills were associated with an observable shift from an N400 to a P600 deflection in response to form violations in code. This result strikingly mimics proficiency-related shifts seen in the second-language learning literature (McLaughlin et al., 2010). Together these studies point to important roles for grammatical sensitivity and the ability to map new meanings in memory in both learning to program and in real-time code comprehension.

These studies raise the related question of how other language aptitude subcomponents that may seem less intuitively related, like phonological coding, may influence programming.[1] To investigate this question, our lab has correlated the MLAT subtests that most closely measure phonological coding (MLAT II), grammatical sensitivity (MLAT IV), and associative memory (MLAT V) with individual differences in programming learning across three studies that ranged in the length of learning, sample-size, language background of the participants, and

[1] Inductive language learning ability is not focused on here as it is not directly measured by any of the MLAT subtests (see Wen et al., 2017 for a discussion). However, inductive learning may have interesting implications in programming as well - particularly when considering analogical transfer of concepts across programming languages (Kao et al., 2022)

programming language learned (Prat et al., 2020; Prat et al., in prep). Contrary to our initial predictions that phonological coding would be less critical in programming than the other language aptitude subcomponents, across all three studies MLAT II was significantly correlated with programming at a magnitude similar to or stronger than the other subtests (see Table 1). These results suggest that phonological skills may also play an important role in learning to program. In the section that follows I will dive deeper into what is known about phonological coding, and its association with programming and related cognitive processes.

The Role of Phonology in Complex Skills

The role of the phonological system in acquiring a second natural language has been well established (e.g., Aliaga-Garcia et al., 2011; Chalmers et al., 2021; Turker et al., 2021, Wen et al., 2017). Phonological skills can be understood broadly in this context as a set of abilities including the ability to: learn new sounds, distinguish between different sounds, map sounds to orthographic symbols and/or to semantic meanings, and recall these sounds or their related mappings when needed (e.g., Leinenger, 2014; Melby-Lervåg, Halaas Lyster, & Hulme, 2012; Saiegh-Haddad, 2019). Prior work has demonstrated behaviorally that measures of phonological skills including better discrimination on onset rime manipulation tasks, better identification of phonemes within words or pseudowords, and better memory of speech-symbol mappings predict more facile second-language learning (e.g., Haigh et al., 2011; Rokham et al., 2020; Saito, 2017; Verhagen, Leseman, & Messer, 2015). Moreover, neuroimaging work has shown that structural biomarkers associated with second language aptitude are heavily concentrated in phonologically related areas including the left inferior frontal gyrus, Heschel's gyrus, left inferior parietal lobule and its connectivity via the arcuate fascicle to related language areas (e.g., Golestani et al., 2002, 2007; Turker et al., 2021). Considering that natural language has speech and auditory

comprehension components it is easy to intuit how phonological processing would be important for second-language learning success.

The role of phonological skills in programming languages is less intuitive, however, several studies hint at the idea that phonological skills may be implicated here as well. Behaviorally, our lab has found a consistent pattern showcasing that MLAT II, the subtest most strongly tied to phonological skills, predicts learning to program in both Python and Java (Prat et al., 2020; Prat et al., in prep). These results suggest that better phonological skills are associated with better programming language acquisition – at least in the early stages of learning. Additionally, the neural network elicited during programming tasks involves regions associated with phonological processing including the left inferior frontal gyrus (e.g., Siegmund et al., 2014; 207; Xu et al., 2021), angular gyrus (e.g., Endres et al., 2021), and supplementary motor area (e.g., Endres et al., 2021). In a recent review of the programming neuroimaging literature, Castelhano and colleagues (2021) noted this similarity saying that: “each of the identified neuroimaging studies regarding programming revealed clusters that were also reliably activated in other studies assessing phonological processing…” (Castelhano et al., 2021, p. 9). However, no study to date has explicitly examined the overlap between these systems within the same sample of participants nor proposed a theoretical basis for *how* phonological skills might mechanistically relate to programming.

An additional line of evidence in support of the idea that phonological skills might be important for programming comes from a line of work demonstrating that the phonological system is implicated in arithmetic (e.g., De Smedt et al., 2010; Hecht et al., 2001; Zhou et al., 2018). A neuroimaging meta-analysis showed significant overlap between the neural networks supporting phonological processing and arithmetic (Pollack & Ashby, 2018). Moreover, the

brain areas that displayed this overlap – the left angular gyrus and posterior portion of the left IFG – converge nicely with areas frequently reported in neuroimaging studies of programming (e.g., Endres et al., 2021; Hishikawa et al., 2023; Ikutani et al., 2020; Liu et al. 2020; Peitek et al., 2020; Peitek et al., 2021; Siegmund et al., 2014; 207; Xu et al., 2021). The relation between phonology and arithmetic is also important to consider in light of the emerging narrative in the literature of whether programming is more 'language-like' or 'math-like' (e.g., Catelhano et al., 2021; Duraes et al., 2016; Graafsma et al., 2023; Liu et al., 2020; Xu et al., 2021). This framing may pose false dichotomies of what it takes to be good at programming. An alternate approach is to consider complex skills – like programming, language, and arithmetic – as an amalgamation of more elemental cognitive computations with the relative weight of these computations differentially producing each skill. Phonological processing may be an important underlying cognitive process that has implications for success in programming, arithmetic, and second language learning alike.

Taken together, prior work has found that similar brain areas are evoked in programming and phonological tasks, and that the phonological system supports other related complex skills like arithmetic and second-language learning. However, no study to date has explicitly examined if the phonological system is involved in programming. Additionally, there is currently no theoretical consensus about *how* phonological processing might support programming mechanistically. The goal of this dissertation is to use a neuroscientific individual differences approach to adjudicate between two theoretical hypotheses about how phonology might support programming. First, I propose that phonological processing may operate as a means to, or as a by-product of, lexical access. Second, I suggest that phonological systems may be utilized to

support inner speech functionalities during code comprehension. It is important to note that these hypotheses may not be mutually exclusive.

Phonology as a Gateway to Lexical Access

The role of the phonological system in silent reading has received ample attention in the natural language literature (for a review see Leinenger, 2014). A number of hypotheses have been proposed about exactly how phonological processing occurs in the context of silent reading. The primary differences between these perspectives concern *when* in the processing cascade from orthography to lexical access phonological processing occurs, and if phonological processing is *necessary* for lexical access.

All models of silent reading begin with the reader seeing an orthographic representation (i.e., the written word) and end with the reader accessing the lexical meaning of the word (i.e., a semantic memory of what the word means). The Activation Verification Model proposes that after encountering the orthographic word, the reader activates the phonological representation (i.e., the speech sound) of the word (Van Orden, 1987; see Figure 1A). The Activation Verification Model proposes that phonology plays an instrumental role in silent reading such that it occurs prior to, and is necessary for, lexical access. In contrast, the Parallel Distributed Processing Model proposes that the role of phonology in silent reading is epiphenomenal (Sidenberg & McClelland, 1989; see Figure 1B). This view suggests that from the orthographic word, the phonological code and the lexical meaning of the word are jointly activated but that the phonological code is not necessary for lexical access. The activation link between orthographic and phonological processing is likely a byproduct of the way reading is taught in early education with a focus on sounding out the words (e.g., Snow & Juel, 2005).

An alternative perspective is that the necessity of phonology for lexical access is variable based on circumstance. The Dual-Route Model proposes that both the reading skill of the individual and the frequency of the word dictate the role of phonological coding (Coltheart, 1980; Coltheart et al., 2001; see Figure 1C). For less skilled readers – or when experienced readers encounter a new or difficult word – phonological coding is necessary and precedes lexical access akin to the processing cascade proposed in the Activation Verification Model. However, in experienced readers, a direct route forms between the orthographic word and the lexical representation negating the need for phonological coding.

Experimental work to delineate between these hypotheses have yielded mixed results. One of the most common methods to study this relationship is to use homophone priming tasks (see Reynolds & Besner, 2005 for a review). Homophones are pairs of words that are semantically different but phonologically the same (e.g., BEAR and BARE). In priming paradigms, participants are presented with one word from the homophone pair and then presented with the second word from the pair later on. If participants respond more quickly to the second word from the pair than they do to other non-primed words, phonological priming is said to have occurred. Researchers have used this approach to study when phonological coding occurs relative to lexical access by comparing priming effects elicited by homophones to those elicited by pseudohomophones (e.g., BRANE and BRAIN), where the prime is a pseudoword with the same phonological sound as the target word. The inference made from this method is that if phonological coding occurs *prior* to lexical access, the priming effect should be a similar magnitude for psuedohomophones and homophones. By and large, the results of studies using this approach have found that homophones and pseudohomophones have equivalent priming effects, supporting theories that suggest phonological coding occurs before lexical access (e.g.,

Drieghe & Brysbaert, 2002; Lesch & Pollastsek, 1998). Studies using neurophysiological recordings have also suggested a very early role for phonological coding with related neural indices occurring within the first 100 ms after stimulus presentation (e.g., Ashby, Sanders, & Kinston 2009; Ashby, 2010; Halderman et al., 2012). Additionally, a meta-analysis of 35 neuroimaging studies found that phonological areas were more active when cognitive load was high (Jobard, Crivello, & Tzourio-Mazoyer, 2003). This finding is in line with the Dual-Route Model of lexical access, such that the phonological route is used when cognitive resources are stressed from task difficulty or lower skill of the individual. Taken together, the silent reading literature suggests that phonological coding occurs either as a facilitator to – or in parallel with – lexical access (Coltheart, 1980; Coltheart et al., 2001).

Implications for Code Comprehension: Phonology as a Gateway to Lexical Access. Modern-day programming languages include a large proportion of English words. As discussed previously, these words are both built-in and defined by the user when using coding best practices. Prior work has also commented on how the cognitive computations underlying comprehension of computer code are similar to those supporting natural language reading (Federenko et al., 2019). Thus, the role that phonological coding plays in the lexical access of individual words in natural language reading may also occur in comprehending written computer code. Additionally, if phonological coding is implicated in this way in code comprehension, then according to the Dual-Route Model the degree to which the phonological system is involved should vary as a function of individual differences in English skill (Coltheart, 1980; Coltheart et al., 2001). At lower levels of English proficiency, individuals should co-activate the phonological representations of the English word in the code enroute from the orthographic representation of the word form to the lexical access of the semantic meaning of the word. With

increased English skill, individuals may take advantage of the direct route between orthography and lexical access, bypassing or minimizing the activation of phonological representations in the process. Importantly, this theory of phonological coding proposes that phonological mechanisms are activated, at least for less-skilled English readers, at the individual token or word level. An alternate, though not mutually exclusive possibility, is that phonological representations are activated at the "chunk" level as a means to verbally rehearse multiple tokens in mind or to articulate verbal thoughts to oneself.

Phonology as an Inner Speech Mechanism

Researchers have long been interested in the ideas of inner speech and verbalized thought, though it has been a slippery concept to measure and define. Two primary theories have been proposed as to how inner speech may function: 1) Baddeley's Multicomponent Model of Working Memory, introduces the phonological loop as a working memory mechanism for rehearsing information in mind (Baddeley & Lewis, 1981; Baddeley, 1986), and 2) Vygotsky's Sociocultural Theory, suggests that inner speech functions as a means to turn external conversations inwards to produce verbal thought (Vygotsky, 1934; 1987). In the context of understanding the role of the phonological system in programming, both of these theories may have some relevance.

Baddeley's Multicomponent Model proposes that working memory is supported by a modality-specific multicomponent system (Baddeley & Lewis, 1981; Baddeley, 1986). In this system, visual information is maintained in the visuospatial sketchpad, verbal information is maintained in the phonological loop, the episodic buffer manages the overlap between these systems, and the central executive functions as a guide that directs capacity-limited attentional resources towards the most relevant information at hand. If enough attentional resources are

devoted to a piece of information in the working memory buffer that information is encoded into long-term memory. Retrieval of the information from long-term memory also requires attention to pull the information back to conscious awareness within the working memory buffer.

The phonological loop is made up of both a passive phonological store and an active articulatory rehearsal mechanism (Baddeley, 1992). The phonological loop's ability to maintain information in working memory is contingent on both the ability to 'articulate' the to-be-remembered information to one's self and the ability to 'hear' the rehearsed information in mind. Thus, the phonological loop is said to contain both a speech production, or 'inner voice', and a covert hearing, or 'inner ear,' component (Buschsbaum, 2013). Together, these components create an attentionally modulated conscious experience of speech sounds used to keep information active (e.g., Perrone-Bertolotti et al., 2014).

Simultaneous overt speech can disrupt the experience of inner speech and its ability to serve as a working memory rehearsal mechanism. Even simple articulatory suppression paradigms where participants are asked to repeatedly say aloud a sound (e.g., "ba ba ba") can disrupt the encoding, maintenance, and retrieval of verbal information in working memory whereas spatial dual-tasks (e.g., finger tapping) do not (e.g., Alderson-Day & Fernyhough, 2015). These articulatory suppression approaches can be used to test whether a process relies on inner speech, with the inference being that if performance on a given task is reduced when participants are engaged in articulatory suppression, then the task elicits inner speech. Studies using this approach have demonstrated that articulatory suppression disrupts linguistic processes including silent reading comprehension (e.g., Coleheart, Avons, & Trollope, 1990) as well as cognitive control processes like task switching (e.g., Emerson & Miyake, 2003; Saeki, 2007) and logical reasoning (e.g., Wallace, Peng, & Williams, 2017). The finding that task switching is also

impaired by articulatory suppression is consistent with the idea that goal maintenance or similar metacognitive processes typically ascribed to the central executive may also draw on verbal resources.

On the other hand, Vygotsky's perspective on inner speech proposed that inner speech functions as a means of turning social conversations inward across a developmental trajectory (Vygotsky, 1934; 1987). Vygotsky believed that linguistic development evolved in stages such that in early childhood speech was fully external, then progressed to "private speech" where individuals engaged in outward speech to themselves, and finally speech was turned fully inwards as a means of self-regulation (Vygotsky, 1934; 1987). More contemporary perspectives have acknowledged that this process is likely not as linear as Vygotsky initially proposed, with private speech remaining for many individuals well into adulthood (Alderson-Day & Fernyhough, 2015). However, Vygotsky's initial theory that inner speech is the product of internalizing an ability that begins as overt speech remains a strong theoretical influence in modern-day understandings about inner speech (Langland-Hassan, 2020). Vgotsky's ideas have also led to theories that the purpose of inner speech may stem from similar root causes as external speech: "in conversations, we pool our cognitive resources with those of others, and thereby solve problems we could not have solved on our own…in learning to converse with ourselves, we learn to bring together resources from different quarters of the mind" (Gauker, 2018, p. 66). Individuals may utilize the resource pooling functionalities of inner speech to cope with difficult task demands and solve novel problems (e.g., Kompa & Mueller, 2022).

As speech becomes more internalized over the course of development, syntactic and semantic shortcuts can be taken in the way that speech is internally represented (Fernyhough & McCarthy-Jones, 2013). A model by Fernyhough (2004) suggests that by default, inner speech

should include a condensed syntactic and semantic structure, but that this inner speech representation can become expanded such that it is more akin to the phonological, syntactic, and semantic elements one expects during a spoken conversation when cognitive resources are stressed by limited capacity or a challenging problem. The idea that under capacity constraints internal representations become more expanded – moving away from pure meaning and towards a complex linguistic representation including phonology – mirrors the logic proposed in the Dual-Route Theory of lexical access at the word level (Coleheart, 1980; Coleheart et al., 2001). In both theories, individuals recruit the phonological system when cognitive resources are stressed either through capacity limits of the individual or through increased task demands.

Testing theories of inner speech can be quite challenging with Vygotsky himself famously noting that "the area of inner speech is one of the most difficult to investigate" (Vygotsky, 1986, p. 226). Indeed, measuring indices of inner speech can be tricky as there is no clear outward behavior to signal whether inner speech is occurring (for a discussion see Langland-Hassan, 2020). Self-report questionnaires that ask participants to describe or rate their inner speech processes are frequently used in this research domain (e.g., Alderson-Day et al., 2018; Burnett, 1996; Brinhaupt, Hein, & Kramer, 2009; Hardy, Hall & Hardy, 2005; McCarthy-Jones & Fenyhough, 2011). However, self-report measures rely on an individual's ability to introspect and accurately report that introspection – abilities that can vary substantially across individuals and contexts. An alternate approach is to use neurophysiological techniques to "peek under the hood" and observe the underlying neural mechanisms that are active during inner speech even in the absence of overt behavioral responses.

Prior neurophysiological work has suggested that inner speech relies on a neural network that is partially – though not fully – overlapping with overt speech comprehension and

production regions (for reviews of the overt speech network see Hickok & Poeppel, 2007 and Price, 2012). Neuroimaging studies have shown similar activation in classic left-lateralized language areas including portions of the inferior frontal gyrus particularly the pars opercularis, inferior parietal lobule, and superior temporal gyrus for inner and overt speech (see Geva et al., 2018 and Perrone-Bertolotti et al., 2014 for reviews). The importance of the inferior frontal gyrus in inner speech is further supported by lesion mapping work in stroke patients which demonstrated that lesions to the left inferior frontal gyrus – in particular, the pars opercularis (BA 44) and the surrounding arcuate fasciculus and the left supramarginal gyrus (BA 40) – were correlated with behavioral deficits in rhyming and homophone judgment tasks (Geva et al., 2011). Additionally, neuroimaging work has supported the theoretical intuition that inner speech increases when cognitive resources are stressed (Fernyhough, 2004). During a working memory paradigm that varied in load and manipulation demands, greater activation was observed as the working memory demands increased in a network of regions including areas implicated in inner speech like the left inferior frontal gyrus, left premotor cortex, and bilateral parietal areas (Marvel & Desmond, 2012).

Inner Speech as a Verbal Code-of-Thought Strategy. While task demands are a key driver of inner speech, they likely do not act in isolation but interact dynamically with individual differences in strategy. Prior work has found that the propensity for engaging in inner speech is highly variable (see Hurlburt, Heavy, & Kelsey, 2013 for a discussion). In one study examining how frequently inner speech occurred, some participants reported never engaging in inner speech whereas others reported engaging in inner speech as much as 75% of the time (Heavy & Hurlburt, 2008). This variability suggests that while inner speech may provide a structure for thinking for some people, this is not a universal experience with some individuals never

engaging in inner speech at all (e.g., Nedergaard & Lupyan, 2024). Rather, the tendency to engage in inner speech may vary as a function of individual differences in the ways that people prefer to internally represent information.

This idea intersects with the code-of-thought literature suggesting that there are differences in individual's tendencies to represent information more verbally or more visually (e.g., Alfred et al., 2020; Kirby et al., 1988; Kraemer, Rosenberg, & Thompson-Schill, 2009), and that these differences are stable within a person across time (Miller et al., 2012). Work in this domain has primarily used self-report questionnaires, which ask participants to introspect on how they represent information (e.g., Antonietti & Giorgetti, 1998; Kozhevnikov et al., 2002; Roebuck & Lupyan, 2020), and experimental attentional-bias tasks, which measure biases towards visual or verbal information (e.g., Alfred et al., 2020; Kraemer et al., 2009), to assess code-of-thought. Prior work in this domain has demonstrated that individuals differ in the ways they attend to and internally represent information in both laboratory and naturalistic contexts (see Alfred & Kraemer, 2017 for a review).

In most cases, people *can* attend to and represent information both verbally and visually (though see Nedergaard & Lupyan, 2024 for an exception). For example, in a naturalistic navigation task, individuals who self-reported having a more verbal code-of-thought tended to better recall landmarks, whereas individuals with more visual codes-of-thought tended to better recall the relative distances between locations (Kraemer et al., 2017). However, when participants in this study were explicitly instructed to pay attention to the landmarks or the distances between locations, there was no difference in recall for either type of information between individuals with visual and verbal codes-of-thought. This study nicely highlights that while individuals may have preferences for how they construct their internal experience, this

construction is often flexible to the demands of a task. From an evolutionary perspective it is easy to see why having variability in the kinds of information individuals notice, like landmarks or distances, would be useful in building a diverse knowledge base at the population level.

Neuroimaging work has suggested that both self-reported and experimentally measured code-of-thought have a biological basis in the brain. In a fMRI study where participants were exposed to both visual and verbal stimuli in the scanner, participants showed evidence of mentally converting the modality of the stimulus when it was misaligned with their preferred code-of-thought (Kraemer et al., 2009). When participants with a more verbal code-of-thought performed the visual task in the scanner, they activated phonological regions in the left supramarginal gyrus. Conversely, when participants with a more visual code-of-thought performed the verbal task in the scanner, they recruited visual areas in the right fusiform gyrus. These results suggest that participants may convert information to their preferred information processing modality.

Key Take-Aways from the Inner Speech Literature. Taken together, prior theoretical and empirical work suggests several key points about the nature of inner speech. First, two primary theoretical perspectives have been put forth as to how inner speech operates, and the phenomena these perspectives described likely coexist to some degree. Baddeley's theories of the phonological loop suggest that internal speech functions as a mechanism by which verbal information is kept active in working memory, whereas Vygotsky and his contemporaries have suggested that inner speech evolves from turning external speech inwards. These two perspectives are not mutually exclusive, rather the working memory view describes a more specific instance in which inner speech is necessarily recruited, whereas Vygotsky's ideas may speak more generally to the ways that internal verbalization can be used towards offloading

cognitive demands and supporting metacognitive processes like goal setting, self-regulation, and introspection. Second, inner speech is a slippery construct that is challenging to measure due to a lack of observable behavioral responses. Neurophysiological methods are well-suited to address these challenges. Third, prior work on the inner speech neural network has implicated a network of regions including portions of the left inferior frontal gyrus, inferior parietal lobule, superior temporal gyrus and in some cases, pre-motor and articulatory planning areas; activation in these areas has been shown to increase with greater cognitive demands. This network of regions is also consistent with many of the areas implicated in neuroimaging studies of computer programming (e.g., Hishikawa et al., 2023; Ikutani et al., 2020; Liu et al. 2020; Peitek et al., 2020; Peitek et al., 2021; Siegmund et al., 2014; Siegmund et al., 2017). Finally, inner speech has been suggested to play an important role in a diverse range of cognitive and linguistic tasks including reading, language acquisition, language comprehension, memory, and metacognitive thinking and reflection (see Alderson-Day & Fernyhough, 2015 and Perrone-Bertolotti et al., 2014 for reviews). However, the tendency to engage in inner speech may vary as a function of individual differences in code-of-thought.

Implications for Code Comprehension: Phonology as a Means to Inner Speech.

Theories of computer programming comprehension suggest two mechanisms by which programmers can parse meaning from code (see Siegmund et al., 2014; 2017 for discussions). Top-down comprehension occurs when a programmer uses past experience to generate predictions, or prior likelihoods, of what a particular source code statement may include, and uses these predictions to guide the understanding of the code (Brooks, 1978; Soloway & Ehrlich, 1984). This top-down method of comprehension builds on Bayesian style logic whereby priors guide the way the programmer expects to encounter code, and are refined to generate stronger

and more accurate predictions with repeated experience. In early stages of learning to program, or when programmers encounter novel or difficult problems, the appropriate priors are not available to guide a top-down prediction. In such cases, programmers must rely on bottom-up comprehension.

Theories of bottom-up comprehension suggest that programmers understand source code by linking together the meanings of individual tokens in the code (Pennington, 1987; Shneidermand & Mayer, 1979). This process is working memory demanding as it requires multiple code tokens and their associated meanings to be kept active and combined in order for the programmer to extract the meaning of the larger source code statement. The phonological loop is well-suited to act as a mechanism for maintaining these code tokens in mind. If the phonological system supports bottom-up code comprehension, then the phonological system should be more engaged for less skilled programmers or when task demands are particularly challenging. This theory is also consistent with the idea that experts can "chunk" together domain-specific information to reduce working memory load (e.g., Gobet & Simon, 1996).

Alternatively, inner speech may function during code comprehension as a characteristic that stems from differences in code-of-thought. Individuals with more verbal codes-of-thought may engage inner speech via the phonological system to support code comprehension whereas those with more visual codes-of-thought may use alternative strategies. A study by Zarnhofer and colleagues (2013) observed similar results using an arithmetic paradigm in the scanner. Participants with more verbal codes-of-thought engaged the left angular gyrus, a key region integrating phonological and semantic information, when completing arithmetic problems; conversely, individuals with more visual codes-of-thought showed greater activation in visual processing areas. Critically, this study also found that code-of-thought was not related to

behavioral differences in intelligence or in arithmetic performance, suggesting this difference in strategy was about more than capacity or skill differences.

In summary, code-of-thought seems to be variable across individuals, stable within an individual, and instantiated in the neural architecture of the brain. Typically, code-of-thought does not relate to performance ability - suggesting that there is not a 'better' or 'worse' code-of-thought. By this view, inner speech may suggest a propensity for verbal thinking such that phonological resources are only used during code comprehension by individuals with a verbal code-of-thought. Moreover, the extent to which an individual represents information using verbal brain areas should be stable across task domains but should not relate to programming skill or other measures of cognitive performance like working memory capacity, however it should relate to self-reported code-of-thought and/or to performance on attentional bias tasks.

The Role of Phonology in Code Comprehension: Interim Summary & Hypotheses

Prior work has suggested that phonological processes may be recruited during programming to some degree. Support for this assertion comes from three converging lines of evidence. First, behavioral work from our lab has demonstrated that second language aptitude is a robust predictor of learning to program and that MLAT II, the language aptitude subtest most closely tied to phonological coding abilities, predicts programming acquisition at a magnitude similar to or stronger than other language aptitude subtests (Prat et al., 2020; Prat et al., in prep). Second, the neural network supporting programming shares noticeable overlap with the network of regions involved in phonological processing including the left inferior frontal gyrus, left supramarginal gyrus, and supplementary motor cortices. This observed overlap between the programming and phonological networks have been noted in prior research (Castelhano et al., 2021; Siegmund et al., 2014) but has not been explicitly investigated. Finally, the two other

complex skills that programming is most commonly believed to derive from, mathematics and natural language, have both been associated with phonological processing at the behavioral and neural level (e.g., Chalmers et al., 2021; De Smedt et al., 2010; Pollack & Ashby, 2018; Turker et al., 2021). Together, these findings suggest that the phonological system may be involved in programming. However, no work to date has explicitly investigated the proposed relationship between phonological processes and programming nor proposed a theoretical basis for how the phonological system might operate during code comprehension. In my dissertation work I propose several hypotheses about how the phonological system may be implicated in code comprehension.

The *Lexical Access Hypothesis*, is based on theories of silent reading which suggest that the phonological codes associated with orthographic words are activated enroute to retrieving the lexical meanings of the words (e.g., Coltheart, 1980; Sidenberg & McClelland, 1989; Van Orden, 1987). Modern day programming languages, like Python, include a large portion of English words that are both built-in and defined by the user when using coding best practices. The *Lexical Access Hypothesis* proposes that phonological codes are activated at the word-level to aid access to the lexical meaning of English words in code, and that the degree to which the phonological system aids lexical access should be tied to individual differences in English skill.

The *Inner Speech Hypothesis*, proposes that the phonological system is activated to support form an internal representation. There are two primary ways that inner speech processes have been conceptualized in the literature, both of which have potential implications for how the phonological system is involved in programming. First, inner speech can function as a means to maintain or manipulate verbal content in working memory via the phonological loop (e.g., Baddeley & Hitch, 1974; Baddeley, 1986, Baddeley, 2000). In this *Inner Speech: Working*

Memory Variant Hypothesis, the phonological loop is engaged during bottom-up code comprehension (e.g., Pennington, 1987; Shneidermand & Mayer, 1979), to keep individual code tokens active in mind long enough to extract meaning from the larger source code statement. Alternatively, inner speech can function as a strategic difference in the way that individuals prefer to represent information to themselves. The *Inner Speech: Code-of-Thought Variant Hypothesis* proposes that individuals vary in the extent to which they represent information verbally or visually. Individuals with more verbal codes-of-thought may selectively recruit the phonological system to support inner speech during code comprehension, whereas those with more visual codes-of-thought may use alternative strategies.

A central challenge of investigating internal verbalization processes is that they occur in the absence of observable behavior. Neuroscience methods are well suited to address this challenge as they allow for the examination of underlying neural processes in the absence of a behavioral output. While a number of recent studies have used neuroimaging methods to examine to the neural network that supports program comprehension at the group-level, relatively few studies to-date have examined how variation in these neural systems relates to individual differences in behavior (Castelhano et al., 2019; Floyd et al., 2017; Hishikawa et al., 2023; Ikutani et al., 2021; Ivanova et al., 2020; Liu et al., 2020; Peitek et al., 2020; 2021). Studies that have considered individual differences have been limited in power and scope due to having relatively small sample sizes (Ns < 30) and only considering programming skills, typically assessed in the scanner, as a behavioral regressor. My dissertation work uses the largest sample size of any study in this research area to date (N = 44) and a wide range of behavioral covariates to put individual differences at the forefront of my research approach.

The Present Study: An Argument for a Neuroscientific Individual Differences Approach

Individuals vary greatly in their ability to acquire and utilize a programming language. For example, in work from our lab where we had novice learners move through a Python learning environment at their own pace, the fastest learners acquired Python at a rate more than double that of the slowest learners (Prat et al., 2020). Even when completing much simpler tasks, like associating a stimulus with a button-response, individuals from the same relatively homogenous sample can vary substantially in the strategies and underlying cognitive mechanisms they employ (e.g., Collins, 2018; Haile, Prat, & Stocco, 2024). Considering the cognitive complexity of programming and its intersection with sociocultural and environmental factors (e.g., Cheryan, Plaut, Handron & Hudson, 2013; Frachtenberg & Kaner, 2019), it is even less likely that a 'one-size-fits-all' model of programming will successfully explain skill variance. Understanding the brain systems that support programming adds an additional layer of complexity and potential for important variance across individuals.

Individual variability in the brain primarily stems from three sources. First, differences in the underlying cognitive processes that participants are engaged in can lead to resulting variability in the brain regions supporting these processes (e.g., Collins, Ciullo, Frank, & Badre, 2017; Kraemer et al., 2009; Miller et al., 2012). This source of variability is important for understanding how different task strategies or compensatory behaviors are employed across individuals.

Second, there can be differences in the anatomical brain structures under investigation. Variation in the size, shape, and composition of brain areas can be predictive of behavioral performance differences (e.g., Kanai & Rees, 2011). For example, higher natural language aptitude has been associated with structural differences including: greater gray matter volume in

left-lateralized language processing areas, greater gyrification in the auditory cortex, and greater myelination of the articulate fascicle which connects language processing regions in the left inferior frontal gyrus to the temporoparietal junction (Turker et al., 2023). However, even when anatomical differences do not directly predict behavioral differences, they are still important to consider methodologically. Most neuroimaging work assumes a generalized averaging approach where cognitive functions are mapped onto anatomical structures based on what regions are elicited during the task at the group-level. However, when anatomy differs across participants, the function ascribed to those regions may not accurately capture the peak functional response for each individual.

Finally, there can be individual differences in the function of brain areas and their connectivity to one another. Individual differences in functional connectivity can be used to predict behavior (e.g., Elliot et al., 2019; Reineburg et al., 2015; Shen et al., 2017) and can be modulated by experiences (e.g., Ellwood-Lowe, Whitfield-Gabrieli, & Bunge, 2021; Rakesh et al., 2021). Additionally, when multiple regions and their connectivity patterns are considered in tandem, these networks can be variable in both size (i.e., how fine grained or large scale) and functional specificity (i.e., how many cognitive processes use the network). The degree to which these functional networks are lateralized to one hemisphere or involve bilateral communication between hemispheres is another interesting source of variability. Individual differences in lateralization have been related to cognitive skills (Chiarello, Welcome, & Leonard, 2012) but are also influenced by factors like handedness (e.g., Eviatar, Hellige, & Zaidel, 1997), which has led to the common practice of excluding left-handed participants from neuroimaging studies.

Considering the multitude of ways that brains can vary is important for both understanding individual differences in behaviors – particularly when those behaviors are

complex in nature – and for informing methodological decisions. One way that individual variability can be accounted for methodologically is to use a functional localizer. The basis of the functional localization approach is to have participants complete a task that elicits a particular cognitive process of interest, and then use the subject-specific activity from that localizer task to define an individualized region-of-interest (ROI) that can be analyzed in a subsequent target task. This approach benefits from being sensitive to individual differences in both anatomy and function, but it is not without its challenges. In practice, it can be difficult to determine how much to constrain the statistical map from the localizer. One approach to this issue is to use theoretically driven search areas and then take the portion of the search area that is most active for each subject as the ROI (Nieto-Castañón & Fedorenko, 2012). However, decisions about how large these search areas should be, how many voxels within a search area should be included in the ROI, and what to do with activation clusters that extend into multiple search areas remain somewhat arbitrary decision points. Despite these challenges, functional localization is a theory-driven way to consider functional and anatomical variation across individual brains and can allow for more inclusive recruitment criteria.

In my dissertation work, I will use a functional localization approach to systematically examine the role of the phonological system in code comprehension. Specifically, I will use a rhyming judgment task (adapted from Yen et al., 2019) to isolate phonological regions and then examine these subject-specific ROIs during a secondary code comprehension task. This individualized approach is well-suited to this question as one of the primary brain areas involved in both phonological and programming tasks – the left inferior frontal gyrus – is known to have substantial variability in the macro and micro anatomy across individuals (e.g., Amunts et al., 1999; Federenko & Blank, 2020). An additional consideration is that many of the areas involved

in the programming neural network can have both language specific and domain general functionalities in extremely close proximity (e.g., Federenko, Duncan, & Kanwisher, 2012), which can make it challenging to determine what role a brain area is playing when it is activated in a programming task. Using a small ROI defined from the region immediately surrounding the peak activation in the phonological localizer will allow me to examine how tightly coupled phonological and programming processes are while minimizing the chance of observing separable – but proximate – brain areas in each task.

An individual differences approach can also help delineate between the hypotheses proposed herein about *when, how,* and *for whom* the phonological system may be involved in program comprehension. The *Lexical Access Hypothesis* suggests that phonological processing occurs at the individual word level due to the large proportion of English words in Python. Under this hypothesis, the extent to which the phonological system is involved in code comprehension should be related to the lexical demands of the stimuli.

The *Inner Speech: Working Memory Variant Hypothesis* proposes that code tokens are maintained in the phonological loop in order for the programmer to extract a larger meaning from the code statement. Under this hypothesis, the degree to which the phonological system is involved in code comprehension should vary as a function of both the stress put on the cognitive system and the programming experience of the participant. Prior work supports the idea that phonological rehearsal is employed when cognitive resources are scarce and information needs to be maintained through brute rehearsal (e.g., Marvel & Desmond, 2012). As such, it is predicted that individuals with lower working memory capacities will be more likely to rely on the phonological loop to support code comprehension. Additionally, theories of code comprehension have suggested that bottom-up comprehension, whereby code tokens are

processed individually and linked together to form a meaning representation, occurs when top-down predictions of what is expected in the code are not available (Pennington, 1987; Shneidermand & Mayer, 1979). Thus, it is predicted that less experienced programmers will be more likely to engage in bottom-up code comprehension and as a result recruit more phonological resources.

The *Inner Speech: Code-of-Thought Variant Hypothesis* proposes that individuals vary in the extent to which they engage in internal verbalization but that this variability is not directly related to better or worse cognitive performance. Under this hypothesis, it is expected that individuals will vary in the extent to which they engage phonological brain areas during code comprehension, but that this variability will not be related to skill or cognitive capacity. Rather, this variability should relate to individual differences in code-of-thought, such that individuals with verbal codes-of-thought will recruit phonological resources more than individuals with visual codes-of-thought.

The present study adds to the rapidly growing literature on how individuals understand computer code. Prior work has demonstrated that individual differences in second language aptitude in general – and its subskills including phonological coding ability – predicts variability in programming acquisition. In this dissertation, I will build on this foundation to examine if phonological processing supports code comprehension in real-time. By analyzing the extent to which phonological activation during code varies as a function of individual differences in behavior, I will be able to gain traction on the theoretical basis of how these two processes may relate. To my knowledge this is the first study to examine the nature of phonological processing in programming by assessing individual differences in the brain areas responsible for phonological skills and computer code comprehension. Additionally, while most work to date

has primarily focused on group-level averages, my approach puts individual differences at the forefront and includes a large sample composed of diverse programming skill levels and language backgrounds. This rich data set combined with an individual differences approach will allow me to delineate between the hypothesized roles of the phonological system in code comprehension proposed herein.

Methods

Participants

Forty-nine healthy adults between the ages of 18 and 50 were recruited for participation in this study (18 females, 31 males, *M* = 23.16 years old). Participants were required to have some prior experience (operationalized as eight hours or more) with Python before participating. All participants were required to be fluent in English, but could have fluency in other natural languages as well. Both right and left-handed individuals were included in the sample (one participant reported being left-handed and one participant reported being ambidextrous). Before the MRI session, one participant withdrew from the study due to feeling claustrophobic in the scanner, one participant was dismissed for not being able to read the screen clearly with the available MRI-safe glasses, and one participant was dismissed due to not meeting the MRI safety-screening criteria. Neuroimaging data was collected from 46 participants (18 female, 28 male, *M* = 23.11 years). One participant was excluded due to having below-chance behavioral performance on one of the scanner tasks and meeting the criteria for excess motion and another participant was excluded due to a technical error with the scanner. The final neuroimaging sample included usable data from 44 participants (17 female, 27 male, *M* = 23.16 years). All

participants provided informed consent in accordance with the standards set forth by the University of Washington Institutional Review Board and were paid for their participation.

While participants' experience with Python and their language backgrounds were not explicitly controlled for, the resulting sample included a diverse range of language and programming experiences. Using the Python skill criteria from Kuo & Prat (2024), the final sample consisted of 26 novices and 20 experts. The language backgrounds of the participants were similarly diverse. Participant language background was classified using the following criteria. Monolinguals were highly proficient in understanding only one language with no other language exposure before age 10 (N = 11). Late bilinguals were highly proficient in understanding two or more languages with exposure to the second language after seven years old (N = 14). Early bilinguals were highly proficient in understanding two or more languages with exposure to both languages before seven years old (N = 18). One additional participant did not meet any of these criteria as they had early second language exposure but were minimally proficient. Of the final sample, 26 participants reported that their first language was English and the remaining 18 participants reported a non-English first language (Chinese: N = 13, Punjabi: N = 2, Bengali: N = 1, Gujarathi: N = 1, and Korean: N = 1).

Measures

Behavioral Measures

Modern Language Aptitude Test (MLAT): Part II Phonemic Script. The Modern Language Aptitude Test (MLAT) is a standardized paper and pencil measure designed to assess one's potential for acquiring a second natural language (Carroll & Sapon, 1959). The MLAT is separated into five subtests that differentially rely on the theorized underlying components of language aptitude (i.e., grammatical sensitivity, phonological coding, etc.). In Part II: Phonemic

Script, participants hear sets of four auditorily presented phonemes and learn to associate them with written representations on their answer sheets. After five sets of four phonemes had been learned, participants returned to the first set they had learned, heard only one phoneme, and had to choose the corresponding written representation on their answer sheet. The task proceeded in this manner with participants learning another five sets of four phonemes and then being tested on each set until participants had completed 30 test items. In total the task took approximately 12 minutes to administer and was always administered in person. Phonological coding skill was measured as the number of items correct out of 30 possible points.

Forward Digit Span. The Forward Digit Span task (Blackburn & Benton, 1957) is a classic measure designed to assess route working memory capacity. In this version of the task, participants viewed serially presented letters for 750 ms with a 250 ms blank screen between each letter, then were tasked with typing the letters they saw, in order, into a text box. The set sizes of the trials ranged from three to nine items, participants completed two trials at each set size and the difficulty of the set sizes increased linearly throughout the task. Working memory capacity was operationalized as the highest set size at which participants were able to recall all items in order on both of the task trials. The task was administered using PsyToolkit (Stoet, 2010; 2017) during the Zoom session.

Nelson Denny Reading Test. The Nelson Denny Reading test is a measure of English ability normed on college students (Brown, 1960). Participants completed the comprehension portion of the test where they were asked to read short passages and answer multiple-choice comprehension questions. In this version of the test, the passages and questions were put into an online survey using Qualtrics (Qualtrics, Provo, UT) and subject responses were collected during the Zoom session. *Nelson Denny Reading Rate* was measured as the number of words the

participant read in the first passages of the test in one-minute. Participants were then given an additional 19 minutes to complete the *Nelson Denny Comprehension Test*. A research assistant alerted the participants over Zoom when their time was up. The comprehension test consisted of seven unique reading passages and 38 multiple-choice questions. English reading skill was operationalized using a normed percentile score based on the number of questions participants answered correctly on the Nelson Denny Comprehension Test.

Language Experience and Proficiency Questionnaire (LEAPQ). The Language Experience and Proficiency Questionnaire (LEAPQ) is a self-report measure of the natural languages a participant knows and their experiences and skills in using each language (Marian, Blumenfeld, & Kaushanskaya, 2007). The LEAPQ was administered online via a Qualtrics survey (Qualtrics, Provo, UT) during the Zoom session. Self-reported English skill was assessed by averaging the scores together from the questions that asked participants about their proficiency in: reading, speaking, and understanding English. Each of these questions was assessed using a 10-point Likert scale.

Python Experience and Proficiency Questionnaire (PEAPQ). The Python Experience and Proficiency Questionnaire is a self-report measure of experience and skill with the programming language Python (Kuo & Prat, 2024). The PEAPQ was administered online via a Qualtics (Qualtrics, Provo, UT) survey during the Zoom session. Self-reported Python skill was assessed by averaging the scores together from the questions that asked participants about their proficiency: writing code, reading code, and debugging code. Each of these questions was assessed using a 10-point Likert scale.

Python Declarative Knowledge Test. The Python Declarative Knowledge test was a 72-item multiple-choice assessment of Python ability (Kuo & Prat, 2024; Prat et al., in prep). The

test was administered via Google Forms during the Zoom session. The first 36 items focused on Python knowledge more closely tied to semantics (*e.g., "What does str(24) do?"*), and the second 36 items focused on Python knowledge more closely tied to syntax (*e.g., "Why will this code generate an error?"*). Participants were given 30-minutes to complete the test and could navigate freely throughout the entire set of questions. Python skill was assessed as the number of items participants got correct out of the total 72 possible points.

Card Sorting Task. The Card Sorting task (Alfred et al., 2020) is a measure of variability in individuals' propensity to attend to visual versus verbal information. In the task, participants are instructed to sort playing cards by suit (i.e., heart, spade, club) by pressing one of three keys on a keyboard. Each card contains both a verbal word (e.g., "heart") and a symbol denoting a suit (e.g., ❤️). In 75% of the trials, the word and symbol are congruent (e.g., ❤️ & "heart") and in 25% of the trials the word and the symbol are incongruent (e.g., ❤️ & "spade"). Participants were not told beforehand that there would be incongruent trials or given any instructions on how to sort them. The position of the word and symbol were counterbalanced such that each appeared on the top half of the screen 50% of the time. Participants completed four blocks of trials, with each block consisting of 48 trials. The trials were not time-limited, however, participants were instructed to respond as quickly and as accurately as possible. The task was administered online using PsyToolKit (Stoet, 2010; 2017) during the Zoom session. A verbal bias score was calculated for each participant using the following formula: (# of incongruent word responses - # of incongruent symbol responses) / (total "correct" incongruent responses), where correctness indicated that participants selected a button corresponding to one of the two suits present in the incongruent trials (e.g., heart or

spade) and *not* the button corresponding to the third suit (e.g., club). Verbal bias score was used as a measure of code-of-thought with more positive scores indicating that a participant had a more verbal bias score and more negative score indicating that a participant has a more visual bias score.

Internal Representations Questionnaire (IRQ). The Internal Representations Questionnaire (IRQ) is a self-report survey that asks participants to rate the extent to which they agree with statements about how they represent information (Roebuck & Lupyan, 2020). The version of the questionnaire employed herein was adapted to include only the questions that loaded most strongly onto the visual and verbal factors. Four additional questions were also added that asked about the extent to which individuals engaged in visual and verbal information processing while writing and debugging code. Including these added questions, the final survey contained 12 items measuring visual information processing, and 14 items measuring verbal information processing (for a full list of questions see Appendix A). For each survey item participants rated their agreement using a 5-point Likert scale ranging from *1-Strongly Disagree* to *5-Strongly Agree*. Participants completed the IRQ questions in PsyToolkit (Stoet, 2010; 2017) directly following the Card Sorting task and were given unlimited time to make their selections. Performance on the IRQ was split into an IRQ-Visual and IRQ-Verbal dimension by averaging the scores for the visual and verbal dimension questions respectively. Scores on these two dimensions were calculated independently from one another.

fMRI Measures

Phonological Localizer Task. The Phonological Localizer task was modeled after the rhyming paradigm developed by Yen and colleagues (2019). Participants completed two types of task trials: *Rhyming* and *Letter Search* trials (see Figure 2). Trials were presented in blocks of six

trials. Blocks began with a cue presented for 2500 ms instructing the participant which condition they would be completing next (RHYME or “D” SEARCH) followed by 3000 ms fixation. Each stimuli pair was presented vertically above and below a central fixation with a prompt at the bottom of the screen reminding participants which button to press for “YES” and “NO”. Participants were given 3750 ms to make their responses, if they responded before this time had elapsed the response prompt on the bottom of the screen would disappear but the stimuli pair and fixation would remain until the allotted 3750 ms had elapsed. A 750 ms fixation was presented between trials within a block. At the end of each block, there was a 12000 ms interblock fixation to allow the hemodynamic response to return to baseline before beginning the next block of trials. The task was broken up into two MRI runs. Each run lasted approximately five-minutes and contained eight blocks (four per condition, presented in alternating order). In total participants completed 48 trials per condition. Before completing the task in the scanner, participants completed a behavioral practice version of the task containing four blocks (two per condition) on a lab-provided computer.

In *Rhyming* trials, participants were tasked with determining if two nonwords rhymed. Nonword stimuli were selected from the stimuli pairs used in the Yen et al (2019) paradigm. Nonword pairs that differed in how the endings of the nonwords were spelled (e.g., VEERY & FEARIE as opposed to VEERY & FEARY) were prioritized in the stimuli set to encourage phonological processing of the stimuli and minimize making decisions based on the visual word form. The original Yen et al. (2019) paradigm included adaptive difficulty levels, in the present study we used a fixed stimuli set where the difficulty of the items based on Yen and colleagues' 7-point categorization ranged from 2 - 4 (M = 3.04).

In *Letter Search Trials*, participants viewed pairs of consonant strings (e.g., DPLS & VDMN) and were tasked with determining if the letter "D" was present in both consonant strings or only in one of the consonant strings. The position of the "D" characters within the constant strings was counterbalanced to ensure the position of the target letter was not predictable across trials.

The character length of all stimuli ranged from 4-6 characters and stimuli length was counterbalanced across conditions (RJ: M = 4.91 characters; LS: M = 4.91 characters). The correctness of the items was counterbalanced such that 50% of the trials were correct and 50% were incorrect for both conditions. Participants indicated their responses using buttons placed in each hand, the handedness to correctness mappings were counterbalanced across participants.

Comprehension Task. The Comprehension task is a novel task developed for this study. Participants complete two types of task trials: *Python* and *Scrambled Word Reading* trials. The task was broken up into four fMRI runs each lasting approximately seven minutes. Each run was split in half by condition, and the order in which the conditions were presented (e.g., the first or second half of the run) varied across runs. Each run contained 16 trials (8 trials per condition) and the total task contained 64 trials (32 trials per condition).

In *Python* trials, participants viewed Python code and were asked to determine if the code successfully compiled to an output and what that output was. In 30% of the trials, there was an error in the presented code meaning that participants should respond using the button corresponding to "NO" to indicate that the code would not successfully compile to an output. Three types of errors were included in the Python trials: syntax errors (e.g., incorrect brackets used), type errors (e.g., treating a variable like a string), and built-in function errors (e.g., incorrectly using a dot notation function). In 30% of the trials, participants completed

comprehension checks to keep them engaged in the task (see Figure 3). These comprehension checks occurred after participants made their initial judgments about whether the code would compile and tasked participants with responding to whether a presented output correctly matched the output that they predicted. The correctness of the presented outputs was counterbalanced such that 50% of the comprehension checks were correct and 50% were incorrect. The comprehension checks were allowed to co-occur with error trials in which case the comprehension check probe would either be "error" in a correct case or a false output in an incorrect case.

In *Scrambled Word Reading* trials, participants viewed lists of words and were asked to think about the meanings of the words and respond if they knew the meanings of all the words. In 30% of the trials, there was a novel nonword (*e.g., aleebo*) in the list of words meaning that participants should respond using the button corresponding to "NO" to indicate they did not know the meanings of all the words. Nonword stimuli were taken from the Yen et al. (2019) stimuli set but did not co-occur with the nonword stimuli used in the Rhyming Judgement task. In 30% of the trials, participants completed a comprehension check after they had made their initial judgment (see Figure 3). During these checks, participants were presented with a target word and asked whether it was thematically related to the words they had seen previously. The relatedness of the target words was counterbalanced such that 50% were related and 50% were unrelated. The word lists included in each trial were developed by taking the Python stimuli *not* used in the current version of the task (i.e., for Version A use the Python stimuli from Version B) and extracting all English words. The English words included in the dataset were relatively high-frequency words (Log HAL frequency: $M = 9.82$, $SD = 2.15$; Word length: $M = 5.36$, $SD = 1.94$) for a full list of the included words and their frequencies, lengths, and occurrences see Appendix

B. Numbers were included in the word lists but all abbreviations and symbols were removed. The word list was then randomized using a random number generator and the words were organized on the screen such that the number of words per line matched the corresponding Python trial it was derived from.

The number of words included in trials ranged from 7-16 words (*M* = 11.67 words) and did not differ between task versions (Version A: *M* = 11.78 words, *SD* = 2.86; Version B: *M* = 11.56 words, *SD* = 2.07). The position where errors/nonwords and comprehension checks occurred within the stimuli set was counterbalanced such that each stimuli position contained an error/nonword twice and a comprehension check twice during the task. Errors and comprehension checks were also balanced across the task such that each grouping of eight condition trials included two trials with errors/nonwords and two trials with comprehension checks (one correct and one incorrect).

Each grouping of eight trials began with a cue presented for 3000 ms instructing the participant of which condition would be coming up next (see Figure 4). This cue was followed by a 10500 ms fixation. Each trial was presented alongside a prompt at the bottom of the screen reminding participants which button to press for “YES” and “NO”. Participants were given 12000 ms to view the stimuli and make their responses, if they responded before the time had elapsed the response prompt on the bottom of the screen would disappear but the stimuli would continue to be presented for the remaining time allotted. If the trial contained a comprehension check, participants would immediately be shown a second screen with a probe of either a code output (Python trials) or a word (Scrambled Word Reading trials) alongside a prompt at the bottom of the screen reminding participants which button to press for “YES” and “NO”. Participants were given 4500 ms to make their responses, if they responded before the duration

had elapsed the response prompt on the bottom of the screen would disappear but the probe would continue to be presented for the remaining time allotted. A 10500 ms fixation was presented between each trial. The hemodynamic response function was modeled at a variety of fixation and trial lengths, the model indicated that this timing was optimal to capture each trial's BOLD response whilst allowing for enough time for participants to complete the trials behaviorally (NeuroElf v1.1 rc2).

Procedure

Participants completed two sessions for this study. One of these sessions took place on Zoom and the other took place in person and included the MRI scan. All participants provided informed consent during their first session before completing any study measures.

In the Zoom session, participants were led through the session by a research assistant. For each task, the research assistant provided verbal instructions to the participant and then shared a link to an online task in the Zoom chat. Participants were instructed to share their screens with the research assistant throughout the session and all Zoom sessions were recorded. During the Zoom session, participants completed the following tasks in order: a demographic questionnaire, a contact information questionnaire that asked about their desire to be contacted for future study participation, the Nelson Denny Reading test, the PEAPQ, the Declarative Knowledge Python test, the Card Sorting task, the Internal Representations Questionnaire, the Forward Digit Span task, and the LEAPQ. In total the Zoom session took approximately 90 minutes to administer.

The in-person session consisted of both behavioral tasks and the MRI. Participants completed the MLAT II measure and behavioral practice versions of the MRI scanner tasks on a lab provided desktop computer. Participants also received a briefing on what to expect in the scanner and completed an MRI safety screening in accordance with the standards set forth by the

Center for Human Neuroscience at the University of Washington. The MRI portion of the session included initial calibration scans, field maps, two runs of the Phonological Localizer each lasting approximately five minutes, four runs of the Comprehension Task each lasting approximately seven minutes, and a structural T1w MPRAGE lasting approximately seven minutes. The intercom system was used to check in with participants between scans and provide instructions for the upcoming scan. In total participants spent around one hour in the scanner and the full session took approximately two hours. After the scan was finished participants completed a debriefing questionnaire, received instructions about subject payment, and were thanked for their participation.

fMRI Acquisition & Analysis

Acquisition

MRI data was collected on a Siemens Prisma 3.0 T scanner at the Center for Human Neuroscience at the University of Washington. The data were aligned to the anterior-posterior commissure. Functional data for both tasks were collected using a gradient echo-planar pulse sequence with interleaved acquisition, a multiband acceleration factor = 3, TR = 1500ms, TE = 30ms, a 74-degree flip angle, and a 216mm FOV. Each volume consisted of 48 slices with an anterior-to-posterior encoding. Each slice was 3 mm thick, with no gap between slices and an in-plane resolution of 3 x 3 mm voxels. Opposing field maps were collected in both the anterior-to-posterior and posterior-to-anterior directions to account for inhomogeneities in the magnetic field using the same pulse sequence parameters as the functional data. A T1w image was collected at the end of the session using the scan parameters from the ABCD project (Casey et al., 2018). For the T1, a Grappa acceleration factor = 2 was used and the in-plane-resolution was 1 x 1 mm voxels.

Analysis

Preprocessing. Data was preprocessed using the default configuration in fMRIprep version 23.2.1. During preprocessing, the functional data was co-registered to the T1w structural scan for each participant, slice-time corrected, corrected for susceptibility distortions using field maps, and normalized to the MNI152NLin2009cASym template. Six motion confound regressors recording head movement in the translational and rotational x, y, and z directions were extracted from the fMRIprep output and regressed out for each participant in the first-level models. Excessive motion was defined using framewise displacement with volumes containing more than 1mm of motion flagged. Runs where 5% or more of the collected volumes had 1mm or more of motion were excluded from analysis (N = 3). Participants who had two or more runs that met the excessive motion criteria were removed from further fMRI analyses (N = 1)[2]. One additional participant was removed due to a technical timing error with the scanner. The final sample of usable data for the MRI tasks included 44 participants.

General Linear Model: Original Timing Approach. First-level models for both the Phonological Localizer and the Comprehension task were computed using the General Linear Model implemented in BrainVoyager 22.4. Second-level group models were performed using the parameter estimates obtained from fitting a first-level linear model to each participant's data. Significant clusters were output using NeuroElf v1.1 rc2. For the Phonological Localizer task, three regressors were used in the model corresponding to: 1) Rhyming trials; 2) Letter Search trials; and 3) Fixation periods. The Rhyming > Letter Search contrast was used to index *phonological processing*. For the Comprehension task, three regressors were used corresponding to: 1) Python trials; 2) Scrambled Word Reading trials; and 3) Fixation periods. The regressors in

[2] The participant removed for excess motion was the same participant removed for low behavioral performance on the scanner tasks.

this model were based on the entire time window that the stimuli was presented on the screen to the participant. The Python > Fixation, Scrambled Word Reading > Fixation, and Python > Scrambled Word Reading contrasts were all examined.

General Linear Model: Response Time Approach. An additional General Linear Model was fit to the Comprehension task data using individual participant response times to the Python and Word Reading trials to define the model durations. This approach followed the same procedure as the original timing model by first fitting first-level models to subject-level data and then computing a second-level group analysis. This approach was included because behavioral analysis demonstrated that response times for the Scrambled Word Reading condition were much faster than for the Python condition (see Figure 5).

This difference in behavior between the conditions was also seen in the BOLD timecourses extracted from the fMRI data. As can be seen in Figure 6A, under the standard timing model the timecourse for the Python condition appears to both have a higher amplitude and a longer latency than the Scrambled Word Reading condition. In Figure 6B, a theoretical model of the hemodynamic response generated using NeuroElf v1.1 rc2 is plotted. This theoretical response demonstrates that when the time on task is increased the amplitude of the waveform also increases. As such, it is challenging to know if differences between the Python and Scrambled Word Reading conditions in the Original model are merely a reflection of spending more time on task in the Python condition. To control for this possibility a GLM was computed using participant response times for each trial.

Subject-Specific Region-of-Interest (ROI) Definition. Phonological processing regions were identified using the subject-specific functional localizer approach (Nieto-Castañón & Fedorenko, 2012). In this method, each participant's first-level phonological processing map

(Rhyming > Letter Search) was compared to theoretically motivated search areas and the subject's peak activity within each search area was extracted to serve as the ROI in subsequent analyses (see Figure 7).

The phonological brain regions used to define the search areas were: 1) left inferior frontal gyrus (left IFG), 2) left parietal, and the right-hemisphere homologs 3) right inferior frontal gyrus (right IFG) and 4) right parietal, as well as 5) pre supplemental motor area (preSMA). The exact search areas were created by finding the peak activation for the left inferior frontal gyrus, left parietal, and preSMA areas in the group-level GLM map for the Rhyming > Letter Search contrast and drawing a spherical search area around the peak. The right hemisphere search areas were created using a homologue of the left hemisphere coordinates. The resulting search areas are described in Table 2.

The full ROI definition procedure is depicted in Figure 8. To identify the subject-specific ROIs, each participant's first-level Rhyming > Letter Search contrast map above a *t* threshold = 2 was compared to the search areas. Within each search area a peak was identified for each participant by searching for the voxel with the highest statistical value. The voxels contiguous to the peak in all directions were surveyed and the most active voxel was selected for inclusion in the ROI. Voxels that were contiguous but not initially selected for the ROI were saved in a que and could be added to the ROI in the future if their statistical value was higher than the newly contiguous voxels. This procedure continued until the ROI contained five voxels. The same procedure was repeated for each search area such that every participant had ROIs for all five areas of interest. If a participant did not have five contiguous voxels above the threshold within the search area they did not receive a ROI for that region - this occurred only for the Right Parietal search area for two participants. The ROI definition procedure was completed using

Matlab Version 23.2. These functionally localized subject-specific ROIs were operationalized as the phonological processing system for each participant and used in subsequent analyses of the Comprehension task.

ROI Correlational Analysis. To examine how the phonological network's role in computer code comprehension varies as a function of individual differences, a ROI GLM was computed for each phonological ROI during the Comprehension task. To do so, the VOI (volume-of-interest) GLM tool in BrainVoyager 22.4 was used with the options for: separate subject predictors, individually defined ROIs, and z-transformation selected. The contrasts between the resulting beta weights: Python > Fixation, Scrambled Word Reading > Fixation, and Python > Scrambled Word Reading were correlated with behavioral measures of interest including: Python Declarative Knowledge test, Nelson Denny Reading Comprehension, Forward Digit Span, MLAT II, IRQ Visual and Verbal, and verbal bias on the Card Sorting task. These ROI GLM analyses and resulting behavioral correlations were computed using both the Original and RT based models.

Exploratory Whole-Brain Correlational Analyses. Exploratory correlational analyses were computed at the whole-brain level using the ANCOVA tool in BrainVoyager 22.4. Each of the behavioral measures of interest: Python Declarative Knowledge test, Nelson Denny Reading Comprehension, Forward Digit Span, MLAT II, IRQ Visual and Verbal, and verbal bias on the Card Sorting task, were entered separately as a covariate and correlated with the Python > Fixation and Python > Scrambled Word Reading contrasts. Correlation maps were thresholded above a correlation value of $r = 0.3$ and an extent threshold of 20 voxels. Significant clusters were output using NeuroElf v1.1 rc2 and visually inspected to ensure they fell within cortical or subcortical areas.

Results

In the sections that follow I will discuss the behavioral and neuroimaging results obtained from the present study at the group and individual-level. First, I will present results showing descriptive and correlational analyses between the behavioral measures of interest. Then I will discuss the neuroimaging results derived from the Phonological Localizer task and the Comprehension task. The discussion of the Comprehension task will be organized around topical sections that deal with: 1) the brain's response to code, 2) how specific this neural response is, 3) how time on task affects the brain's response, and 4) areas of differentiation that emerge when Python and Scrambled Word Reading trials are explicitly contrasted. Within each of these topical sections I will start by describing the group-level data, then I will discuss how individual differences in behavior modulates these responses using correlational analyses.

Behavioral Results

Descriptive statistics

Descriptive statistics for all behavioral measures of interest are provided in Table 3.

Phonological Localizer. Behavioral performance on the Phonological Localizer task was lower and more variable in the Rhyming Judgement condition (M = 0.80, SD = 0.15, *range* = 0.47 - 0.97) than in the Letter Search condition (M = 0.98, SD = 0.06, *range* = 0.59 - 1.0). A paired two-tailed t-test revealed that this difference in accuracy was statistically significant [$t(45) = 8.65$, $p < 0.001$; see Figure 9]. The Rhyming Judgement condition also had slower response times (M = 1897.04, SD = 349.53, *range* = 1111.67 - 2625.17) than the Letter Search (M = 1320.56, SD = 286.79, *range* = 923.97 - 2219.68) condition [$t(45) = 12.89$, $p < 0.001$; see Figure 9]. One participant performed below chance on the Phonological Localizer and was removed

from all subsequent neuroimaging analyses, this participant also met the criteria for excess head motion.

The lower performance on the Phonological Localizer task was partially driven by bilingual participants, the majority of whom had a logographic language as their first language (L1; see Figure 10). A mixed 2 (L1) x 2 (Condition) ANOVA demonstrated that the main effect of participant first language [$F(10, 34) = 14.54$, $p < 0.001$], the main effect of task condition [$F(1, 34) = 298.31$, $p < 0.001$], and the interaction between L1 and task condition [$F(10, 34) = 13.29$, $p < 0.001$] were all significant. This pattern of results suggests that while participants with English as their L1 had higher accuracy than participants with a non-English L1 on both types of trials, the magnitude of this performance benefit for English L1 participants was greater in the Rhyming condition.

Comprehension Task. Behavioral performance on the Comprehension task is depicted in Figure 10. Total accuracy was calculated by combining accuracies on the initial trial response with comprehension check probe accuracies. Total accuracy was higher on Scrambled Word Reading trials ($M = 0.96$, $SD = 0.07$, *range* = 0.59 - 1.0) than on Python trials ($M = 0.78$, $SD = 0.09$, *range* = 0.53-0.97), though the variability between the conditions was similar (see Figure 11A). A two-tailed paired t-test revealed that this difference in total accuracy between the task conditions was statistically significant [$t(45) = 10.72$, $p < 0.001$]. A paired two-tailed t-test revealed a similar pattern when accuracy was calculated on only the comprehension check probes (see Figure 11B), such that accuracy was higher for Scrambled Word Reading ($M = 0.89$, $SD = 0.13$, *range* = 0.50 - 1.0) than for Python ($M = 0.76$, $SD = 0.15$, *range* = 0.38 - 1.0) comprehension check probes [$t(45) = 4.38$, $p < 0.001$]. Considering that participants completed relatively few of these comprehension check probes ($N = 12$ per condition), the comprehension

check time window was not modeled in the neuroimaging analysis, and individual differences in programming skill was a key covariate in the planned analyses, I opted not to removed participants with low accuracy on the comprehension check probes, and instead focused on assessing behavioral performance using total accuracy.

Response times from the Comprehension task showcased a similar pattern (see Figure 11C). A paired two-tailed t-test demonstrated that participants responded faster on Scrambled Word Reading trials (M = 3161.88, SD = 837.39, *range* = 1598.50 – 5040.97) than they did on Python (M = 5911.46, SD = 908.53, *range* = 4248.34 – 7484.88) trials [$t(45 = 21.08, p < 0.001$]. There was also a significant difference in response time for responding to the comprehension check probes (see Figure 11D), such that responses were faster for Scrambled Word Reading probes (M = 1697.79, SD = 399.62, *range* = 1037.25 – 2673.62) than they were for Python (M = 1957.61, SD = 463.70, *range* = 1074.12 – 3063.88) probes [$t(45) = 3.05, p = 0.004$]. Together, the accuracy and response time data suggest that the Python condition was more difficult than the Scrambled Word Reading condition. This is not particularly surprising, as I intentionally recruited a wide range of Python experience levels and all participants were recruited from an English-speaking university and were therefore quite fluent in English.

Behavioral Correlational Analyses

A full table of bivariate correlations for all measures of interest are provided in Table 4.

Bivariate Correlations Between Python Measures. The majority of the measures that indexed programming skill were correlated with one another (see Figure 12). This pattern provides convergent validity that the Python condition from the MRI Comprehension task was tapping into the same variation in skill as the more established self-report and evaluative behavioral measures. Total accuracy on the Python trials from the MRI task was significantly

correlated with self-reported Python skill, as measured by the PEAQ, (r = 0.53, p < 0.001, *pfdr* < 0.001) and performance on the behavioral Python Declarative Knowledge test (r = 0.71, p < 0.001, *pfdr* < 0.001). Response times on Python trials were also significantly correlated with PEAPQ (r = -0.44, p = 0.002, *pfdr* = 0.007) and marginally correlated with performance on the Python multiple choice test (r = -0.26, p = 0.08, *pfdr* = 0.12).

Bivariate Correlations Between English Measures. The behavioral measures related to English proficiency also showed convergent validity. A full summary of the Pearson correlations between English measures are depicted in Figure 13. Total accuracy on the Scrambled Word Reading condition from the MRI task was significantly correlated with self-reported English Skill as measured by the LEAPQ (r = 0.48, p = 0.001, *pfdr* = 0.005), English reading skill as measured by the Nelson Denny Comprehension test (r = 0.40, p = 0.006, *pfdr* = 0.02), and response times on the Scrambled Word Reading trials (r = - 0.40, p = 0.006, *pfdr* = 0.02). Response times on the Scrambled Word Reading trials were also significantly correlated with the LEAPQ (r = -0.30, p = 0.045, *pfdr* = 0.09) and Nelson Denny Comprehension test (r = -0.37, p = 0.013, *pfdr* = 0.03).

Neuroimaging Results: Phonological Localizer Group-Level GLM (Rhyming > Letter Search)

The group-level Phonological Localizer task activity is depicted in Figure 14 and a full list of the observed clusters are reported in Table 5.

As anticipated, the phonological processing network, operationalized using the Rhyming > Letter Search contrast, consisted primarily of language-related brain regions including the left inferior temporal gyrus, portions of the left parietal lobule including the supramarginal gyrus, the striatum, and temporal regions including the visual word form area (VWFA). Activation in the

left inferior frontal gyrus, a region implicated in both overt and inner speech production, was particularly robust consistent with prior results using similar rhyming judgment tasks (e.g., Hoeft et al., 2007; Poldrack et al., 2001; Yen et al., 2019). Right hemisphere homologues of some of these areas including the inferior frontal gyrus, inferior parietal lobule, and striatum were also present at the group level though the size of these clusters were smaller and the statistical magnitude weaker than their left hemisphere counterparts. This pattern is consistent with visual analysis of first-level individual subject maps where a great deal of variability can be observed particularly in the right hemisphere (see Figure 15). This variability may be due to the inclusive recruitment criteria of this work and specifically to the decisions to not control for handedness or natural language background of the participants. This variability also highlights the need for analysis approaches that are sensitive to individual differences in function and anatomy such as the subject-specific functional localization approach to ROI selection employed herein.

However, it can be theoretically challenging to decide how much to constrain variability in the subject-level data. In the present study, theoretically motivated search areas were used to limit individual subject data to regions that should be related to phonological processing theoretically. These search areas were created by drawing a sphere around the peak voxel activation observed at the group-level (see Figure 7). Then, the peak activation for each individual within the search area was identified and the ROI was created using the search procedure depicted in Figure 8. Within each search area, there was variability between subjects in terms of which voxel had the highest statistical value (see Figure 16). The spread in peak activation within each search area highlights just how variable activity patterns in these areas can be across subjects and the benefit of using a functional localizer to individually define ROIs based on functional differences.

Comprehension Task: Operational Hypotheses

Analysis of the Comprehension task primarily aimed to: 1) understand if the phonological system is involved in real time computer code comprehension, and 2) gain traction on the theoretical mechanism linking the phonological system to code comprehension. Towards the first aim, I examine how the phonological system responds during code comprehension and how performance on a behavioral measure of phonological coding (MLAT II) relates to individual differences in these responses. Towards the second aim, three primary theoretical hypotheses are proposed. In this section, I will briefly summarize each of these theories and discuss the results that are expected under each theory at the operational level. It is important to note that support for these hypotheses is not mutually exclusive.

The *Lexical Access Hypothesis* suggests that the role of phonology in computer programming is largely epiphenomenal in nature and occurs as a by-product of accessing the lexical meanings of the many English words embedded in code. Under this theory, several operational hypotheses can be proposed and evaluated. First, because the English words included in the Scrambled Word Reading condition are the same as those in the Python condition (i.e., Scrambled Word Reading Version A words are the same as Python Version B words), it is expected that if lexical access is driving the relation between phonology and programming the patterns of activity at the group-level should look similar between the Python and Scrambled Word Reading conditions in phonological processing areas. Second, it is expected that if phonology is aiding lexical access individual differences in English reading skill should modulate the extent to which this relationship is observed. Therefore, it is expected that Nelson Denny Reading Comprehension scores will be correlated with activation in phonological areas in both the Python and Scrambled Word Reading conditions. Under the Dual-Route model of

lexical access (Coltheart, 1980; Coltheart et al., 2001), it is expected that the direction of these correlations with English skill will be negative, such that less-skilled readers require more phonological brain activation to access the lexical meanings of the words in both the Python and Scrambled Word Reading conditions. In contrast to the *Lexical Access Hypothesis* which suggests that the relation between programming comprehension and phonology is epiphenomenal and occurs early in processing, both of the proposed *Inner Speech Hypotheses* suggests a more central role of phonology in higher-level processing as either a means to compensatorily offload information in the face of difficulty task demands and/or a means of representing a problem space in a code concurrent with one's code-of-thought.

The *Inner Speech: Working Memory Variant Hypothesis* suggests that the phonological system is used to cope with cognitively taxing task demands. Considering the participant sample was composed of students at an English-speaking university who were all fluent in English, it was anticipated that the Python condition would be more difficult than the Scrambled Word Reading condition. The behavioral accuracy and response time results supported this prediction (see Figure 11). Therefore, under this theory, it is expected that phonological areas will be more active in the Python condition than in the Scrambled Word Reading condition at the group-level due to the Python condition having more difficult task demands. The individual differences analyses can also shed light on this theory. If the phonological system is engaged as a way to compensate in cognitively taxing contexts, it is expected that individual differences in programming skill, assessed using the behavioral Python Declarative Knowledge test, should be related to variability in phonological network activation during the Python condition. Specifically, it is predicted that less skilled programmers will find the Python condition more challenging and as a result offload information into the phonological loop, resulting in greater

recruitment of neural areas implicated in phonological processing. It is also likely that a relationship between individual differences in working memory capacity, as assessed by the Forward Digit Span, and the degree to which individuals rely on the phonological system during code comprehension will be observed under this theory. However, predicting the direction of such a relationship is challenging as higher working memory capacity could mean a higher threshold for resources to be taxed, resulting in a negative correlation with phonological activity during programming, or higher working memory capacity could lead to more brute force maintenance, resulting in a positive correlation with phonological activity during code comprehension. While the *Inner Speech: Working Memory Variant Hypothesis* centers around the idea that inner speech is used to compensate in the face of task demands, the *Inner Speech: Code-of-Thought Variant Hypothesis* suggests that inner speech is used selectively by some participants with more verbal strategic habits.

The *Inner Speech: Code-of-Thought Variant Hypothesis* suggests that phonological processing may be used as a strategy during code comprehension selectively for individuals with a propensity for representing information using a verbal code. Unlike the *Inner Speech: Working Memory Variant Hypothesis*, the *Inner Speech: Code-of-Thought Variant Hypothesis* suggests that inner speech is about strategy not a response to cognitive constraints. The degree to which code-of-thought influences performance should relate to how much opportunity there is for participants to employ diverse strategies. The Python condition is more complex and has more opportunities for participants to represent information using visual or verbal codes, whereas the Scrambled Word Reading condition more stringently encourages a verbal processing strategy. As such, it is expected that participants with more verbal codes-of-thought, as indexed by a higher score on the IRQ-Verbal and/or a greater verbal bias score on the Card Sorting task, will rely

more on phonological brain areas during code comprehension. Conversely, participants with more visual codes-of-thought, as indexed by a higher score on the IRQ-Visual and/or a lower verbal bias score on the Card Sorting task, should rely on phonological brain areas less during code comprehension and may instead recruit more visuospatial processing resources.

The Brain on Code: What Happens in the Brain During Computer Code Comprehension?

In this section, I will explore what the results of the present study showcase about the areas of the brain involved in computer code comprehension at the group-level and how variability in these brain responses relate to individual differences in the behavioral measures of interest.

Group-Level General Linear Model Results (Python > Fixation): Original GLM

The group-level statistical map depicting neural activation during programming comprehension is depicted in Figure 17. The resulting network of regions included robust activation of the frontoparietal network consistent with patterns of activation observed in tasks that elicit the multiple-demand network (*e.g.,* Camilleri et al., 2018; Duncan, 2010), and with prior studies on computer programming (e.g., Endres et al., 2021; Hishikawa et al., 2023, Ikutani et al., 2021; Ivanova et al., 2020; Liu et al., 2020; Pietek et al., 2021; Siegmund et al., 2014; Siegmund et al., 2017; Xu et al., 2021). A full list of activation clusters observed during code comprehension are provided in Table 6.[3] Critical to the question of how the phonological system is involved in code comprehension, code comprehension activity at the group-level overlapped with the results of the Phonological Localizer in many key areas (see Figure 18). The overlap in

[3] Note that the clusters provided in the table are thresholded at a more stringent level of $t = 5$ (FDR corrected $p < 0.001$) to reduce all regions being connected to one another. The statistical map of this activity in Figure 16 is thresholded at the standard level of $t = 3$ (FDR corrected $p < 0.05$).

neural representations between these two tasks suggests that the neural mechanisms underpinning phonological processing and code comprehension may be shared to some degree.

Individual-Level Region-of-Interest Correlations (Python > Fixation): Original GLM.

Correlations between code comprehension activity in the subject-specific phonological ROIs and the behavioral measures of interest are reported in Table 7A.

To assess if phonological coding skill related to real-time engagement of the phonological network during code comprehension, performance on the MLAT II was correlated with code comprehension activity in the phonological ROIs. A negative trend was observed between performance on the MLAT II and code comprehension activity in the left parietal ROI (r = -0.29, p = 0.05; see Figure 19). This result is consistent with the idea that individuals who are more skilled at phonological coding behaviorally have greater neural efficiency in an area implicated in generating early-stage phonological codes (*e.g.*, Junker et al., 2023).

There were no significant correlations between Nelson Denny Reading Comprehension and individual differences in code comprehension activity in phonological ROIs (ps > 0.10). This result is inconsistent with the predictions made by the *Lexical Access Hypothesis*.

Support for the *Inner Speech: Working Memory Variant Hypothesis* was mixed. Inconsistent with this hypothesis, no significant correlations were observed between performance on the Python Declarative Knowledge test and code comprehension activity in any of the phonological ROIs (ps > 0.10). However, a trending positive correlation was observed between Forward Digit Span and code comprehension activity in the right parietal ROI (r = 0.28, p = 0.07; see Figure 20). This result is consistent with the idea that higher working memory individuals may store more informational units in working memory to aid comprehension. However, it is noteworthy that the correlations in all other ROIs - though not significant - were in

the negative direction. While some participants did engage portions of the right parietal lobe during the Phonological Localizer - this was the least consistently activated region. This raises the possibility that the right parietal ROI may be playing a non-phonological role in its relation to working memory capacity and may be aiding maintenance of the probe-related information more so than comprehension per se.

The observed correlational analyses also provided partial support for the *Inner Speech: Code-of-Thought Hypothesis*. Negative correlations were observed between IRQ-Visual and code comprehension activation in the preSMA (r = -0.33, p = 0.03), left inferior frontal gyrus (r = -0.26, p = 0.09), and left parietal (r = -0.33, p = 0.03) ROIs (see Figure 21). This negative relationship suggests that individuals with more visual codes-of-thought are using phonological areas to support code comprehension *less* than individuals with less of a visual processing style. This is consistent with the hypothesis that differences in strategy may influence the extent to which individuals engage in inner speech processes during Python trials. However, the expected positive correlation between the behavioral measures of verbal code-of-thought – IRQ-Verbal and verbal bias on the Card Sorting task – were not significantly correlated with code comprehension activity in any of the phonological ROIs (ps > 0.10).

Exploratory Whole-Brain Correlations (Python > Fixation): Original GLM.

To examine the relationship between the behavioral measures of interest and individual differences in code comprehension activity, exploratory correlational analyses were computed at the whole-brain level. These results are reported in full in Table 8.

Whole-brain correlations between MLAT II and code comprehension activity provided converging support for the pattern seen in the ROI results. Namely, MLAT II was negatively correlated with code comprehension activity in the left supramarginal gyrus (BA 40), the portion

of the inferior parietal lobule most implicated in generating early phonological codes (e.g., Graves et al., 2023; Junker et al., 2023; Stoeckel et al., 2009). This result further supports the idea that individuals who perform well on phonological coding measures behaviorally are more efficient at generating phonological codes. Additionally, MLAT II was positively correlated with code comprehension activity in the bilateral angular gyrus (BA 39) suggesting that individuals with better phonological coding skills engage this area more. The angular gyrus is also part of the inferior parietal lobule but is commonly ascribed a semantic function (e.g., Binder et al., 2009; Price, 2010), and specifically is implicated in linking semantic knowledge to phonological codes (Junker et al., 2023). This result may reflect that individuals who can more efficiently generate phonological codes are more likely to use these codes and link them to semantic meanings to aid further processing.

Whole-brain analysis of the relationship between Nelson Denny Comprehension and code comprehension activity revealed positive correlations in a network of semantic related regions including bilateral temporal regions and the anterior-ventral portion of the left inferior frontal gyrus. This pattern of activity is consistent with the idea that individuals who are more skilled in English pay more attention to the semantic meanings of the words embedded in the code. With regard to the *Lexical Access Hypothesis*, this pattern of results is consistent with the ROI results showing no relation between English skill and code comprehension activity in regions implicated in phonological processing.

Correlational analyses between performance on the Python Declarative Knowledge test and code comprehension activity at the whole-brain level revealed a set of negative correlations in frontostriatal areas, and positive correlations in visual and salience-based areas. This pattern of results, consistent with what was seen in the ROIs, did not support the *Inner Speech: Working*

Memory Variant Hypothesis. However, the primarily negative correlations between programming skill and neural activation in cognitive control areas is consistent with the more general theoretical notion that greater skill should lead to a reduction in cognitive load as evidenced by a more efficient neural network. The whole-brain correlations with Forward Digit Span also failed to support the *Inner Speech: Working Memory Variant Hypothesis* at the whole-brain level.

The whole-brain analyses provided further support for the *Inner Speech: Code-of-Thought Hypothesis*. Consistent with the results observed in the ROI analyses, individuals who scored higher on the IRQ-visual showed negative correlations with phonological processing areas including the left inferior and supramarginal gyri. The whole-brain analyses were able to further add to this picture by showing that IRQ-visual was also positively correlated with activity in bilateral portions of the visual cortex. These results suggest that individuals with more visual strategic habits not only engage phonologically related areas *less* when they read code, but they also engage visual areas *more*.

The whole-brain correlational results for IRQ-Verbal supports the notion that individual differences in strategic habits influence the neural network supporting code comprehension. IRQ-verbal was positively correlated at the whole-brain level with programming activity in the left angular gyrus, a region implicated in integrating phonological and semantic codes. However, the whole-brain correlational analyses with verbal bias on the Card Sorting task did not find significant correlations in phonological areas. Together, this pattern of results is consistent with the idea that individuals may represent programming information using more verbal or more visual neural pathways based on differences in the way that they prefer to represent information.

Interim Summary: The Brain on Code.

In this section I have explored what the neural network involved in the Python condition looks like and how variation in this network at the individual level can be used to gain traction on the theoretical hypotheses of interest. The key takeaways from this section are as follows. First, programming activity at the group-level was elicited in the expected frontoparietal network and showed considerable overlap with the group-level phonological activity elicited from the Phonological Localizer task. Second, behavioral phonological coding skills, as indexed by the MLAT II, were associated with both more efficient processing in brain areas associated with generating phonological codes and greater recruitment of regions that integrate phonological and semantic information. This result supports the idea that behavioral phonological skills influence the way the phonological system is deployed in real-time code comprehension. Third, neither individual differences in English nor Python skill were related to the degree to which phonological areas were active during code comprehension, this pattern of results is inconsistent with the predictions generated by the *Lexical Access Hypothesis* and *Inner Speech: Working Memory Variant Hypothesis,* respectively. Finally, both the ROI and whole-brain results provided support for the *Inner Speech: Code-of-Thought Hypothesis*. During code comprehension, individuals with more visual codes-of-thought engaged phonological processing areas less and visual areas more, whereas individuals with more verbal codes-of-thought showed greater activation in areas that link phonological and semantic information.

The Brain on Words: Is the Observed Neural Network Specific to Code, or Also Recruited During Word Reading?

In this section, I will explore if the brain regions involved in code comprehension are specific to a code neural network or whether these resources jointly support word reading. Additionally, I examine if individual differences in behavior relate to variability in Scrambled

Word Reading activity in similar or distinct ways from the individual differences correlations observed during the Python condition.

Group-Level General Linear Model Results (Scrambled Word Reading > Fixation)

The Scrambled Word Reading condition revealed activation in the frontoparietal task positive network at the group-level (see Figure 22 and Table 9). The statistical map of word reading activity looked very similar to that obtained during code comprehension, though the magnitude of the statistical map was more robust in the code comprehension condition particularly in regions like the inferior frontal gyrus.

Like in code comprehension, word reading activity also overlapped substantially with the group-level results obtained from the Phonological Localizer task (see Figure 23). However, this overlap was less robust compared to the overlap observed with code comprehension activity. These results could indicate support for the *Inner Speech: Working Memory Variant Hypothesis* whereby participants engage phonological areas more when task demands get more difficult, as the Python condition is more cognitively taxing than the Scrambled Word Reading condition. However, these results could also be a consequence of the Scrambled Word Reading condition regressor having a poorer fit than the Python condition regressor due to the differences in time on task between the two conditions (see Figures 5 and 6). I explore this possibility in more depth in a subsequent analysis where I model the GLM for both conditions based on individual participant response times at the trial level.

Individual-Level Region-of-Interest Correlations (Scrambled Word Reading > Fixation)

Several of the relationships observed between phonological ROI activity and behavior for Python were also observed for Scrambled Word Reading. For a full report of the resulting correlational analyses see Table 7B. Like in the Python condition, MLAT II negatively trended

with word reading activity in the left inferior frontal (r = -0.29, p = 0.06) and left parietal (r = -0.29, p = 0.06) ROIs (see Figure 24). This pattern suggests that better phonological skills were associated with greater neural efficiency in phonological processing regions across both the Python and Scrambled Word Reading conditions.

Individual differences in working memory capacity also showed a similar correlational pattern with ROI activity in both conditions, such that individuals with greater Forward Digit Span scores showed greater word reading activity in the right parietal ROI (r = 0.31, p = 0.04; see Figure 25). This is consistent with the location and direction observed in the Python correlational analyses. This result may be indicative of higher capacity individuals holding more information in working memory because they have the ability to do so. In the case of the Scrambled Word Reading condition, this may have meant that higher working memory capacity individuals were attempting to hold the words themselves in working memory to compare the comprehension check probes against, rather than extracting a semantic gist. While the correlations observed for between IRQ-Visual and phonological activity in the Python condition did not reach significance for the Scrambled Word Reading condition, they were directionally consistent with the correlations observed for Python.

The word reading ROI analyses also revealed several correlations which emerged selectively for the Scrambled Word Reading, but not Python, conditions. Individual differences in Nelson Denny Reading Comprehension scores were negatively correlated with word reading activity in the left inferior frontal gyrus (r = -0.31, p = 0.04), left parietal (r = -0.48, p = 0.001), right parietal (r = -0.35, p = 0.02), and preSMA (r = -0.37, p = 0.01) ROIs (see Figure 26). These results are consistent with the idea that more skilled readers may need to engage phonological areas less when they read words. While this pattern of result makes sense theoretically, poor fit

of the Scrambled Word Reading regressor in the GLM model may also be a factor driving these results. Behavioral analysis showed that Scrambled Word Reading response times were negatively correlated with Nelson Denny Reading Comprehension (r = -0.37, p = 0.01; see Figure 13 and Table 4), meaning that issues of poor regressor fit for the Scrambled Word Reading condition likely disproportionately affected more skilled readers. If this is the case, it is anticipated that the correlations between Nelson Denny Reading Comprehension and phonological activity for the Scrambled Word Reading condition will be reduced under the response-time based GLM.

Another novel pattern that appeared in the Scrambled Word Reading condition analyses was that performance on the Python Declarative Knowledge test trended negatively with activity in the left (r = 0.29, p = 0.06) and right (r = -0.26, p = 0.09) parietal ROIs (see Figure 27). Considering that the words used in the Scrambled Word Reading condition were all derived from homologous Python trials, this result may reflect better programmers being more attune to these code related words. In particular, more skilled programmers may be better able to filter out common Python words (e.g., if, for, else, print) that are unlikely to relate to the semantic comprehension check probes present on Scrambled Word Reading trials. If so, this reduction in cognitive load may explain the corresponding decrease in parietal activity seen for more skilled programmers during the Scrambled Word Reading condition.

Interim Summary: The Brain on Words

In this section I have examined whether the neural network elicited during the Python condition is specific to code or whether similar results are observed when participants read words in the Scrambled Word Reading condition. The group-level statistical map for the Scrambled Word Reading condition involved a similar network of regions as the Python condition and

showed overlap in similar areas with the Phonological Localizer task, though the magnitude of these effects was more robust for Python. This result supports the theory that when cognitive demands increase more neural resources - including those in phonological areas - are brought online to manage task difficulty. However, this result should be interpreted with caution as there were large differences in response times between the Python and Scrambled Word Reading conditions, which may have led to the Scrambled Word Reading condition being disproportionately poorly captured by the Original GLM. This possibility is discussed in greater depth in the next section. At the individual-level, many of the correlations observed between behavior and activation in the phonological ROIs during the Python condition were also present in the Scrambled Word Reading condition to some degree. This suggests that the relation between activation in phonological areas during code comprehension may generalize more broadly to other domains including word reading.

What's Time Got to Do With It? Comparing the Original and Response Time General Linear Models

One of the challenges of designing the Comprehension task employed herein was how to contend with differences in the difficulty of the task conditions. As is reflected in the behavioral results, the Python condition was more difficult and took longer than the Scrambled Word Reading condition (see Figure 5 and Figure 11). While this difference in difficulty between the conditions has interesting implications for testing the *Inner Speech: Working Memory Variant Hypothesis*, it also poses methodological challenges in properly modeling the resulting neural responses (see Figure 6). In the Original GLM, both the Python and Scrambled Word Reading conditions were modeled using the same fixed time window that the stimuli was presented for (i.e., 12s), this allowed the difficulty of the task conditions to vary freely. However, response

times on the trials were much shorter than 12 seconds in both conditions, which may have led the Original GLM to poorly fit the data (see Figure 5 and Table 1). While this may have been an issue in both conditions, it likely disproportionately affected the Scrambled Word Reading condition where response times were significantly faster than in the Python condition (see Figures 5 and 11C). Fitting the GLM to participant response times for every trial alleviates some of these concerns. In this section I explore how the results change when a GLM that is specific to the response time for each task trial is used.

Group-Level General Linear Model Results: Response Time Model (Python > Fixation)

The code comprehension activity obtained from the RT GLM looked very similar to that of the Original model activating the expected frontoparietal network (see Figure 28). For a full list of significant clusters active in the response time model see Table 10.

Group-Level General Linear Model Results: Response Time Model (Scrambled Word Reading > Fixation)

As expected, under the RT model the activation pattern observed for the Scrambled Word Reading condition was more robust at the group-level than the response observed under the Original GLM (see Figure 29). For a full list of significant regions active in the response time model for the Scrambled Word Reading condition see Table 11.

A condition-level comparison of the activation maps generated by the RT model is depicted in Figure 30. Using the RT based GLM did reduce the differences between the condition maps to some degree. However, there were still some notable differences between task conditions, with activation being more robust for the Python condition in left frontoparietal and temporal areas. These differences are explored further by directly contrasting the conditions under both the RT and Original models in an upcoming section.

Individual-Level Region-of-Interest Correlations: Response Time Model (Python > Fixation)

Compared to the Original model, the RT model ROI correlations were relatively similar in location and direction, though there were some shifts in the magnitudes of the correlations (see Table 12A). The negative trend seen between MLAT II and code comprehension activity in the left parietal under the Original model ($r = 0.29$, $p = 0.053$; see Figure 19) was also observed in the RT model at a similar magnitude ($r = 0.26$, $p = 0.095$; see Figure 31). This is consistent with the interpretation that individuals with better behavioral phonological coding skills show greater efficiency in left parietal areas implicated in generating phonological codes. Also consistent with what was observed under the Original model, neither Nelson Denny Comprehension nor performance on the Python Declarative Knowledge test correlated significantly with programming activity in any of the phonological ROIs (ps > 0.1) under the RT model. Taken together, correlations between phonological ROI activity during the Python condition and performance on the MLAT II, Nelson Denny Comprehension test, and the Python Declarative Knowledge test were relatively similar when the Original versus RT models were used.

However, for other behavioral measures there were magnitude differences in the ROI correlations between the two models. Under the RT model, the negative correlation between Forward Digit Span and code comprehension activity in left parietal strengthened to reach statistical significance ($r = -0.38$, $p = 0.01$; see Figure 32) relative to the weaker but directionally consistent pattern observed under the Original model ($r = -0.16$, $p = 0.31$; see Figure 20). This result is consistent with the prediction made by the *Inner Speech: Working Memory Variant Hypothesis* that lower working memory capacity individuals are more likely to engage inner speech mechanisms to cope with challenging task demands. Conversely, the positive trend observed in the Original model between Forward Digit Span and code comprehension activity in

the right parietal ROI (r = 0.28, p = 0.07; see Figure 20) weakened in the RT model (r = 0.12, p = 0.12; see Figure 32). This is likely a reflection of how these regions are differentially recruited during the early comprehension and later maintenance phases of the task.

Likewise, magnitude shifts between the models for the correlational analyses involving the code-of-thought measures were observed. Under the RT model, the magnitude of the negative correlations between IRQ-Visual and code comprehension activity weakened (preSMA: r = -0.28, p = 0.07; left IFG: r = -0.22, p = 0.16; left parietal: r = -0.09, p = 0.55; see Figure 33) relative to those observed under the Original model (preSMA: r = -0.33, p = 0.03; left IFG: r = -0.26, p = 0.09; left parietal: -0.33, p = 0.03; see Figure 21).

For the verbal bias metric derived from the Card Sorting task this pattern reversed, such that there was a stronger positive correlation between verbal attentional bias and code comprehension activity in the preSMA and left parietal ROIs for the RT (preSMA: r = 0.26, p = 0.08; left parietal: r = 0.31, p = 0.04; see Figure 34) relative to the Original (preSMA: r = 0.19, p = 0.22; left parietal: r = 0.23, p = 0.13; see Table 7A) GLM. While some changes emerged under the RT model, the overall pattern of results still supported the *Inner Speech: Code-of-Thought Variant Hypothesis*.

Individual-Level Region-of-Interest Correlations: Response Time Model (Scrambled Word Reading > Fixation)

Accounting for differences in response times reduced the magnitude of the correlations observed between behavior and word reading activity in the phonological ROIs. However, the direction of the correlations remained consistent across models (see Table 12B). This pattern was exemplified in the correlation between Nelson Denny Comprehension and word reading activity. While under the Original model, there was a negative correlation between Nelson Denny

Comprehension and word reading activity in most of the phonological ROIs (see Figure 26) this relationship attenuated in all ROIs under the RT model and only remained significant in left parietal ROI (r = -0.34, p = 0.02; see Figure 35). This result is consistent with concern that the Scrambled Word Reading condition was disproportionally underfit by the Original model due to the difference in response times between the conditions (see Figure 5). The reduction in strength for the correlations involving Nelson Denny Comprehension is consistent with the idea that model fit issues were especially problematic for more skilled readers. Behavioral analysis demonstrating that Nelson Denny Comprehension was significantly correlated with response times on Scrambled Word Reading trials lends further credibility to this idea (see Figure 13 and Table 4).

The negative trend observed in the Original model between performance on the Python Declarative Knowledge test and word reading activity in the right parietal ROI (r = -0.26, p = 0.09; see Figure 27) strengthened slightly in the RT model (r = -0.32, p = 0.04; see Figure 36). All other correlations that were significant or trending with word reading activity under the Original model were reduced in magnitude to become non-significant under the RT model.

Exploratory Whole-Brain Correlations (Python > Fixation): Response Time Model

Exploratory whole-brain correlational analyses between code comprehension activity and the behavioral measures of interest were also conducted using the RT model (see Table 13). Overall, compared to the correlations detected under the Original model, the RT model produced more robust correlations with virtually all of the behavioral measures examined, though the general patterns of these correlations were relatively consistent across models. This provides converging support for the idea that the RT model better captures the true nature of the underlying neural responses generated during the Comprehension task (see Table 14).

In the RT model exploratory whole-brain correlations, MLAT II was negatively correlated with a distributed network of regions including a large cluster in the left supramarginal gyrus consistent with the negative trend observed between the left parietal ROI and MLAT II in the ROI analysis. MLAT II was also positively correlated with the bilateral angular gyrus, the left precuneus, and the left superior frontal gyrus. These results are consistent with the pattern observed in the Original GLM and suggest diverging roles for the supramarginal gyrus and angular gyrus within the inferior parietal lobule. The relation of these areas to individual differences on the MLAT II, suggests that behavioral phonological skills are related to the way the phonological system is deployed during code comprehension.

Nelson Denny Reading Comprehension was positively correlated with areas related to semantic retrieval in bilateral temporal areas. These results are consistent with the idea that more skilled readers utilize the semantic meanings of the words in the code to aid programming comprehension. In the RT model, correlations between phonological brain areas and Nelson Denny Comprehension were not observed at the whole-brain level. This is consistent with the ROI results obtained under both models and the whole-brain results from the Original GLM exploratory analysis. Together, the results of these analyses are inconsistent with the predictions generated by the *Lexical Access Hypothesis*.

Conversely, the whole-brain correlations did provide some support for the predictions of the *Inner Speech: Working Memory Variant Hypothesis*. Individual differences in working memory capacity, as indexed by performance on the Forward Digit Span task, was negatively correlated with a network of regions including areas like the left inferior frontal gyrus and supplementary motor areas that are involved in the planned articulation of inner speech supported by the phonological loop. The direction of these correlations is consistent with the idea

that higher working memory capacity individuals need to offload information to the phonological loop less and are more efficient in their abilities to comprehend code. This is consistent with the hypothesis that having a larger cognitive capacity should reduce the need for inner speech and correspondingly for recruitment of the phonological system. On the other hand, inconsistent with the *Inner Speech: Working Memory Variant Hypothesis*, individual differences in programming skill, as indexed by performance on the Python Declarative Knowledge test, was not related to the recruitment of the phonological network. The correlations with programming skill under the RT model demonstrated the same general pattern as those obtained from the Original model. Namely, that better performance on the Python Declarative Knowledge test was primarily negatively correlated with activation is a distributed network of bilateral frontostriatial areas and positively correlated in a small number of clusters primarily in visual salience areas. Compared to the Original model, the RT model correlations with Python skill were more robust in both the size of the clusters and the number of regions implicated but still did not include phonological areas. Taken together, the correlational analyses involving Forward Digit Span and the Python Declarative Knowledge test provided mixed support of the *Inner Speech: Working Memory Variant Hypothesis* under both the Original and RT models.

The exploratory whole-brain analyses using the RT model also found support for the *Inner Speech: Code-of-Thought Hypothesis*. Verbal bias on the Card Sorting task was positively correlated with activation in phonologically relevant regions including the left inferior frontal gyrus and the left supramarginal gyrus, as well as with areas implicated in semantic retrieval and language processing more generally like the left angular gyrus and bilateral portions of the temporal lobe. These results suggest that individuals with a greater attentional bias towards verbal information tend to recruit the phonological network more than individuals with a greater

bias towards visual information. This interpretation was also supported by the correlational results with self-reported code-of-thought. IRQ-Verbal was positively correlated with a network of regions that included semantic areas in the left angular gyrus and left temporal lobe and motor planning areas like preSMA. IRQ-Visual was negatively correlated with a primarily left-lateralized network that included the left supramarginal gyrus. However, IRQ-Visual was also positively correlated with activation in bilateral visual cortical areas. This result is consistent with the idea that individuals with more visual codes-of-thought utilize phonological areas *less,* and instead use visual areas *more*. Taken together, both the results of the RT and Original models support the *Inner Speech: Code-of-Thought Hypothesis.*

Interim Summary: The Role of Model Timing

The results reported in this section explore the role that time on task can have at both a methodological and theoretical level. From a methods perspective, thinking about the time participants spent on the task had implications for how well the proposed models fit the neural data. Participants performed faster on Scrambled Word Reading trials and this difference in speed meant that the associated neural response for the Scrambled Word Reading trials was likely not sustained throughout the full time window modeled in the Original GLM. This was also a problem to a lesser extent in the Python condition, as response times were still considerably shorter on average than the time window used in the Original GLM. As such, modeling the GLM to trial level response times improved the fit of the GLM for both of the Comprehension task conditions (see Table 14).

Theoretically, the difference between the models also provides a coarse insight into differences in the underlying cognitive processes occurring during each time window. Under the RT model, the truncated time window more selectively captures the response associated with

initial comprehension and the related decision as to whether: 1) the code compiles to an output (i.e., Python trials), or 2) the participant knows the meanings of all the words (i.e., Scrambled Word Reading trials). Conversely, in the Original model the longer time window includes neural activity associated with these processes but also activity related to maintaining 1) the code output (i.e., Python trials), or 2) the semantic gist of the word list (i.e., Scrambled Word Reading trials), in preparation for a potential comprehension check probe. As such, correlations that strengthen under the RT model likely reflect cognitive skills that are more closely tied to the earlier comprehension process whereas those that weaken under the RT model may be more involved in supporting later task-specific maintenance demands.

Taken together, while the individual differences correlations with behavior did vary in magnitude to some extent between the Original and RT model, the overall directionality and topography of these correlations remained largely consistent. The greatest difference between the model results came when the differential activity between the Python and Scrambled Word Reading conditions were directly compared, this is discussed at length in the next section.

Unpacking Functional Specificity: Understanding Points of Divergence Between the Code Comprehension and Word Reading Neural Networks (Python > Scrambled Word Reading)

In this section I examine the areas of differentiation that emerge when the Python and Scrambled Word Reading conditions are compared directly at the group-level, and how individual differences in the magnitude of difference between conditions relates to variation in behavior. The group-level GLM and ROI analyses are reported in this section based on both the Original and RT GLMs. Exploratory whole-brain analyses are reported selectively for the RT model, as it provides a better direct comparison of the two conditions.

Group-Level General Linear Model Results: Original Model (Python > Scrambled Word Reading)

The group-level difference between the Python and Scrambled Word Reading conditions is depicted in Figure 37. Under the Original GLM, the resulting statistical map included greater responses for the Python condition in a network of frontoparietal and frontotemporal regions (see Table 15 for a full list of significant clusters). While activity differences were bilateral to some degree, the differential response was more robust in the left-hemisphere. Moreover, the areas that were significant in the differential map were also present in both the code comprehension (Python > Fixation) and word reading (Scrambled Word Reading > Fixation) statistical maps. This suggests that the differential activity is not a result of the Python condition recruiting unique areas not used in the Scrambled Word Reading condition, but rather that there are differences in the magnitude of recruitment of these areas between conditions. This may also be a product of the Python condition having longer response times on average and, as a result, have a greater neural response in the Original GLM based on differences in model fit between the conditions as described in the previous section. Examining the differential activity in the RT model will shed further light on this possibility.

Group-Level General Linear Model Results: Response Time Model (Python > Scrambled Word Reading)

Under the RT model, the difference in the statistical maps between the Python and Scrambled Word Reading conditions was substantially reduced (see Figure 37; for a full list of significant clusters see Table 16). The regions that continued to show significant differentiation between conditions under the RT model included left lateralized frontoparietal areas, the left striatum, and bilateral portions of the fusiform gyri and precuneus. The reduction in differential

activity when response times were accounted for is consistent with the ideas discussed in the prior section on the role of timing. The overlap between the differential activity in the Comprehension task and the Phonological Localizer is depicted in Figure 38. While under the Original model there was substantial overlap between differential activity on the Comprehension task and the Phonological Localizer, this overlap was reduced under the RT model. However, even under the RT model there was some overlap between the MRI tasks in left inferior frontal and left parietal areas.

Individual-Level Region-of-Interest Group-Level GLM: Original Model

Group-level comparisons of the ROI activity obtained during the Python and Scrambled Word Reading conditions are shown in Figure 39 for the Original GLM. Paired t-test comparisons demonstrated significantly greater activity for Python than Scrambled Word Reading in all ROIs [left IFG: $t(43) = 8.61$, $p < 0.001$; left parietal: $t(43) = 8.36$, $p < 0.001$; right IFG: $t(43) = 6.30$, $p < 0.001$; right parietal: $t(41) = 8.33$, $p < 0.001$[4]; preSMA: $t(43) = 6.34$, $p < 0.001$]. This difference in the ROI activity at the group-level is consistent with the patterns observed in the whole-brain GLM and suggests that, under the Original model, there was greater activity in phonological ROIs during the Python condition than the Scrambled Word Reading condition. This is consistent with the *Inner Speech: Working Memory Variant Hypothesis*'s prediction that phonological areas are recruited to a greater extent when task demands are more challenging.

Individual-Level Region-of-Interest Correlations: Original Model (Python > Scrambled Word Reading)

[4] Note that two subjects did not have a peak that met the threshold criteria in the Phonological Localizer with the Right Parietal search area. As such, these participants did not have a Right Parietal ROI in any of the ROI analyses hence the smaller degrees of freedom in this analysis

The magnitude of the differential activity between the Python and Scrambled Word Reading conditions was correlated under the Original model with individual differences in the behavioral measures of interest (see Table 7C). More skilled readers, as evidenced by higher Nelson Denny Comprehension scores, had greater neural differentiation between the Python and Scrambled Word Reading conditions in the preSMA ($r = 0.32$, $p = 0.03$), left parietal ($r = 0.45$, $p = 0.002$), and right parietal ($r = 0.31$, $p = 0.04$) ROIs (see Figure 40). This result was likely driven by the relationship between English skill and the Scrambled Word Reading condition, as significant negative correlations were observed between Nelson Denny Comprehension and word reading activity (Scrambled Word Reading > Fixation; see Table 7B) but not between Nelson Denny Comprehension and code comprehension activity (Python > Fixation; see Table 7A). As discussed previously, the observed negative correlations between English reading skill and Scrambled Word Reading activity likely reflected greater neural efficiency during the word reading process consistent with the idea that more skilled readers could more automatically access the meanings of the words.

Differential activity between the conditions was also positively correlated with Python skill (see Table 7C). Specifically, it was observed that individuals who performed better on the Python Declarative Knowledge test showed a trend towards having greater differential activity in the left parietal ROI ($r = 0.28$, $p = 0.06$; see Figure 41). However, this may have been partially driven by an outlier circled in red in Figure 41. Similarly to the correlations observed with Nelson Denny Comprehension, the differential activity correlation with Python Declarative Knowledge test performance appeared to be driven by the Scrambled Word Reading condition (see Table 7B vs Table 7A). The working interpretation of this negative relationship between coding skill and word reading activity in parietal areas is that more skilled programmers have

greater familiarity with the code related words in the Scrambled Word Reading condition and can therefore more effectively filter out some of the common filler words (e.g., if, for, print) that are unlikely to relate to the comprehension check probes.

Finally, a significant negative correlation was observed between differential activity and Forward Digit Span in the left inferior frontal gyrus ROI (r = -0.33, p = 0.03; see Figure 42 and Table 7C). This correlation did not seem to be directly driven by either the Python nor the Scrambled Word Reading condition (see Table 7A and 7B). Left inferior frontal gyrus activation was only significantly related to working memory capacity when the difference between the Python and Scrambled Word Conditions was considered. This pattern of results is consistent with the *Inner Speech: Working Memory Variant Hypothesis* whereby individuals with lower working memory capacity, as evidenced by lower span scores, may need to engage compensatory phonological loop mechanisms to a greater degree than higher-capacity individuals. Considering that the Python condition was significantly more difficult than the Scrambled Word Reading condition, lower working memory capacity individuals should need to engage compensatory phonological loop resources more in the Python condition relative to the Scrambled Word Reading condition. The location of this correlation is also noteworthy. Theoretically, the left inferior frontal gyrus is a key region implicated in generating internal articulations (*e.g.*, Junker et al., 2023; Mathur et al., 2020). Additionally, out of all the ROIs derived from the phonological localizer, the left inferior frontal gyrus ROI was both the most robust area activated in the localizer at the group level (see Figure 14) and the area most commonly activated across individuals (see Figure 15). Taken together, this pattern of results suggests that when cognitive demands are taxed, such as in the more difficult Python condition,

lower working memory capacity individuals bring inner speech mechanisms online to offload some of the cognitive demands associated with the task.

Individual-Level Region-of-Interest Group-Level GLM: Response Time Model

Group-level comparisons of the ROI activity obtained under the RT model during the Python and Scrambled Word Reading conditions are shown in Figure 43. Compared to the Original model, the magnitude of the differences between the Python and Scrambled Word Reading conditions was reduced for all ROIs under the RT model. Despite this reduction in differential magnitude, paired t-test comparisons still revealed significantly greater activity for Python than Scrambled Word Reading in the left parietal [$t(43) = 2.21$, $p = 0.03$] and right parietal [$t(41) = 2.10$, $p = 0.04$] ROIs. For the left inferior frontal gyrus [$t(43) = 1.46$, $p = 0.15$], and preSMA [$t(43) = -0.30$, $p = 0.76$] ROIs the difference between the Python and Scrambled Word Reading conditions attenuated to no longer reach significance though the direction of the effect remained consistent with the Original model. For right inferior frontal gyrus [$t(43) = -2.00$, $p = 0.052$], the direction of the effect reversed from the Original model result such that Scrambled Word Reading activity trended towards being greater than Python activity, though this difference did not reach statistical significance. Taken together, these results suggest that even when response times are accounted for, the Python condition recruits more phonological activation in parietal areas than the Scrambled Word Reading condition consistent with the predictions made by the *Inner Speech: Working Memory Variant Hypothesis*.

Individual-Level Region-of-Interest Correlations: Response Time Model (Python > Scrambled Word Reading)

The correlations between differential activity in the phonological ROIs and behavior under the RT model are reported in Table 12C. Relative to the results obtained from the Original

model, the correlations between Nelson Denny Comprehension and differential activity were reduced in the RT model and only remained significant in the left parietal ($r = 0.33$, $p = 0.03$) ROI (see Figure 44). This reduction in correlation strength replicated what was seen in the comparison of the Original and RT model correlations for the word reading activity condition (Scrambled Word Reading > Fixation) alone (see Table 12B). This result provides converging support for the interpretation that the Scrambled Word Reading condition is driving the differential activity effect with regard to English skill, such that higher English skill leads to more efficiency in phonological processing areas during word reading. This pattern is more robust in the Original model but still present in the RT model after accounting for differences in response times.

The correlational results between differential activity in the phonological ROIs and programming skill strengthened in magnitude under the RT model. Specifically, performance on the Python Declarative Knowledge test was positively correlated with differential activity between the Python and Scrambled Word Reading conditions in the preSMA ($r = 0.26$, $p = 0.098$), left inferior frontal ($r = 0.28$, $p = 0.07$), and right parietal ($r = 0.30$, $p = 0.05$) ROIs (see Figure 45 and Table 12C). These correlations were driven by the Scrambled Word Reading condition, where word reading activity was negatively correlated with performance on the Python Declarative Knowledge test (see Table 12B). The working interpretation for this pattern across analyses is that more skilled programmers are more familiar with the code related words in the Scrambled Word Reading condition and as a result show more neural efficiency in their initial comprehension of the word list. This interpretation is further supported by the fact that these correlations strengthen under the RT model which more selectively models the early

comprehension portion of the stimulus time window where this efficiency bias would be most likely to occur.

Under the RT model, the relationship between Forward Digit Span and differential activity in the left inferior frontal gyrus attenuated to no longer reach significance (Original: r = -0.33, p = 0.03; RT: r = -0.13, p = 0.41) though the direction of the correlation remained consistent.

Exploratory Whole-Brain Correlations: Response Time Model (Python > Scrambled Word Reading)

Exploratory whole brain correlations between the differential activity on the Comprehension task and the behavioral measures of interest are reported in Table 17. These exploratory correlations are reported selectively for the RT model, which better captures both task conditions by equating differences in their timings within the GLM. The Python Declarative Knowledge test, Nelson Denny Comprehension, and Forward Digit Span measures were minimally correlated with differential activity at the whole brain level - producing only a few small clusters of activation none of which were in phonologically relevant areas. This suggests that neither programming skill (see Table 17A), English skill (see Table 17B), nor working memory capacity (see Table 17D) led to drastic differences in the way the brain responded to comprehending Python compared to the stimuli in the Scrambled Word Reading condition.

The whole-brain correlation with MLAT II did reveal several positive correlations with differential task activity in left-lateralized language areas including the inferior frontal gyrus, superior temporal gyrus, and the visual word form area (see Table 17C). This pattern of results in consistent with the idea that individuals with better phonological skills behaviorally use language processing centers during code comprehension. The fact that these correlations were observed

with differential activity suggests that the propensity of more phonologically skilled individuals to bring online language mechanisms may be greater when task demands are more challenging as opposed to employed equally in the Python and Scrambled Word Reading conditions.

The code-of-thought measures were also correlated with differential activity at the whole-brain level. IRQ-Verbal was negatively correlated with differential activity in a distributed network of primarily bilateral frontoparietal areas (see Table 17E). IRQ-Visual was negatively correlated with differential activity in a smaller network of primarily left lateralized regions (see Table 17F). Verbal bias on the Card Sorting task was positively correlated with a network of areas including the left angular gyrus, left putamen, bilateral frontal areas, and the cerebellum (see Table 17G). Together, the code-of-thought measures correlating with differential activity suggest that the propensity for engaging in a verbal or visual strategic habit may interact dynamically with the demands of the task.

Areas of Differentiation Interim Summary

Comparing the Python and Scrambled Word Reading conditions directly provided the opportunity to think about how the phonological system may be differentially involved in diverging task contexts. The reported analyses also highlight the dynamic way that model timing can influence the results. Under the Original model, code comprehension activity was greater than word reading activity in all of the phonological ROIs. This pattern was also seen in the whole-brain GLM results, where activation was greater for Python than Scrambled Word Reading in a network of regions including bilateral frontoparietal areas, bilateral temporal regions, the insula, and the striatum. Under the RT model, the differential activity between the conditions dissipated. In the ROI comparison, phonological activity was still significantly greater for Python than for Scrambled Word Reading in the bilateral parietal ROIs, but no longer

reached significance in the bilateral frontal or preSMA ROIs. Likewise, the whole-brain GLM showed a reduction in differential activity under the RT model. Together, these results demonstrate that the involvement of the phonological system is not qualitatively unique to Python, yet there are quantitative differences in how the phonological system is recruited to support Python relative to the Scrambled Word Reading condition even after controlling for time on task. This is consistent with the prediction made by the *Inner Speech: Working Memory Variant Hypothesis* that phonological activity would be greater under the more challenging demands of the Python task.

Discussion

This dissertation explored the relation between the phonological system and computer code comprehension through several related goals. First, I examined if the phonological system is involved in supporting computer code comprehension in real-time. While prior work has noted that neural areas implicated in phonological processing are also observed in brain responses to comprehending computer code, no prior study had explicitly explored the overlap between phonological and programming language neural responses in the same group of participants. Second, I examined whether or not the phonological network's involvement in computer code comprehension related to individual differences in a behavioral measure of phonological coding skill. This extended prior work demonstrating that behavioral phonological skills predict variation in learning to program (Prat et al., 2020; Prat et al., in prep). Lastly, I tested three theoretical hypotheses to explain when, for whom, and how the phonological system is involved in computer programming comprehension. The *Lexical Access Hypothesis* proposed that the relation between phonology and programming was epiphenomenal and arose as a byproduct of accessing the lexical meanings of English words embedded in code. Both the inner speech

hypotheses proposed that programmers engaged phonological mechanisms to support internal speech representations, but made different predictions about the factors that would drive variability in the propensity for inner speech. The *Inner Speech: Working Memory Variant Hypothesis* suggested that variability in inner speech was driven by task demands and individual differences in contending with them, whereas the *Inner Speech: Code-of-Thought Hypothesis* suggested that variability in inner speech was driven by strategic differences in the ways individuals construct their internal realities. Each of these theoretical hypotheses produced a set of operational predictions. In the sections that follow I will explore how my results supported or failed to support these predictions and what these patterns as a whole can elucidate about the theoretical mechanisms linking phonology and computer code comprehension.

Evidence for the Phonological System's Involvement in Code Comprehension

Prior work suggests that phonological skills may relate to programming abilities; however, this dissertation was the first to explicitly investigate this relationship in a code comprehension task. The results of this study found that the phonological system is active in supporting real-time computer code comprehension. Both the whole brain (see Figures 17, 18, and 30) and ROI (see Figures 39 and 43) analyses showed considerable overlap between the phonological network elucidated by the Phonological Localizer task and the brain areas involved in code comprehension. Moreover, the network of regions recruited during code comprehension was also largely consistent with the results reported in prior neuroimaging studies (i.e., Endres et al., 2021; Ikutani et al., 2021; Ivanova et al., 2020; Liu et al., 2020; Peitek et al., 2020; Siegmund et al., 2014; Xu et al., 2021). However, engagement was also observed in the same neural areas during the Scrambled Word Reading condition (see Figures 22, 23, and 30), albeit to a lesser degree.

The results of the present study demonstrated that individual differences in behavioral phonological skills influenced the way the phonological network was deployed during code comprehension. Specifically, a negative trend was observed between code comprehension activity in the left parietal ROI and performance on the MLAT II (see Figures 24 and 31). The left parietal ROI targeted the portion of the left supramarginal gyrus that was most active for each participant during the Phonological Localizer task. The exploratory whole brain results confirmed the negative correlation between the left supramarginal gyrus and MLAT II and also found bilateral positive correlations with the angular gyrus (see Tables 8C and 13C).

Recent work using dynamic causal modeling has proposed a directional model to explain how phonological codes arise in the brain and the relative roles that the supramarginal and angular gyri play in this process (Junker et al., 2023). In the best fitting model, the left supramarginal gyrus converts the orthographic representation into its initial phonological code. Then bidirectional neural connections link the left supramarginal gyrus to the left pars opercularis in the inferior frontal gyrus where the phoneme representations are combined into more complex phonological representations. The pars opercularis is bidirectionally connected to the angular gyrus, which in turn binds the phonological code to a semantic meaning representation. Neurostimulation work also supports differential roles for the supramarginal gyrus and angular gyrus in phonological and semantic processing respectively (Stoeckel et al., 2009). In light of this functional specificity, rhyming tasks that use nonword stimuli, lacking semantic content, typically do not elicit activity in the angular gyrus (see Geva, 2018 for a discussion). Thus, it is unsurprising that the Phonological Localizer task employed herein did not elicit activity in the angular gyrus, and as a result, correlations with the angular gyrus were only detectable in the exploratory whole-brain analysis. Taken together this pattern of results suggests

that individuals with better phonological skills, as indexed by higher performance on the MLAT II, are able to generate more efficient phonological codes via the left supramarginal gyrus and recruit the angular gyrus to link these phonological codes to semantic representations.

Several past studies from our lab have shown a positive correlation between MLAT II and learning to program (Prat et al., 2020; Prat et al., in prep), however, in the present study this relationship was not observed (see Table 4). This is not necessarily at odds with our prior findings as in the present study programming skill and experience were allowed to vary freely making it challenging to parse exactly how much experience produced the Python skills observed in the present sample. In light of this variability, it is even more notable that individual differences on the MLAT II influenced the way the phonological system was deployed in the brain during code comprehension. If programming experience level was more explicitly controlled for, these patterns would likely strengthen and may relate to programming skills behaviorally to replicate the prior findings from our early learning studies (Prat et al., 2020; Prat et al., in prep).

Taken together, the present study observed that the phonological system is involved in real-time code comprehension at the group-level. Moreover, individual differences in behavioral phonological skills influenced the way that these phonological systems were deployed during code comprehension. These results raise the related question of *why* the phonological system is involved in code comprehension. Several theoretical hypotheses were put forth in this dissertation to address this question. In the sections that follow I will explore each of these theories in turn and discuss if the operational predictions generated by them are supported by the study's results.

Phonology Is Not Just About Lexical Access

The *Lexical Access Hypothesis* proposes that the relationship between phonology and programming comprehension was largely epiphenomenal in nature. Thus, the Lexical Access Hypothesis predicted that the phonological involvement in the Python and Scrambled Word Reading conditions should be equal, as the lexical demands were equated across the two conditions. Additionally, I predicted based on models of silent reading (e.g., Coltheart.,1988; 1993; Jobard et al., 2003) that the extent to which phonology is tied to lexical access should vary as a function of individual differences in English skill. The results of the present study failed to support both of these operational predictions, suggesting that the role of phonology in code comprehension is not about lexical access.

Inconsistent with the prediction that phonological involvement should be equal for the Python and Scrambled Word conditions, activation in phonological ROIs was greater in the Python condition than the Scrambled Word Reading condition (see Figure 39). This pattern was also observable in the group-level whole-brain results where there was greater activation in the Python condition than in the Scrambled Word Reading condition in a distributed network of regions including phonological areas like the left supramarginal gyrus and left pars opercularis (see Figure 37). While the difference in phonological system recruitment between conditions was greater under the Original GLM that allowed time on task to vary freely, the same pattern was also observable under the RT model in a more fine-tuned network within the parietal cortex (see Figure 43). Together, the group-level ROI and whole-brain results fail to support the operational prediction that phonological activation is equivalent in conditions where lexical access demands are equated.

Second, the *Lexical Access Hypothesis* predicted that individual differences in English skill should dictate the ease of lexical access and the resulting tendency to engage phonological mechanisms (Coltheart, 1980; Coltheart et al., 2001). Both the ROI and whole-brain results failed to support this prediction.

During code comprehension, variability in Nelson Denny Comprehension scores was not significantly correlated with phonological activity at either the ROI (see Tables 7A and 12A) or whole-brain level (see Tables 8B and 13B). However, significant positive correlations – the opposite direction of what would be predicted by the *Lexical Access Hypothesis* – were observed between code comprehension activity and Nelson Denny Comprehension scores in semantic retrieval areas including bilateral portions of the temporal lobe (see Tables 8B and 13B). These results suggest that more skilled readers utilize the semantic meanings of words to aid code comprehension. This result is consistent with recent work from our lab demonstrating that English semantics influences the way that programmers understand code – even when the code is syntactically correct (Kuo & Prat, 2024). The present study extends this finding to suggest that the extent to which English semantics are used to aid computer code comprehension varies as a function of English reading skill.

During Scrambled Word Reading, Nelson Denny Comprehension was negatively correlated with phonological activity in the left inferior frontal gyrus, left parietal, right parietal, and preSMA ROIs under the Original model (see Table 7B). This result suggests while reading skill does not relate to phonological involvement during Python trials, in Scrambled Word Reading trials more-skilled readers engage phonological areas less than less-skilled readers. However, the magnitude of this correlation attenuated under the RT model, and only remained significant in the left parietal ROI (see Table 12B). This drop in correlation magnitude between

the Original and RT models may reflect better readers responding faster on Scrambled Word Reading trials, and as a result have a disproportionately worse fit model for the Scrambled Word Reading regressor under the Original GLM. This reduction in correlation strength may have also been driven by the cognitive demands of preparing for the comprehension probes, which are minimized in the RT model. The Scrambled Word Reading comprehension check probes required participants to make semantic comparisons (see Figure 3), a process that is more strongly tied to English skill than the initial word-nonword judgment. While the Original model included the neural activity associated with the earlier word-nonword and later semantic relatedness judgements, the RT model selectively focused on the earlier process. If English skill is driving phonological recruitment on the later semantic judgment portion of the task this may have also contributed to the weaker correlation observed under the RT model.

Taken together, the results of this dissertation failed to support both operational predictions of the *Lexical Access Hypothesis*. Despite the lexical demands of the Python and Scrambled Word Reading conditions being carefully equated, both the ROI and whole-brain results support the conclusion that the phonological system was differentially involved in the task conditions. Additionally, individual differences in English skill were not related to the degree to which the phonological system was involved in code comprehension. However, English skill did relate to how semantic regions were recruited during code comprehension, which suggests that individual differences in English skill do play a non-phonological role in how programmers understand code. This result also raises the related possibilities that either the words included in the stimuli were too high frequency or the participants were too skilled in English (i.e., all being enrolled at an English-speaking university) to elicit activation of the longer phonological route proposed by the Dual-Route model (Coltheart, 1988; 1993; see Figure

1C). It is possible that if the stimuli were manipulated to include lower frequency words, individual differences in English skill may have led to differential recruitment of phonological areas in accordance with the Dual-Route theory. However, in the present study the observed activation in phonological areas during Python trials does not appear to be related to lexical access demands. These results suggest that the role of the phonological system in code comprehension is not merely an epiphenomenal byproduct of accessing the meanings of the English words in the code.

Phonology as a Mechanism to Support Inner Speech Processes

An alternative hypothesis explored at length in the sections that follow is that individuals may use phonological mechanisms to engage in inner speech whilst they comprehend code. Philosophers, psychologists, and laypeople alike have long understood that most people experience the phenomenon of "hearing an inner voice" or "talking to oneself" (for discussions see Kompa & Mueller, 2022; Munroe, 2023; and Russo, 2019). In this dissertation, I examined whether the extent to which individuals engaged in inner speech might be used to scaffold internal thought by exploring how task demands and strategic habits influenced phonological recruitment during code comprehension.

Compensatory Inner Speech: Evidence for Verbalizing in Response to Task Demands

Inner speech can be used as a mechanism to contend with difficult task demands (e.g., Marvel & Desmond, 2012). The *Inner Speech: Working Memory Hypothesis* suggests that the phonological system is used in code comprehension to support inner speech, which is deployed in a compensatory way to deal with challenging task demands. From this theory, three key operational predictions emerge. These predictions are discussed subsequently in relation to the results of the present study.

First, because the Python trials were more cognitively taxing than the Scrambled Word Reading trials, a larger phonological response was expected for Python than for Scrambled Word Reading. This prediction was supported in both the group-level ROI and whole-brain results. In the ROI analyses, activation was greater in all of the phonological ROIs for the Python condition relative to the Scrambled Word Reading condition (see Figure 39). Likewise, the whole-brain results found greater activation for Python than for Scrambled Word Reading in a distributed network of regions including phonological areas like the left inferior frontal gyrus and left supramarginal gyrus (see Figures 22 and 37). The areas that displayed this differentiation between the Python and Scrambled Word Reading conditions overlapped with the group-level statistical map from the Phonological Localizer task (see Figure 38). When response times were considered using the RT GLM, the magnitude of the difference in phonological brain responses between the Python and Scrambled Word Reading conditions was reduced in both the ROI and whole-brain analyses (see Figures 37 and 38). However, under the RT model, phonological activation was still greater for the Python condition than for the Scrambled Word Reading condition in a smaller network of regions (see Figure 37 and 43). Together these results support the prediction that individuals use phonological resources to support inner speech in contexts where task demands are more challenging.

Second, how demanding a task is should be tied to individual differences in the domain knowledge required for that task. If inner speech was used to compensate for task demands, then more skilled programmers would need to engage in compensatory inner speech less than less-skilled programmers. The results of the present study failed to support this prediction. Relatively few studies have investigated how individual differences in programming skill relate to the deployment of neural resources when completing coding-related tasks in the scanner. Those that

have, found evidence that the degree of activation in the bilateral inferior frontal gyrus during programming is related to programming skill level (Ikutani et al., 2021; Petiek et al., 2020), and that more programming experience increases the neural response in right inferior frontal gyrus (Hishikawa et al., 2023). The correlational results of the present study did not converge with these prior studies, possibly because of the wider variability of programming skill levels in the present sample. In the ROI results, programming skill was not significantly correlated with phonological activity during code comprehension in any of the ROIs (see Tables 7A and 12A). In the exploratory whole-brain correlational analyses, programming skill was correlated negatively with some areas in frontostriatal regions and positively correlated with the precuneus and portions of the right visual cortex (see Tables 8A and 13A). While these correlations were not in regions typically implicated in phonological processing or inner speech, the general directionality of these correlations in cognitive control areas is consistent with the notion that greater skill should lead to more efficiency in neural processing. This pattern of results was contrary to the prediction that novice programmers would lack the ability to use top-down schemas to understand patterns in code, and as a result would rely on effortful bottom-up comprehension of individual code tokens linked together via the phonological loop (Pennington, 1987; Shneidermand & Mayer, 1979). It is possible that the Python stimuli were not complex enough or followed schematic patterns too consistently to engage the expected bottom-up comprehension in less-skilled programmers. However, behavioral performance on the in scanner programming task nicely mapped onto the other behavioral measures of programming skill and showed similar degrees of variability which partially ameliorates this concern (see Figure 12 and Tables 1 & 4). Taken together, these results fail to support the operational prediction that

programming skill should relate to the extent that individuals engage in inner speech during code comprehension.

Lastly, the demands of a task should be relative to the domain-general cognitive capabilities of the individual. Thus, it was predicted that variation in working memory capacity would be related to the extent to which individuals utilize compensatory inner speech mechanisms in the face of task demands. The results of the present study partially supported this prediction. In the ROI analyses, greater working memory capacity was positively correlated with both code comprehension activity and word reading activity in the right parietal ROI under the Original GLM (see Figures 20 & 25 and Table 7). All other ROI correlations, though not significant, trended in a negative direction. When response times were controlled for in the RT GLM, the positive correlation between working memory capacity and phonological activity in the right parietal ROI got weaker and no longer reached significance. Conversely, the negative correlation between working memory capacity and phonological activity in the left parietal ROI got stronger and reached significance under the RT model (see Figure 32 and Table 12). Together, this pattern of results suggests that the left and right parietal ROIs are differentially supporting different phases of the task.

Under this interpretation, the left parietal ROI supports inner speech processes to aid code comprehension. Individuals with lower working memory capacity tend to engage this compensatory system more to cope with task demands that tax their limited capacities. The observation that this correlation gets stronger under the RT model is in line with the idea that the left parietal ROI may be more strongly implicated in the code comprehension process rather than maintaining the probe related information. This interpretation also makes sense with respect to the directionality of the left hemisphere results, the left supramarginal gyrus is well established

to be involved in generating the phonological codes that make up inner speech, and activation in this area has been shown to increase with task demands (e.g., Geva, 2018; Price, 2010). For lower working memory capacity individuals, the Python condition is likely more cognitively taxing requiring greater recruitment of neural resources that support inner speech. This relationship weakens when the model includes activation that is not directly relevant to the early comprehension process in the Original GLM.

Conversely, the correlation between right parietal and working memory capacity is stronger under the Original GLM suggesting that the right parietal ROI may be involved in maintaining the probe-related information in working memory. It is well established that the parietal cortex is sensitive to working memory load and is implicated in storing information in working memory (e.g., Eriksson et al., 2015; McNab & Klingberg, 2008). Higher working memory capacity individuals may be utilizing their larger working memory capacities to maintain more information that could aid their ability to answer the comprehension probes. It is somewhat unclear why this later maintenance process is observed in the right hemisphere. While the left inferior parietal lobule is strongly implicated in verbal working memory processes (see Emch et al., 2019 for a meta-analysis), right parietal areas have been linked more strongly to visuospatial working memory (e.g., Nee et al., 2013; Owen et al., 2005) which could suggest that the maintenance of the probe related information may occur via alternate mechanisms. Alternatively, the use of the right hemisphere to maintain probe information may reflect the characteristics of the sample which included bilinguals many of whom had logographic first languages, as the language network of logographic speakers tends to be more bilateral (e.g., Tan et al., 2005). This possibility is supported by the fact that, behaviorally, working memory

capacity was negatively correlated with English skill (see Table 4), suggesting that the higher working memory participants may have disproportionately been logographic language speakers.

Taken together, the results of this dissertation provide mixed evidence for the hypothesis that phonological mechanisms support inner speech in order to contend with task demands. At the group-level, activation in the phonological system was greater for the more challenging Python condition than for the less demanding Scrambled Word Reading condition; this remained true in a more fine-tuned network when response times were accounted for in the RT model. I also found that individual differences in working memory capacity predicted the extent to which the phonological system was involved in code comprehension. Together these results suggest a relationship between the capacity of an individual and their need to use inner speech mechanisms to offload demands associated with task difficulty. However, no relationship was observed between programming skill, assessed outside the scanner, and phonological system recruitment in the present study which contradicted one of the operational predictions made by the *Inner Speech: Working Memory Variant Hypothesis*. This suggests that the tendency to engage in inner speech may be more related to individual differences in domain general constraints and strategies rather than skill-specific expertise. This idea is explored further with respect to how variation in strategy relates to inner speech propensity in the subsequent section.

Strategic Inner Speech: Evidence for Verbalizing as a Feature of Individual Differences in Information Processing

The *Inner Speech: Code-of-Thought Hypothesis* proposed that variability in how individuals represent their inner experience will dictate the degree to which inner speech mechanisms are engaged during code comprehension. From this theory, two operational predictions were made.

First, it was expected that individuals with more verbal codes-of-thought would engage phonological areas *more* than individuals with less verbal codes-of-thought. By and large, this prediction was supported by the results of the present study. In the ROI analyses, greater attentional bias towards verbal information in the Card Sorting task was positively correlated with code comprehension activity in the preSMA and left parietal ROIs (see Table 12 and Figure 34). This result suggests that individuals whose attention is more biased towards verbal information are utilizing neural resources consistent with an inner speech strategy during code comprehension. Verbal bias on the Card Sorting task was also positively correlated with programming activity in the whole-brain analysis in phonological and semantic areas (see Table 13G). Significant correlations between programming activity in the phonological ROIs and IRQ-Verbal were not observed in the ROI analyses (see Table 7A and 12A). However, IRQ-Verbal was positively correlated with activation in the left angular gyrus in the exploratory whole-brain analysis, suggesting a relationship between the extent to which individuals represent internal information verbally and the use of semantic brain areas during code comprehension (see Tables 8E and 13E). Taken together, the results of the present study support the operational prediction that greater tendencies towards verbal information processing underpins the extent to which individuals engage in inner speech during code comprehension.

A second related prediction is that individuals with more visual codes-of-thought will engage inner speech mechanisms *less* when they comprehend code. This prediction was partially supported by the ROI results. Individuals who self-reported a greater propensity towards visual strategic habits on the IRQ-Visual exhibited less activity in the preSMA and left parietal ROIs during code comprehension (see Figure 34 and Table 12A). The exploratory whole-brain analyses further supported this conclusion, as IRQ-Visual was negatively correlated with a

distributed network of regions including the a key phonological area, the left supramarginal gyrus (see Tables 8F and 13F). The whole brain results also found that IRQ-Visual was positively correlated with activation in bilateral visual cortex (see Tables 8F and 13F). This result is consistent with the interpretation that individuals with more visual codes-of-thought recruit visual neural resources to aid code comprehension instead of utilizing verbal networks. This pattern of results nicely replicates prior work in the arithmetic domain, demonstrating that individuals with more visual codes-of-thought, engage visual cortex during mental calculation, whereas individuals with more verbal codes-of-thought recruit verbal areas like the left angular gyrus to support the same mental calculation process (Zarnhofer et al., 2013).

Taken together, the results of the present study support the *Inner Speech: Code-of-Thought Hypothesis*. Individuals with more verbal codes-of-thought tended to engage the phonological system more during code comprehension whereas those with more visual codes-of-thought engaged the phonological system less and utilized visual areas more. This pattern of results is consistent with past neuroimaging studies on code-of-thought demonstrating that information processing style can lead to differential recruitment of neural networks linked to language and visuospatial functions (e.g., Alfred et al., 2020; Kraemer et al., 2009; Zarnhofer et al., 2013).

On Talking to Myself: Insights on When, for Whom, and How Inner Speech Occurs

The results of the present study support the idea that task demands, as well as individual differences in cognitive capacity and strategy influence the extent to which the phonological system is engaged during code comprehension. Phonological areas were more active in the Python than Scrambled Word Reading condition, suggesting that on average people tended to engage in inner speech more when the task was more difficult. This is consistent with prior work

suggesting that individuals engage phonological brain areas more when working memory demands increase and cognitive resources are taxed (e.g., Marvel & Desmond, 2012). This is also consistent with the philosophical notion that the purpose of inner speech is to provide a structure for problem solving (e.g., Kompa & Mueller, 2022), and that inner speech becomes more elaborative and phonological in nature when cognitive resources are taxed (e.g., Alderson-Day & Fernyhough, 2015). The results of the present study support the prediction that *when* cognitive resources are scarce, inner speech can provide a buffering function to keep multiple pieces of information active long enough to extract a larger meaning representation. Individual differences results further supported this conclusion, such that higher working memory capacity individuals relied less on phonological areas during code comprehension than lower working memory capacity individuals. This finding highlights the fact that task demands are always relative to an individual's capacity to do that task. Likewise, the strategy employed to contend with task demands may be relative to an individual's preferred code-of-thought. This raises a related possibility that the phonological system is recruited to contend with challenging task demands *selectively* for individuals that have more verbal codes-of-thought.

The phenomenon of engaging an inner voice is something most - but not all - people experience (Nedergaard & Lupyan, 2024). However, the frequency and contexts within which individuals engage in inner speech is highly variable (e.g., Heavy & Hurlburt, 2008; Morin, Uttl, & Hamper, 2011). Contexts with greater task complexity leave more room for individual differences in performance and strategy to emerge (see Ackerman, 2007 for a discussion). As such, the Python condition likely left more room for individual differences in strategy to vary freely than the Scrambled Word Reading condition which more strictly required participants to engage in semantic processing. Prior work has demonstrated that when task demands explicitly

require participants to attend to more visual or verbal features, individual differences in code-of-thought did not meaningfully impact performance (Kraemer et al., 2017). However, when participants were not given instructions on what to attend to, code-of-thought did lead to differences in more verbal or more spatial content being better remembered (Kraemer et al., 2017). In the present study, individual differences in code-of-thought were differentially related to verbal and visuospatial neural recruitment during code comprehension. This pattern was not observed to a similar magnitude during the Scrambled Word Reading trials, which likely left less room to engage in visuospatial strategies. Taken together, the results reported herein support the notions that individual differences in cognitive capabilities and strategic habits jointly influence *who* engages in inner speech, and that the demands of the task can constrain the interplay between these factors.

The results of the present study also provide insights into *how* this interplay is instantiated at the neural level. While it is challenging to know if an individual is engaging in inner speech behaviorally, the neuroimaging methods employed herein provided a unique methodological advantage to understanding the relation between the phonological system and code comprehension. The results of the present study found that inner speech processes emerge through dynamic connections between phonological and semantic neural regions during code comprehension. However, mapping the purpose of this inner speech at a cognitive level remains somewhat challenging. Many of the phonological regions active during code comprehension, including the left supramarginal gyrus and left inferior frontal gyrus, have been implicated in studies that examine the phonological system's involvement in cognition across varying levels of processing including: lower-level word reading (e.g., Papoutsi et al., 2009, Stoeckel et al., 2009), verbal working memory maintenance via articulation mechanisms in the phonological loop (e.g.,

Emch et al., 2019; Marvel & Desmond, 2012; Perrachione et al., 2017), and inner dialogue to support self-regulation and higher-level reasoning (e.g., Morin & Michaud, 2007; Yoshioka et al., 2023). Together these divergent literatures raise interesting questions about how to consider the level, or levels, of processing at which the phonological system is bought online during code comprehension. This is partially addressed by the hypotheses of the present study which diverge from one another with respect to the level of processing in which the phonological system is most tightly coupled to code comprehension. In the *Lexical Access Hypothesis*, the phonological system was hypothesized to come online in lower level lexical processing at the word-level. Conversely, the *Inner Speech Hypotheses* suggested that the phonological system supports higher level processing to: 1) maintain multiple code tokens in mind and parse their collective meaning via the phonological loop and/or, 2) to support an internal dialogue to guide problem solving strategies.

On Understanding the Dynamic Interaction Between Demands, Capabilities, and Strategies: Open Questions and Future Directions

It is somewhat unclear in the present study whether inner speech is supporting a passive phonological rehearsal or a more active dialogue to aid problem solving. Using Baddely's working memory framework this tension speaks to the question of whether inner speech is solely a function of the phonological loop or whether phonological representations are being generated to support high-level planning akin to the central executive (e.g., Baddely, 2003; 2010). While Baddely's components of the mind do not neatly map on to isolated neural regions, broadly, the phonological loop tends to be associated with the left inferior frontal gyrus and portions of the left temporoparietal junction including the supramarginal gyrus (e.g., Buchsbaum & D'Esposito, 2008; Papagno et al., 2017) whereas the central executive tends to be associated with cognitive

control areas in the dorsolateral prefrontal cortex and anterior cingulate (e.g., Funahashi, 2017; Kerns et al., 2004; Miller & Cohen, 2001). In the present study, greater activation was observed for the Python than Scrambled Word Reading condition in both phonological and cognitive control areas (see Figure 37, and Tables 15 & 16). It is unclear the extent to which these responses influenced one another. Methods that examine functional connectivity between phonological and cognitive control areas, and the directionality of these connections, may be able to shed further light on the possibility that phonological codes are used to support cognitive control processes during code comprehension. Future work using less verbalizable stimuli may also be advantageous, as this should remove the need for route phonological rehearsal while still allowing inner speech that serves a function in aiding problem solving to be observed.

The extent to which this central executive function is supported by verbal codes may differ as a function of individual differences in code-of-thought. By this view individuals with more verbal codes-of-thought may use phonological codes to support an internal voice that drives cognitive control processes, whereas individuals with more visual codes-of-thought may use alternative methods to support cognitive control. Vygotsky's notion that inner speech is part of the typical development trajectory (e.g., Vygotsky, 1934; 1987) and studies finding that the majority of people engage in inner speech to some degree (Heavy & Hurlburt, 2008), may jointly reflect the phenomenon that verbal codes-of-thought are more prominent than visual codes-of-thought at a population level. Prior work from our lab using the Card Sorting task found that 61% of people were biased towards verbal information (Czerwonka, Mottarella, & Prat, 2024). Using a non-overlapping sample, the results of the present study replicated this finding with 60% of participants responding more strongly to verbal information on the Card Sorting task. If a larger proportion of the population uses verbal codes to support internal thought representations,

then the observation of seeing phonological recruitment increase in response to challenging task demands may be a function of individuals utilizing whatever resources are most compatible with their code-of-thought. This may mean that individuals with more visual codes-of-thought use visuospatial strategies – instead of inner speech – to contend with challenging task demands.

In the present study, participants were recruited with an individual differences methodology in mind meaning that there was wide variability in the measures. This variability in combination with the sample size made the present data ill-suited to robust group-level analysis of how differences between code-of-thought profiles led to recruitment of diverging neural areas to contend with task difficulty. However, the intersection between strategy and task demands is an important area for future work. Utilizing a paradigm that systematically varied the difficulty of stimuli within conditions (i.e., using harder or easier Python problems and using lower and higher frequency words), and recruiting participants with more specific code-of-thought profiles would be necessary to gain further traction on this question.

Limitations

The results of the present study provide novel insights into how the phonological system supports code comprehension. However, there are several key limitations to consider while interpreting these results.

One of the main challenges of the task design employed herein was how to pull apart the overlapping cognitive demands of the Comprehension task. The Python condition required participants to comprehend code, detect syntax errors, and maintain in mind a predicted output to compare against a comprehension check probe. The rationale behind this task design was to encourage true comprehension whilst having participants respond to every trial to stay engaged in the task. I considered including comprehension check probes on every trial, but eventually

opted not to do this as it would have either substantially increased the length of the task or reduced the number of trials. The complexity of this task design makes it challenging to deduce exactly which underlying cognitive process is related to the observed brain activity at a given time. Controlling for response times with the RT model, provided a coarse method for minimizing activation associated with maintaining probe related content. However, participants may still be maintaining the probe content to some degree before making their initial response. Additionally, the early time window captured in the RT model almost certainly included neural responses associated with both error detection and comprehension processes as these demands overlapped in time. Prior work on the functional specificity of the neural regions observed during the Comprehension task provides some insights into which regions were more closely tied to the comprehension compared to the error detection process. For example, prior work has strongly implicated the insula as a key area to detect errors in code (Castelhano et al., 2019; Duares et al., 2016), thus the insula activity observed during the Comprehension task likely was related to the error detection component of the task. However, it remains challenging to fully disentangle how much of the neural response observed in each area was specific to comprehension. Future work using comprehension check probes for every trial and excluding the syntax error manipulation could help to further delineate between the portions of the code neural network involved in error detection as opposed to comprehension.

A related challenge was that there were significant differences in response times between the MRI task conditions. In the Comprehension task, participants took significantly longer to complete the Python trials than the Scrambled Word Reading trials (see Figures 5 and 11C). The RT model was introduced to account for these differences in time on task, but this remains a limitation of the task design. There were also differences in response times between the Rhyming

and Letter Search conditions in the Phonological Localizer task (see Table 1). The blocked design of the Phonological Localizer was not well suited to creating a response time based model as there was not adequate time between trials for the hemodynamic response to fall back to baseline. This may have contributed to differences in the neural activity observed between conditions. However, this was not a primary concern as the primary purpose of the Phonological Localizer was to isolate for each participant where the differentiation between conditions was largest within a theoretically constrained search area. If there were differences in the fit of conditions, they likely would be relatively consistent across the voxels included in the search area, and thus would not be expected to dictate which specific voxels were selected for inclusion in the subject-specific ROI. Using a self-paced task design and modeling both MRI tasks based on individual response times, potentially with the inclusion of a jittered interstimulus interval, would be a good future direction to alleviate some of these limitations.

Another limitation of the present study was that there was not a condition in the Comprehension task that was complex but not verbal. The inclusion of a condition with visuospatial task demands like a complex mental rotation task may have provided a nice baseline to help differentiate if phonological brain areas were being engaged selectively in verbal conditions or were involved in representing an internal dialogue to guide strategies across task-domains. This manipulation would have helped to shed light on the question of if inner speech is supporting a verbally coded central executive. Additionally, examining a visuospatial task would provide some interesting opportunities to look at the interaction between code-of-thought and task demands in more depth. I see this as an important avenue for future research.

Conclusion

This dissertation was the first study to examine the relation between phonology and computer code comprehension. The results reported herein support the conclusion that participants engaged in inner speech when comprehending code, and failed to support the hypothesis that phonology was epiphenomenally related to code comprehension through lexicosemantics. While the phonological network was also observed to some extent during word reading, activation in this network was more robust during code comprehension, consistent with the idea that individuals used inner speech mechanisms to contend with challenging task demands. This conclusion was further supported by the finding that higher working memory capacity individuals engaged less phonological resources during code comprehension, suggesting that individuals with higher cognitive capacities need to compensatorily offload information into an inner speech representation less than lower capacity individuals. Results also supported the idea that baseline differences in the way that individuals construct their internal representations of the world influenced the extent to which participants engaged neural areas implicated in verbal or visuospatial processing. Together, these results suggest that inner speech emerges as a function of both one's capacity to contend with task demands and one's strategic habit for how they represent information. It is unclear from the present results exactly how these factors interact to give rise to inner speech, and I see this is an important avenue for future research. Taken together, this dissertation adds to the burgeoning literature on how individuals understand computer code and suggests that internal speech representations can aid code comprehension for some individuals.

References

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science*, *16*(5), 235–239. https://doi.org/10.1111/j.1467-8721.2007.00511.x

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, *65*, 48–58. https://doi.org/10.1016/j.concog.2018.07.001

Alfred, K. L., Hayes, J. C., Pizzie, R. G., Cetron, J. S., & Kraemer, D. J. M. (2020). Individual differences in encoded neural representations within cortical speech production network. *Brain Research*, *1726*, 146483. https://doi.org/10.1016/j.brainres.2019.146483

Alfred, K. L., & Kraemer, D. J. M. (2017). Verbal and visual cognition: Individual differences in the lab, in the brain, and in the classroom. *Developmental Neuropsychology*, *42*(7–8), 507–520. https://doi.org/10.1080/87565641.2017.1401075

Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, *47*, 1. https://doi.org/10.2478/psicl-2011-0002

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B. M., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *The Journal of*

Comparative Neurology, *412*(2), 319–341. https://doi.org/10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7

Antonietti, A., & Giorgetti, M. (1998). The Verbalizer-Visualizer Questionnaire: A review. *Perceptual and Motor Skills*, *86*(1), 227–239. https://doi.org/10.2466/pms.1998.86.1.227

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review*, *17*(1), 95–100. https://doi.org/10.3758/PBR.17.1.95

Ashby, J., Sanders, L. D., & Kingston, J. (2009). Skilled readers begin processing sub-phonemic features by 80ms during visual word recognition: Evidence from ERPs. *Biological Psychology*, *80*(1), 84–94. https://doi.org/10.1016/j.biopsycho.2008.03.009

Austin, H. S. (1987). Predictors of Pascal programming achievement for community college students.

Baddeley, A. (1986). *Working memory.* Clarendon Press/Oxford University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Baddeley, A., Eldridge, M., & Lewis, V. J. (1981). The role of subvocalization in reading. *The Quarterly Journal of Experimental Psychology*, *33*, 439–454.

Bennedsen, J., & Caspersen, M. E. (2006). Abstraction ability as an indicator of success for learning object-oriented programming? *ACM SIGCSE Bulletin*, *38*(2), 39–43. https://doi.org/10.1145/1138403.1138430

Bennedsen, J., & Caspersen, M. E. (2019). Failure rates in introductory programming: 12 years later. *ACM Inroads*, *10*(2), 30–36. https://doi.org/10.1145/3324888

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies.

Cerebral Cortex (New York, NY), *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blackburn, H. L., & Benton, A. L. (1957). Revised administration and scoring of the Digit Span Test. *Journal of Consulting Psychology*, *21*(2), 139–143. https://doi.org/10.1037/h0047235

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Brinthaupt, T. M., Hein, M. B., & Kramer, T. E. (2009). The Self-Talk Scale: Development, factor analysis, and validation. *Journal of Personality Assessment*, *91*(1), 82–92. https://doi.org/10.1080/00223890802484498

Brooks, R. (1978). Using a behavioral theory of program comprehension in software engineering. In *Proceedings of the International Conference on Software Engineering (ICSE)* (pp. 196–201). IEEE.

Brown, J. I. (1960). *The Nelson-Denny Reading Test*. Houghton Mifflin.

Buchsbaum, B. R. (2013). The role of consciousness in the phonological loop: Hidden in plain sight. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00496

Buchsbaum, B. R., & D'Esposito, M. (2008). The Search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, *20*(5), 762–778. https://doi.org/10.1162/jocn.2008.20501

Burnett, P. C. (1996). Children's self-talk and significant others' positive and negative statements. *Educational Psychology*, *16*(1), 57–67. https://doi.org/10.1080/0144341960160105

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (ed.), *Individual Differences and Universals in Language Learning Aptitude*. Newbury House, 83–118.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. W. Stansfield (eds.), *Language Aptitude Reconsidered*. Prentice Hall, 11–29.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Carroll, J. & Sapon, S. (1959). Modern Language Aptitude Test. San Antonio, TX: Psychological Corporation.

Castelhano, J., Duarte, I. C., Duraes, J., Madeira, H., & Castelo-Branco, M. (2021). Reading and calculation neural systems and their weighted adaptive use for programming skills. *Neural Plasticity*, *2021*, 1–13. https://doi.org/10.1155/2021/5596145

Castelhano, J., Duarte, I. C., Ferreira, C., Duraes, J., Madeira, H., & Castelo-Branco, M. (2019). The role of the insula in intuitive expert bug detection in computer code: An fMRI study. *Brain Imaging and Behavior*, *13*(3), 623–637. https://doi.org/10.1007/s11682-018-9885-1

Chalmers, J., Eisenchlas, S. A., Munro, A., & Schalley, A. C. (2021). Sixty years of second language aptitude research: A systematic quantitative literature review. *Language and Linguistics Compass*, *15*(11), e12440. https://doi.org/10.1111/lnc3.12440

Cheryan, S., Plaut, V. C., Handron, C., & Hudson, L. (2013). The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex Roles: A Journal of Research*, *69*(1–2), Article 1–2. https://doi.org/10.1007/s11199-013-0296-x

Chiarello, C., Welcome, S. E., & Leonard, C. M. (2012). Individual differences in reading skill and language lateralisation: A cluster analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, *17*(2), 225–251. https://doi.org/10.1080/1357650X.2011.561860

Collins, A. G. E. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of Cognitive Neuroscience*, *30*(10), 1422–1432. https://doi.org/10.1162/jocn_a_01238

Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *The Journal of Neuroscience*, *37*(16), 4332–4342. https://doi.org/10.1523/JNEUROSCI.2700-16.2017

Coltheart, M. (1980). Reading, phonological encoding, and deep dyslexia. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep Dyslexia* (pp. 197–226). Routledge & Kegan Paul.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Coltheart, V., Avons, S. E., & Trollope, J. (1990). Articulatory suppression and phonological codes in reading for meaning. *The Quarterly Journal of Experimental Psychology Section A*, *42*(2), 375–399. https://doi.org/10.1080/14640749008401227

Czerwonka, M., Mottarella, M., & Prat, C. S. (May, 2024). Visual, Verbal and Balanced Processing Styles: Exploring the Effects of Attentional Biases on Decision Making Under Conflict. Poster presentation given at Mary Gates Symposium, University of Washington

De Smedt, B., Taylor, J., Archibald, L., & Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Developmental Science, 13*(3), 508–520. https://doi.org/10.1111/j.1467-7687.2009.00897.x

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (1st ed., pp. 589–630). Wiley. https://doi.org/10.1002/9780470756492.ch18

Drieghe, D., & Brysbaert, M. (2002). Strategic effects in associative priming with words, homophones, and pseudohomophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 951–961. https://doi.org/10.1037/0278-7393.28.5.951

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Duraes, J., Madeira, H., Castelhano, J., Duarte, C., & Branco, M. C. (2016). WAP: Understanding the Brain at Software Debugging. *2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)*, 87–92. https://doi.org/10.1109/ISSRE.2016.53

Ellwood-Lowe, M. E., Whitfield-Gabrieli, S., & Bunge, S. A. (2021). Brain network coupling associated with cognitive performance varies as a function of a child's environment in the ABCD study. *Nature Communications*, *12*(1), 7183. https://doi.org/10.1038/s41467-021-27336-y

Emch, M., von Bastian, C. C., & Koch, K. (2019). Neural Correlates of Verbal Working Memory: An fMRI Meta-Analysis. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00180

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*(1), 148–168. https://doi.org/10.1016/S0749-596X(02)00511-9

Endres, M., Karas, Z., Hu, X., Kovelman, I., & Weimer, W. (2021). Relating Reading, Visualization, and Coding for New Programmers: A Neuroimaging Study. *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, 600–612. https://doi.org/10.1109/ICSE43902.2021.00062

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In *Psychology of Learning and Motivation* (Vol. 44, pp. 145–199). Elsevier. https://doi.org/10.1016/S0079-7421(03)44005-X

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, *88*(1), 33–46. https://doi.org/10.1016/j.neuron.2015.09.020

Eviatar, Z., Hellige, J. B., & Zaidel, E. (1997). Individual differences in lateralization: Effects of gender and handedness. *Neuropsychology*, *11*(4), 562–576. https://doi.org/10.1037/0894-4105.11.4.562

Fedorenko, E., & Blank, I. A. (2020). Broca's area is not a natural kind. *Trends in Cognitive Sciences*, *24*(4), 270–284. https://doi.org/10.1016/j.tics.2020.01.001

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), Article 21. https://doi.org/10.1016/j.cub.2012.09.011

Fedorenko, E., Ivanova, A., Dhamala, R., & Bers, M. U. (2019). The language of programming: A cognitive perspective. *Trends in Cognitive Sciences*, *23*(7), 525–528. https://doi.org/10.1016/j.tics.2019.04.010

Fernyhough, C. (2004). Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, *22*, 49 – 68. http://dx.doi.org/10.1016/j.newideapsych.2004 .09.001

Fernyhough, C., & McCarthy-Jones, S. (2013). Thinking aloud about mental voices. In F. Macpherson & D. Platchias (Eds.), *Hallucination: Philosophy and psychology* (pp. 87–104). MIT Press.

Floyd, B., Santander, T., & Weimer, W. (2017). Decoding the representation of code in the brain: An fMRI study of code review and expertise. *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*, 175–186. https://doi.org/10.1109/ICSE.2017.24

Frachtenberg, E., & Kaner, R. D. (2022). Underrepresentation of women in computer systems research. *PLoS ONE*, *17*(4), e0266439. https://doi.org/10.1371/journal.pone.0266439

Funahashi, S. (2017). Working memory in the prefrontal cortex. *Brain Sciences*, *7*(12), 49. https://doi.org/10.3390/brainsci7050049

Gauker, C. (2018). Inner speech as the internalization of outer speech. In P. Langland-Hassan & A. Vicente (Eds.), *Inner speech: Nature and functions. Oxford*, Oxford University Press.

Geva, S. (2018). Inner speech and mental imagery: A neuroscientific perspective. In P. Langland-Hassan & A. Vicente (Eds), *Inner Speech: New Voices* (pp. 105-130), Oxford. https://doi.org/10.1093/oso/9780198796640.003.0005

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion–symptom mapping. *Brain*, *134*(10), 3071–3082. https://doi.org/10.1093/brain/awr232

Gilbert & Moore (1981). Grace Murray Hopper. In *Particular Passions*. Clarkson Potter Inc. 1-7

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*(3), 225–255. https://doi.org/10.1080/741942359

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*(3), 575–582. https://doi.org/10.1093/cercor/bhk001

Golestani N, Paus T, Zatorre RJ. (2002). Anatomical correlates of learning novel speech sounds. *Neuron, 35*(5), 997-1010. https://doi.org/10.1016/s0896-6273(02)00862-0

Graafsma, I. L., Robidoux, S., Nickels, L., Roberts, M., Polito, V., Zhu, J. D., & Marinus, E. (2023). The cognition of programming: Logical reasoning, algebra and vocabulary skills predict programming performance following an introductory computing course. *Journal of Cognitive Psychology*, *35*(3), 364–381. https://doi.org/10.1080/20445911.2023.2166054

Graves, W. W., Purcell, J., Rothlein, D., Bolger, D. J., Rosenberg-Lee, M., & Staples, R. (2023). Correspondence between cognitive and neural representations for phonology, orthography, and semantics in supramarginal compared to angular gyrus. *Brain Structure and Function*, *228*(1), 255–271. https://doi.org/10.1007/s00429-022-02590-y

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. https://doi.org/10.1016/j.tics.2005.07.004

Haigh, C. A., Savage, R., Erdos, C., & Genesee, F. (2011). The role of phoneme and onset-rime awareness in second language reading acquisition. *Journal of Research in Reading*, *34*(1), 94–113. https://doi.org/10.1111/j.1467-9817.2010.01475.x

Haile, T. M., Prat, C. S., & Stocco, A. (2024). One size does not fit all: Idiographic computational models reveal individual differences in learning and meta-learning strategies. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12730

Halderman, L. K., Ashby, J., & Perfetti, C. A. (2012). Phonology: An early and integral role in identifying words. In J. S. Adelman (Ed.), *Visual word recognition: Volume 1: Models and methods, orthography and phonology* (pp. 207–228). Psychology Press.

Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, *17*(3), 798–810. https://doi.org/10.1016/j.concog.2007.12.006

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*(2). https://doi.org/10.1006/jecp.2000.2586

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5). https://doi.org/10.1038/nrn2113

Hishikawa, K., Yoshinaga, K., Togo, H., Hongo, T., & Hanakawa, T. (2023). Changes in functional brain activity patterns associated with computer programming learning in novices. *Brain Structure and Function*, *228*, 1691–170. https://doi.org/10.1007/s00429-023-02674-3

Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, *22*(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Ikutani, Y., Kubo, T., Nishida, S., Hata, H., Matsumoto, K., Ikeda, K., & Nishimoto, S. (2021). Expert programmers have fine-tuned cortical representations of source code. *Eneuro*, *8*(1), ENEURO.0405-20.2020. https://doi.org/10.1523/ENEURO.0405-20.2020

Ivanova, A. A., Srikant, S., Sueoka, Y., Kean, H. H., Dhamala, R., O'Reilly, U.-M., Bers, M. U., & Fedorenko, E. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. *eLife*, *9*, e58906. https://doi.org/10.7554/eLife.58906

Jenkins, T. (2002). On the difficulty of learning to program. *Proceedings of the 3rd Annual Conference of the LTSN Centre for Information and Computer Sciences 4*, 53–58.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage*, *20*(2), 693–712. https://doi.org/10.1016/S1053-8119(03)00343-4

Junker, F. B., Schlaffke, L., Lange, J., & Schmidt-Wilcke, T. (2023). The angular gyrus serves as an interface between the non-lexical reading network and the semantic system: Evidence from dynamic causal modeling. *Brain Structure and Function*, *229*(3), 561–575. https://doi.org/10.1007/s00429-023-02624-z

Kao, Y., Matlen, B., & Weintrop, D. (2022). From One Language to the Next: Applications of Analogical Transfer for Programming Education. *ACM Transactions on Computing Education*, *22*(4), 1–21. https://doi.org/10.1145/3487051

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, *12*(4), 231–242. https://doi.org/10.1038/nrn3000

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science*, *303*(5660), 1023–1026. https://doi.org/10.1126/science.1089910

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, *13*(2), 169–184. https://doi.org/10.1016/0361-476X(88)90017-3

Klaus, J., & Hartwigsen, G. (2019). Dissociating semantic and phonological contributions of the left inferior frontal gyrus to language production. *Human Brain Mapping*, *40*(11), 3279–3287. https://doi.org/10.1002/hbm.24597

Kompa, N. A., & Mueller, J. L. (2022). Inner speech as a cognitive tool—Or what is the point of talking to oneself? *Philosophical Psychology*, 1–24. https://doi.org/10.1080/09515089.2022.2112164

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: evidence for two types of visualizers. *Cognition and Instruction*, *20*(1), 47–77. https://doi.org/10.1207/S1532690XCI2001_3

Kraemer, D. J. M., Rosenberg, L. M., & Thompson-Schill, S. L. (2009). The neural correlates of visual and verbal cognitive styles. *The Journal of Neuroscience*, *29*(12), 3792-3798. https://doi.org/10.1523/JNEUROSCI.4635-08.2009

Kraemer, D. J. M., Schinazi, V. R., Cawkwell, P. B., Tekriwal, A., Epstein, R. A., & Thompson-Schill, S. L. (2017). Verbalizing, visualizing, and navigating: The effect of strategies on

encoding a large-scale virtual environment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 611–621. https://doi.org/10.1037/xlm0000314

Krueger, R., Huang, Y., Liu, X., Santander, T., Weimer, W., & Leach, K. (2020). Neurological divide: An fMRI study of prose and code writing. *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, 678–690. https://doi.org/10.1145/3377811.3380348

Kuo, C.-H., Mottarella, M., Haile, T., & Prat, C. S. (2022). Predicting programming success: How intermittent knowledge assessments, individual psychometrics, and resting-state EEG predict Python programming and debugging skills. *International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 1–6. https://doi.org/10.23919/SoftCOM55329.2022.9911411

Kuo, C.-H., & Prat, C. (2024). Computer programmers show distinct, expertise-dependent brain responses to violations in form and meaning when reading code. *Scientific Reports*, *14*(5404). https://doi.org/10.1038/s41598-024-56090-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Langland-Hassan, P. (2021). Inner speech. *WIREs Cognitive Science*, *12*(2), e1544. https://doi.org/10.1002/wcs.1544

Leeper, R. R., & Silver, J. L. (1982). Predicting success in a first programming course. ACM SIGCSE Bulletin, 14(1), 147-150.

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin*, *140*(6), 1534–1555. https://doi.org/10.1037/a0037830

Li, L. (2022). Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond. *Information Systems Frontiers*, 1–16. https://doi.org/10.1007/s10796-022-10308-y

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Liu, Y.-F., Kim, J., Wilson, C., & Bedny, M. (2020). Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. *eLife*, *9*, e59340. https://doi.org/10.7554/eLife.59340

Margulieux, L. E., Morrison, B. B., & Decker, A. (2020). Reducing withdrawal and failure rates in introductory programming with subgoal labeled worked examples. *International Journal of STEM Education*, *7*(1), 19. https://doi.org/10.1186/s40594-020-00222-7

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marvel, C. L., & Desmond, J. E. (2012). From storage to manipulation: How the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain and Language*, *120*(1), 42–51. https://doi.org/10.1016/j.bandl.2011.08.005

Mathur, A., Schultz, D., & Wang, Y. (2020). Neural bases of phonological and semantic processing in early childhood. *Brain Connectivity*, *10*(5), 212–223. https://doi.org/10.1089/brain.2019.0728

McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults.

Consciousness and Cognition, *20*(4), 1586–1593. https://doi.org/10.1016/j.concog.2011.08.005

McKee, A. (2023, October 12th). Coding best practices and guidelines for better code. *Datacamp*. https://www.datacamp.com/tutorial/coding-best-practices-and-guidelines

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(s2), 123–150. https://doi.org/10.1111/j.1467-9922.2010.00604.x

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. https://doi.org/10.1038/nn2024

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, M. B., Donovan, C.-L., Bennett, C. M., Aminoff, E. M., & Mayer, R. E. (2012). Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals. *NeuroImage*, *59*(1), 83–93. https://doi.org/10.1016/j.neuroimage.2011.05.060

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: Inner speech use during self-related processing. *Brain Research Bulletin*, *74*(6), 387–396. https://doi.org/10.1016/j.brainresbull.2007.06.013

Morin, A., Uttl, B., & Hamper, B. (2011). Self-reported frequency, content, and functions of inner speech. *Procedia - Social and Behavioral Sciences*, *30*, 1714–1718. https://doi.org/10.1016/j.sbspro.2011.10.331

Mottarella, M., Mortimore, K., & Prat, C. S. (2024). Exploring programming aptitude: Comparing the predictive utility of language aptitude subskills for Python and Java learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*, 2553-2559.

Munroe, W. (2023). Thinking through talking to yourself: Inner speech as a vehicle of conscious reasoning. *Philosophical Psychology*, *36*(2), 292–318. https://doi.org/10.1080/09515089.2022.2042505

Nedergaard, J. S. K., & Lupyan, G. (2024). Not everybody has an inner voice: Behavioral consequences of anendophasia. *Psychological Science*, *35*(7), 780–797. https://doi.org/10.1177/09567976241243004

Nedergaard, J. S. K., Wallentin, M., & Lupyan, G. (2023). Verbal interference paradigms: A systematic review investigating the role of language in cognition. *Psychonomic Bulletin & Review*, *30*(2), 464–488. https://doi.org/10.3758/s13423-022-02144-7

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2013). A meta-analysis of executive components of working memory. *Cerebral Cortex*, *23*(2), 264–282. https://doi.org/10.1093/cercor/bhs007

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646–1669. https://doi.org/10.1016/j.neuroimage.2012.06.065

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

Papagno, C., Comi, A., Riva, M., Bizzi, A., Vernice, M., Casarotti, A., Fava, E., & Bello, L. (2017). Mapping the brain network of the phonological loop. *Human Brain Mapping*, *38*(6), 3011–3024. https://doi.org/10.1002/hbm.23569

Papoutsi, M., De Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. *Cerebral Cortex*, *19*(9), 2156–2165. https://doi.org/10.1093/cercor/bhn239

Parkinson, J., & Cutts, Q. (2022). Relationships between an early-stage spatial skills test and final CS degree outcomes. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, 293–299. https://doi.org/10.1145/3478431.3499332

Peitek, N., Apel, S., Parnin, C., Brechmann, A., & Siegmund, J. (2021). Program comprehension and code complexity metrics: An fMRI study. *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, 524–536. https://doi.org/10.1109/ICSE43902.2021.00056

Peitek, N., Siegmund, J., Apel, S., Kastner, C., Parnin, C., Bethmann, A., Leich, T., Saake, G., & Brechmann, A. (2020). A look into programmers' heads. *IEEE Transactions on Software Engineering*, *46*(4), 442–462. https://doi.org/10.1109/TSE.2018.2863303

Pena, C. M., & Tirre, W. C. (1992). Cognitive factors involved in the first stage of programming skill acquisition. *Learning and Individual Differences, 4*(4), 311–334. https://doi.org/10.1016/1041-6080(92)90017-9

Pennington, N. (1987). Stimulus structures and mental representations in expert comprehension of computer programs. *Cognitive Psychology, 19*(3), 295–341. https://doi.org/10.1016/0010-0285(87)90007-7

Perrachione, T. K., Ghosh, S. S., Ostrovskaya, I., Gabrieli, J. D. E., & Kovelman, I. (2017). Phonological working memory for words and nonwords in cerebral cortex. *Journal of Speech, Language, and Hearing Research*, *60*(7), 1959–1979. https://doi.org/10.1044/2017_JSLHR-L-15-0446

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research, 261*, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Pollack, C., & Ashby, N. C. (2018). Where arithmetic and phonology meet: The meta-analytic convergence of arithmetic and phonological processing in the brain. *Developmental Cognitive Neuroscience, 30*, 251–264. https://doi.org/10.1016/j.dcn.2017.05.003

Prat, C. S., Madhyastha, T. M., Mottarella, M. J., & Kuo, C.-H. (2020). Relating natural language aptitude to individual differences in learning programming languages. *Scientific Reports, 10*(1), 3817. https://doi.org/10.1038/s41598-020-60661-8

Prat, C. S., Yamasaki, B. L., Kluender, R. A., & Stocco, A. (2016). Resting-state qEEG predicts rate of second language learning in adults. *Brain and Language, 157–158*, 44–50. https://doi.org/10.1016/j.bandl.2016.04.007

Prat, C. S., Yamasaki, B. L., & Peterson, E. R. (2019). Individual differences in resting-state brain rhythms uniquely predict second language learning rate and willingness to communicate in adults. *Journal of Cognitive Neuroscience*, *31*(1), 78–94. https://doi.org/10.1162/jocn_a_01337

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Quille, K., & Bergin, S. (2018). Programming: Predicting student success early in CS1. a re-validation and replication study. *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 15–20. https://doi.org/10.1145/3197091.3197101

Rakesh, D., Seguin, C., Zalesky, A., Cropley, V., & Whittle, S. (2021). Associations between neighborhood disadvantage, resting-state functional connectivity, and behavior in the adolescent brain cognitive development study: The moderating role of positive family and school environments. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(9), 877–886. https://doi.org/10.1016/j.bpsc.2021.03.008

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage*, *104*, 69–78. https://doi.org/10.1016/j.neuroimage.2014.09.045

Reynolds, M., & Besner, D. (2005). Basic processes in reading: A critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin & Review*, *12*(4), 622–646. https://doi.org/10.3758/BF03196752

Robins, A. (2010). Learning edge momentum: A new account of outcomes in CS1. *Computer Science Education*, *20*(1), 37–71. https://doi.org/10.1080/08993401003612167

Robison, M. K., & Unsworth, N. (2017). Individual differences in working memory capacity and resistance to belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *70*(8), 1471–1484. https://doi.org/10.1080/17470218.2016.1188406

Roebuck, H., & Lupyan, G. (2020). The Internal Representations Questionnaire: Measuring modes of thinking. *Behavior Research Methods*, *52*(5), 2053–2070. https://doi.org/10.3758/s13428-020-01354-y

Rokhman, M. F., Lintangsari, A. P., & Perdhani, W. C. (2020). EFL learners' phonemic awareness: A correlation between English phoneme identification skill toward word processing. *Journal of English Educators Society*, *5*(2), 135–141. https://doi.org/10.21070/jees.v5i2.467

Russo, A. (2019). Thinking out loud or speaking in loud: A review on inner speech. *Journal of Psychosocial Systems*, *3*(2), 53–65. https://doi.org/10.23823/jps.v3i2.60

Saeki, E. (2007). Phonological loop and goal maintenance: Effect of articulatory suppression in number-size consistency task. *Psychologia*, *50*(2), 122–131. https://doi.org/10.2117/psysoc.2007.122

Saiegh-Haddad, E. (2019). What is phonological awareness in L2? *Journal of Neurolinguistics*, *50*, 17–27. https://doi.org/10.1016/j.jneuroling.2017.11.001

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, *67*(3), 665–693. https://doi.org/10.1111/lang.12244

Sasaki, M. (2012). The Modern Language Aptitude Test (Paper-and-Pencil Version). *Language Testing*, *29*(2), 315–321. https://doi.org/10.1177/0265532211434015

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses.* Peter Lang.

Sauter, V. L. (1986). Predicting computer programming skill. *Computers & Education*, *10*(2), 299–302. https://doi.org/10.1016/0360-1315(86)90031-X

Scarlett, R. (2023, March 7th). Why Python keeps growing, explained. *Github Blog.* https://github.blog/developer-skills/programming-languages-and-frameworks/why-python-keeps-growing-explained/

Shneiderman, B., & Mayer, R. (1979). Syntactic/semantic interactions in programmer behavior: A model and experimental results. *International Journal of Parallel Programming*, *8*(3), 219–238. https://doi.org/10.1007/BF00977789

Sebastian, R., Laird, A. R., & Kiran, S. (2011). Meta-analysis of the neural representation of first language and second language. *Applied Psycholinguistics*, *32*(4), 799–819. https://doi.org/10.1017/S0142716411000075

Sedgewick, R. & Wayne, K. (2016). *Computer science: An interdisciplinary approach* (1st ed). Addison-Wesley Professional.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

Shute, V. J. (1991). Who is Likely to Acquire Programming Skills? *Journal of Educational Computing Research*, *7*(1), 1–24. https://doi.org/10.2190/VQJD-T1YD-5WVB-RYPJ

Shute, V. J., & Kyllonen, P. C. (1990). Modeling individual differences in programming skill acquisition. *Air Force Human Resources Laboratory, Air Force Systems Command*.

Seidenberg, M.S, & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Siegmund, J., Kästner, C., Apel, S., Parnin, C., Bethmann, A., Leich, T., Saake, G., & Brechmann, A. (2014). Understanding understanding source code with functional magnetic resonance imaging. *Proceedings of the 36th International Conference on Software Engineering*, 378–389. https://doi.org/10.1145/2568225.2568252

Siegmund, J., Peitek, N., Parnin, C., Apel, S., Hofmeister, J., Kästner, C., Begel, A., Bethmann, A., & Brechmann, A. (2017). Measuring neural efficiency of program comprehension. *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*, 140–150. https://doi.org/10.1145/3106237.3106268

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (ed.), 69–95.

Skehan, P. (2012). Language aptitude. In S. Gass & A. Mackey (eds.), 381–395

Snow, C. E., & Juel, C. (2005). Teaching Children to Read: What Do We Know about How to Do It? In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 501–520). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757642.ch26

Soloway, E., & Ehrlich, K. (1984). Empirical studies of programming knowledge. *IEEE Transactions on Software Engineering*, *10*(5), 595–609. https://doi.org/10.1109/TSE.1984.5010283

Stoeckel, C., Gough, P. M., Watkins, K. E., & Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex*, *45*(9), 1091–1096. https://doi.org/10.1016/j.cortex.2008.12.004

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Sun, Y., Zhuang, F., Zhu, H., Zhang, Q., He, Q., & Xiong, H. (2021). Market-oriented job skill valuation with cooperative composition neural network. *Nature Communications*, *12*(1), 1992. https://doi.org/10.1038/s41467-021-22215-y

Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human Brain Mapping*, *25*(1), 83–91. https://doi.org/10.1002/hbm.20134

TIOBE Software. (2024). TIOBE index. https://www.tiobe.com/tiobe-index/

Turker, S., Kuhnke, P., Eickhoff, S. B., Caspers, S., & Hartwigsen, G. (2023). Cortical, subcortical, and cerebellar contributions to language processing: A meta-analytic review of 403 neuroimaging experiments. *Psychological Bulletin*, *149*(11–12), 699–723. https://doi.org/10.1037/bul0000403

Turker, S., Seither-Preisler, A., & Reiterer, S. M. (2021). Examining individual differences in language learning: A neurocognitive model of language aptitude. *Neurobiology of Language*, *2*(3), 1–27. https://doi.org/10.1162/nol_a_00042

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726. https://doi.org/10.1038/35094573

U.S. Bureau of Labor Statistics. (2022). Computer and information technology field of degree. Occupational Outlook Handbook. https://www.bls.gov/ooh/field-of-degree/computer-and-information/computer-and-information-technology-field-of-degree.html

Van Deusen, A. (2023, January 31st). Python popularity: The rise of a global programming language. *Flatiron School.* https://flatironschool.com/blog/python-popularity-the-rise-of-a-global-programming-language/

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*(3), 181–198. https://doi.org/10.3758/bf03197716

Verhagen, J., Leseman, P., & Messer, M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning*, *65*(2), 417–448. https://doi.org/10.1111/lang.12101

Vygotsky, L. S. (1934/1987). Thinking and speech. *The collected works of Lev Vygotsky (Vol. 1).* Plenum Press.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067). https://doi.org/10.1038/nature04171

Wallace, G. L., Peng, C. S., & Williams, D. (2017). Interfering with inner speech selectively disrupts problem solving and is linked with real-world executive functioning. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3456–3460. https://doi.org/10.1044/2017_JSLHR-S-16-0376

Wang, L., Shi, D., Geng, F., Hao, X., Chanjuan, F., & Li, Y. (2022). Effects of cognitive control strategies on coding learning outcomes in early childhood. *The Journal of Educational Research*, *115*(2), 133–145. https://doi.org/10.1080/00220671.2022.2074946

Wen, Z. (Edward), Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31. https://doi.org/10.1017/S0261444816000276

Xu, S., Li, Y., & Liu, J. (2021). The neural correlates of computational thinking: Collaboration of distinct cognitive components revealed by fMRI. *Cerebral Cortex*, *31*(12), 5579–5597. https://doi.org/10.1093/cercor/bhab182

Yen, M., DeMarco, A. T., & Wilson, S. M. (2019). Adaptive paradigms for mapping phonological regions in individual participants. *NeuroImage*, *189*, 368–379. https://doi.org/10.1016/j.neuroimage.2019.01.040

Yoshioka, A., Tanabe, H. C., Nakagawa, E., Sumiya, M., Koike, T., & Sadato, N. (2023). The role of the left inferior frontal gyrus in introspection during verbal communication. *Brain Sciences*, *13*(1), 111. https://doi.org/10.3390/brainsci13010111

Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Ninaus, M., Reishofer, G., & Ischebeck, A. (2013). Individual differences in solving arithmetic word problems. *Behavioral and Brain Functions*, *9*(1), 28. https://doi.org/10.1186/1744-9081-9-28

Zhou, X., Li, M., Li, L., Zhang, Y., Cui, J., Liu, J., & Chen, C. (2018). The semantic system is involved in mathematical problem solving. *NeuroImage*, *166*, 360–370. https://doi.org/10.1016/j.neuroimage.2017.11.017

Tables

Table 1. Summary of correlations between language aptitude and learning to program in past studies

Study	Programming Language (hours learned)	Participants	Correlation: MLAT II (Phonological Coding) x Programming knowledge	Correlation: MLAT IV (Grammatical Sensitivity) x Programming knowledge	Correlation: MLAT V (Associative Memory) x Programming knowledge
Prat et al. (2020)	Python (7.5hrs)	English Monolinguals (N = 37)	$r = 0.46^{**}$	$r = 0.39^{**}$	$r = 0.39^{**}$
Prat et al. (in prep)	Python (16hrs)	English Monolinguals (N = 49) Chinese/English Bilinguals (N = 46)	$r = 0.45^{**}$	$r = 0.47^{**}$	$r = 0.37^{**}$
Prat et al. (in prep)	Java (8hrs)	Native English Speakers (N = 37)	$r = 0.49^{**}$	$r = 0.45^{**}$	$r = 0.24$

Note. $^{**}p < 0.01$, $^{*}p < 0.05$. Programming knowledge corresponds to declarative knowledge test score

Table 2. ROI search area coordinates and sizes

	Center MNI Coordinates			
Region	x	y	z	Radius
Left inferior frontal gyrus	-47	7	25	12
Right inferior frontal gyrus	47	7	25	12
Left parietal	-44	-41	43	10
Right parietal	44	-41	43	10
preSMA	-2	22	49	10

Table 3. Descriptive statistics for behavioral measures of interest

Measures	*N*	*M*	*SD*	*Range*
1. Python Acc (Comprehension Task)	46	0.78	0.09	0.53 – 0.97
2. Python Probe Acc (Comprehension Task)	46	0.76	0.15	0.38 – 1.0
3. Python Declarative Knowledge Test	49	51.65	9.26	24.0 – 65.0
4. Python Self-Rated Proficiency (PEAPQ)	49	5.42	2.05	2.0 – 10.0
5. Word Reading Acc (Comprehension Task)	46	0.96	0.07	0.59 – 1.0
6. Word Reading Probe Acc (Comprehension Task)	46	0.89	0.13	0.50 – 1.0
7. Nelson Denny Reading Comprehension	49	58.02	23.25	2.0 – 97.0
8. English Self-Rated Proficiency (LEAPQ)	49	9.24	1.22	5.0 – 10.0
9. MLAT II	48	25.65	2.74	19.0 – 30.0
10. Card Sort Verbal Bias Score	49	0.22	0.79	-1.0 – 1.0
11. IRQ Verbal	49	3.60	0.52	2.64 – 4.64
12. IRQ Visual	49	3.57	0.57	1.83 – 4.75
13. Forward Digit Span	49	6.47	1.49	4.0 – 9.0
14. Python RT (Comprehension Task)	46	5911.46	908.53	4248.34 – 7484.88
15. Python Probe RT (Comprehension Task)	46	1957.61	463.70	1074.12 – 3063.88
16. Word Reading RT (Comprehension Task)	46	3161.88	837.39	1598.50 – 5040.97
17. Word Reading Probe RT (Comprehension Task)	46	1697.79	399.62	1037.25 – 2673.62

Table 4. Bivariate correlations between behavioral measures of interest

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Python Acc	-															
2. Python Probe Acc	**0.34***	-														
3. Python Declarative Knowledge Test	**0.71****	**0.32***	-													
4. PEAPQ	**0.53****	0.19	**0.62****	-												
5. Word Reading Acc	0.03	0.26	0.15	-0.13	-											
6. Word Reading Probe Acc	0.06	0.06	0.02	-0.15	-0.02	-										
7. Nelson Denny Comprehension	-0.12	-0.11	0.15	0.10	**0.40***	0.13	-									
8. LEAPQ	-0.06	-0.06	-0.11	0.01	**0.48****	0.12	**0.57****	-								
9. MLAT II	0.14	-0.04	0.15	-0.04	0.19	0.06	0.03	-0.06	-							
10. Card Sort Verbal Bias	-0.05	**0.27**$^{+}$	-0.09	0.15	-0.06	-0.23	-0.15	-0.05	-0.19	-						
11. IRQ Verbal	-0.04	0.03	-0.11	-0.01	0.09	0.05	-0.05	0.09	0.13	-0.05	-					
12. IRQ Visual	0.09	0.02	-0.04	-0.18	0.24	-0.03	-0.05	0.13	0.09	-0.10	0.21	-				
13. Digit Span	0.23	-0.02	**0.28**$^{+}$	0.06	-0.10	-0.06	**-0.34***	**-0.40****	0.04	-0.15	**-0.29***	**-0.31***	-			
14. Python RT	-0.03	-0.05	**-0.26**$^{+}$	**-0.44****	-0.07	0.01	-0.12	-0.04	-0.01	-0.12	**-0.32***	0.05	0.03	-		
15. Python Probe RT	-0.21	-0.23	**-0.32***	**-0.27**$^{+}$	0.08	-0.15	0.07	0.09	-0.08	-0.02	-0.15	0.06	-0.09	**0.41***	-	
16. Word Reading RT	0.04	0.11	-0.07	-0.02	**-0.40***	-0.05	**-0.37***	**-0.30***	-0.25	-0.20	-0.18	**-0.31***	**0.34***	**0.49****	0.15	-
17. Word Reading Probe RT	0.07	0.09	0.08	0.003	-0.10	0.03	-0.07	-0.22	-0.09	-0.17	-0.13	-0.15	**0.30***	**0.36***	0.11	**0.54***

Note. $^{+}p < 0.10$, $^{*}p < 0.05$ uncorrected, $^{**}p < 0.05$ fdr corrected

Table 5. Phonological Localizer activity: Group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
(A) Rhyming > Letter Search						
Left inferior frontal gyrus	9	1723	5.20	-47	7	25
Left superior frontal gyrus / SMA	8	489	4.68	-2	22	49
Left inferior parietal lobule	40	594	4.59	-44	-41	43
Left inferior temporal gyrus	20	477	4.18	-50	-56	-11
Left inferior temporal gyrus	20	53	3.83	-32	-5	-38
Left superior temporal gyrus	42	69	3.86	-56	-32	7
Left cingulate	24	55	4.85	-5	-2	31
Left putamen	-	194	4.80	-20	4	7
Left cerebellum	-	64	5.07	-8	-74	-32
Left cerebellum	-	51	3.90	-23	-74	-44
Right inferior frontal gyrus	47	806	4.41	31	28	-2
Right inferior parietal lobule	40	91	3.75	46	-38	52
Right caudate	-	85	3.92	13	19	7
Right cerebellum	-	469	4.55	25	-74	-47
(B) Letter Search > Rhyming						
Left superior temporal gyrus	22	348	3.96	-56	-8	7
Left middle temporal gyrus	21	280	3.93	-50	7	-26
Left cerebellum	-	89	3.90	-26	-77	-32
Left cerebellum	-	49	3.66	-35	-44	-41
Right medial frontal gyrus	6	74	3.49	4	-8	73
Right middle frontal gyrus	8	264	4.16	28	31	46
Right inferior parietal lobule	40	1492	4.57	49	-44	25
Right superior parietal lobule	7	73	3.50	28	-65	46
Right precuneus	7	5626	4.66	1	-56	46
Right middle temporal gyrus	21	253	3.72	52	-2	-20
Right cerebellum	-	208	3.95	13	-44	-47
Right cerebellum	-	72	4.26	28	-80	-29

Note. FDR corrected $p < 0.05$ (t threshold = 3); extent threshold = 30 voxels

Table 6. Python > Fixation: Original group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
Left middle frontal gyrus	6	821	7.45	-44	1	37
Left superior frontal gyrus	6	323	6.70	-5	7	55
Left inferior frontal gyrus	47	93	8.00	-32	25	1
Left superior temporal gyrus	22	87	6.14	-59	-35	7
Left cerebellum	-	49	8.71	-17	-38	-44
Right middle frontal gyrus	9	327	6.37	46	10	37
Right middle frontal gyrus	6	146	6.59	34	-2	-61
Right inferior frontal gyrus	47	114	8.05	34	28	1
Right cingulate gyrus	24	96	6.89	7	1	31
Right posterior cingulate	30	31	5.87	25	-68	10
Right superior parietal	7	4488	7.68	28	-62	40
Right parahippocampal gyrus	28	950	7.22	25	-23	-5
Right fusiform gyrus	37	37	6.03	34	-35	-26
Right cerebellum	-	48	8.64	22	-35	-44

Note. FDR corrected $p < 0.001$ (t threshold = 5); extent threshold = 30 voxels

Table 7. Original group-level GLM model beta weight contrasts correlated with behavior

ROI	Python Declarative Knowledge Test	Nelson Denny Comprehension Test	MLAT II	Forward Digit Span	IRQ Verbal	IRQ Visual	Card Sort Verbal Bias
(A) Python > Fixation							
preSMA	-0.08	0.001	-0.09	-0.05	-0.10	**-0.33***	0.19
Left IFG	-0.06	-0.05	-0.22	-0.18	0.16	**-0.26**$^{+}$	0.06
Left Parietal	0.005	-0.02	**-0.29**$^{+}$	-0.16	-0.15	**-0.33***	0.23
Right IFG	0.06	0.10	-0.05	-0.10	-0.20	-0.12	0.17
Right Parietal	0.02	0.003	-0.03	**0.28**$^{+}$	0.007	-0.19	0.07
(B) Word Reading > Fixation							
preSMA	-0.13	**-0.37***	-0.20	-0.04	-0.07	-0.19	0.09
Left IFG	-0.06	**-0.31***	**-0.29**$^{+}$	0.13	0.04	-0.25	0.06
Left Parietal	**-0.29**$^{+}$	**-0.48*****	**-0.29**$^{+}$	-0.09	0.06	-0.20	-0.01
Right IFG	-0.01	-0.04	-0.18	-0.02	-0.24	0.03	0.14
Right Parietal	**-0.26**$^{+}$	**-0.35***	-0.21	**0.31***	-0.03	-0.18	0.08
(C) Python > Word Reading							
preSMA	0.03	**0.32***	0.07	-0.02	-0.04	-0.17	0.11
Left IFG	-0.02	0.22	-0.003	**-0.33***	0.16	-0.08	0.02
Left Parietal	**0.28**$^{+}$	**0.45*****	0.01	-0.05	-0.19	-0.11	0.23
Right IFG	0.10	0.19	0.10	-0.13	-0.07	-0.20	0.12
Right Parietal	0.25	**0.31***	0.14	0.12	0.03	-0.10	0.03

Note. $^{+}p < 0.10$, $^{*}p < 0.05$ uncorrected, $^{**}p < 0.01$ uncorrected, $^{***}p < 0.05$ fdr corrected

Table 8. Python > Fixation: whole-brain correlations: Original GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean r value	Peak MNI coordinates		
				x	*y*	*z*
(A) Python Declarative Knowledge Test						
Right precuneus	7	36	0.36	13	-80	58
Right middle occipital gyrus	18	64	0.36	37	-86	-2
Left premotor	6	67	-0.37	-5	-8	55
Left superior frontal gyrus	8	38	-0.37	-5	31	43
Left middle frontal gyrus	46	65	-0.37	-53	31	22
Right postcentral gyrus	3	145	-0.37	43	-23	52
Right putamen	-	52	-0.35	28	7	4
Right cerebellum	-	47	-0.36	10	-89	-38
(B) Nelson Denny Comprehension						
Left inferior frontal gyrus	45	39	0.37	-50	34	4
Left middle temporal gyrus	21	52	0.36	-62	-29	-2
Left precuneus	31	72	0.37	-8	-71	28
Right superior temporal gyrus	22	47	0.37	46	-23	-5
Left precentral gyrus	6	36	-0.37	-62	-17	43
Left superior parietal lobule	7	64	-0.38	-26	-50	58
(C) MLAT II						
Left superior frontal gyrus	6	90	0.36	-11	25	64
Left precuneus	31	33	0.37	-8	-62	22
Left angular gyrus	39	57	0.35	-50	-74	37
Right angular gyrus	39	90	0.38	61	-62	37
Left superior frontal gyrus	6	86	-0.37	-35	1	70
Left middle frontal gyrus	46	44	-0.36	-41	31	25
Left pre SMA	8	32	-0.37	-5	22	49
Left supramarginal gyrus	40	170	-0.36	-56	-41	46
Left middle occipital gyrus	19	42	-0.38	-29	-83	22
Right precentral gyrus	4	48	-0.38	28	-23	55
Right postcentral gyrus	1	44	-0.36	55	-20	55
Right supramarginal gyrus	40	34	-0.35	55	-41	55
Right putamen	-	186	-0.36	19	13	-8
Right cerebellum	-	43	-0.36	40	-65	-50
Right cerebellum	-	42	-0.37	13	-80	-50
(D) Forward Digit Span						
Right cuneus	19	37	0.35	10	-95	31
Left superior parietal lobule	7	88	-0.37	-35	-71	49
Left thalamus	-	93	-0.36	-2	-26	10
Left cerebellum	-	58	-0.35	-29	-29	-38
Right visual cortex	18	61	-0.35	4	-77	1
Right cerebellum	-	108	-0.37	4	-95	-20

(E) IRQ - Verbal						
Left angular gyrus	39	33	0.36	-44	-62	37
Left precuneus	7	126	0.38	-5	-62	67
Right cerebellum	-	31	0.37	7	-95	-38
(F) IRQ - Visual						
Left superior parietal	7	56	0.37	-47	-74	44
Left visual cortex	18	30	0.35	-17	-80	-11
Right visual cortex	18	56	0.37	16	-74	-8
Right visual cortex	17	51	0.35	7	-95	-17
Left superior frontal gyrus	11	109	-0.37	-20	49	-14
Left middle frontal gyrus	19	119	-0.37	-17	31	40
Left supramarginal gyrus	40	77	-0.36	-62	-32	49
Right superior frontal gyrus	6	36	-0.36	4	10	61
Right inferior frontal gyrus	44	46	-0.38	58	13	10
Right anterior cingulate	32	31	-0.38	10	34	16
Right precentral gyrus	4	44	-0.35	40	-20	61
Right postcentral gyrus	1	156	-0.37	55	-20	49
Right middle temporal gyrus	19	42	-0.35	43	-77	19
Right middle temporal gyrus	21	41	-0.35	61	-59	1
Right inferior temporal gyrus	20	30	-0.38	55	-23	-35
Right thalamus	-	32	-0.36	16	-8	7
(G) Card Sorting Verbal Bias						
Left superior temporal gyrus	41	20	0.37	-53	-20	7
Left cerebellum	-	23	0.36	-44	-74	-35
Left cerebellum	-	62	0.35	-45	-72	-35
Cingulate gyrus	31	247	0.37	-20	-38	40
Right superior frontal gyrus	9	28	0.35	43	37	28
Right inferior parietal lobule	40	73	0.36	64	-32	46
Right supramarginal gyrus	40	43	0.36	55	-14	13
Right cerebellum	-	43	0.36	34	-38	-41
Right cerebellum	-	253	0.37	40	-77	-23
Right cerebellum	-	24	0.34	28	-41	-59

Note. Uncorrected threshold: $r > 0.3$ ($p < 0.05$); extent threshold = 30 voxels

Table 9. Scrambled Word Reading > Fixation: Original group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
Left middle frontal gyrus	6	420	6.55	-44	1	37
Left inferior frontal gyrus	45	71	6.68	-29	28	7
Left middle temporal gyrus	21	50	5.70	-47	-44	7
Left superior parietal lobule	7	302	6.39	-23	-65	46
Left parietal	7	58	6.22	-29	-71	50
Left visual cortex	18	844	7.36	-14	-101	-5
Left caudate	-	112	6.34	-17	-5	22
Left thalamus	-	42	6.12	-11	-11	10
Right middle frontal gyrus	9	226	6.04	46	10	34
Right superior frontal gyrus	6	346	6.43	1	13	52
Right superior frontal gyrus	6	94	6.21	34	-2	67
Right inferior frontal gyrus	47	128	6.89	34	28	4
Right cingulate gyrus	23	30	6.29	4	-26	28
Right posterior cingulate	30	30	5.85	22	-59	7
Right precuneus	19	465	6.89	31	-65	40
Right cuneus	17	858	7.48	13	-95	-2
Right parahippocampal gyrus	28	79	7.14	25	-23	-5
Right caudate	-	106	6.11	16	-2	19
Right cerebellum	-	44	7.21	1	-53	-35
Right cerebellum	-	96	6.36	28	-62	-26

Note. FDR corrected $p < 0.001$ (*t* threshold = 5); extent threshold = 30 voxels

Table 10. Python > Fixation: RT group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
Left middle frontal gyrus	6	654	7.21	-44	1	37
Left insula	13	71	7.06	-29	28	4
Left parahippocampal gyrus	28	993	7.32	-23	-23	-5
Left superior temporal gyrus	22	91	6.27	-59	-35	7
Left superior parietal lobule	7	1740	7.64	-23	-62	40
Left putamen	-	231	6.73	-20	10	4
Left cerebellum	-	61	8.64	-17	-38	-44
Left cerebellum	-	100	6.77	-26	-59	-29
SMA	6	322	6.67	-8	10	52
Cingulate gyrus	24	147	6.98	7	1	31
Right middle frontal gyrus	9	200	6.16	46	10	37
Right insula	13	88	7.43	34	28	1
Right premotor cortex	6	127	6.56	25	-2	52
Right fusiform gyrus	37	49	6.06	37	-38	-23
Right superior parietal lobule	7	1422	7.64	28	-62	40
Right cerebellum	-	703	7.29	1	-53	-38
Right cerebellum	-	48	8.69	19	-38	-44

Note. FDR corrected $p < 0.001$ (t threshold = 5); extent threshold = 30 voxels

Table 11. Scrambled Word Reading > Fixation: RT group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
Left inferior frontal gyrus	45	114	7.06	-29	28	4
Left superior temporal gyrus	42	89	6.17	-56	-32	7
Left lingual gyrus	19	39	5.89	-23	-62	4
Left visual cortex	17	1750	7.47	-14	-95	-5
Left cerebellum	-	50	8.41	-20	-35	-44
Striatum / subcortical	-	840	6.89	25	-26	-2
Cingulate gyrus	24	156	7.06	7	1	31
Posterior cingulate	30	39	5.85	16	-62	10
preSMA	6	558	6.93	4	10	52
Right superior frontal gyrus	6	106	6.13	34	-2	67
Right middle frontal gyrus	46	42	5.49	46	40	16
Right middle frontal gyrus	9	141	6.07	46	10	37
Right postcentral gyrus	1	115	5.82	58	-17	52
Right insula	13	181	7.47	34	25	10
Right visual cortex	18	1001	7.34	22	-89	-8
Right cerebellum	-	43	8.57	19	-38	-44
Right cerebellum	-	684	7.04	1	-53	-38
Right cerebellum	-	188	6.48	31	-62	-26

Note. FDR corrected $p < 0.001$ (t threshold = 5); extent threshold = 30 voxels

Table 12. RT GLM model beta weight contrasts correlated with behavior

ROI	Python Declarative Knowledge Test	Nelson Denny Comprehension	MLAT II	Forward Digit Span	IRQ Verbal	IRQ Visual	Card Sort Verbal Bias
(A) Python > Fixation							
preSMA	-0.01	-0.02	-0.06	-0.13	-0.07	**-0.28**$^{+}$	**0.26**$^{+}$
Left IFG	-0.01	-0.09	-0.10	-0.17	0.23	-0.22	0.14
Left Parietal	-0.15	-0.15	**-0.26**$^{+}$	**-0.38***	0.06	-0.09	**0.31***
Right IFG	0.03	0.11	0.05	-0.24	-0.16	0.001	0.24
Right Parietal	-0.16	-0.05	0.03	0.12	0.11	-0.04	0.15
(B) Word Reading > Fixation							
preSMA	-0.21	-0.19	-0.17	-0.17	0.05	-0.17	0.04
Left IFG	-0.21	-0.12	-0.17	-0.05	0.18	-0.14	0.05
Left Parietal	-0.07	**-0.34***	-0.21	-0.22	0.14	-0.04	0.05
Right IFG	0.06	0.20	-0.01	-0.20	-0.13	0.08	0.08
Right Parietal	**-0.32***	-0.16	-0.10	0.21	0.06	-0.10	0.09
(C) Python > Word Reading							
preSMA	**0.26**$^{+}$	0.18	0.15	0.08	-0.14	-0.08	0.24
Left IFG	**0.28**$^{+}$	0.06	0.13	-0.13	0.01	-0.06	0.10
Left Parietal	-0.05	**0.33***	0.05	-0.05	-0.14	-0.03	0.22
Right IFG	-0.06	-0.20	0.12	-0.03	-0.02	-0.15	**0.29**$^{+}$
Right Parietal	**0.30**$^{+}$	0.22	**0.26**$^{+}$	-0.16	0.11	0.13	0.11

Note. $^{+}p < 0.10$, $^{*}p < 0.05$ uncorrected, $^{**}p < 0.01$ uncorrected, $^{***}p < 0.05$ fdr corrected

Table 13. Python > Fixation whole-brain correlations: RT GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean r value	Peak MNI coordinates		
				x	*y*	*z*
(A) Python Declarative Knowledge Test						
Left superior frontal gyrus	8	21	0.36	-29	49	40
Left cuneus	19	45	0.36	-11	-98	34
Right medial frontal gyrus	10	26	0.36	4	52	7
Left medial frontal gyrus	8	53	-0.36	-8	31	43
Left middle frontal gyrus	46	93	-0.36	-47	22	28
Left postcentral gyrus	3	41	-0.34	-20	-29	55
Left superior parietal lobule	7	63	-0.36	-35	-68	49
Left basal ganglia	-	100	-0.35	-26	1	4
Left cerebellum	-	101	-0.36	-20	-56	-17
Right middle frontal gyrus	46	21	-0.37	49	34	28
Right medial frontal gyrus	6	91	-0.38	7	-2	55
Right postcentral gyrus	3	256	-0.37	43	-23	55
Right basal ganglia	-	71	-0.36	25	-11	1
Right cerebellum	-	25	-0.34	16	-92	-35
Right cerebellum	-	27	-0.35	37	-71	-47
(B) Nelson Denny Comprehension						
Left middle temporal gyrus	21	27	0.37	-59	-29	1
Left precuneus	31	63	0.36	-8	-71	28
Right inferior frontal gyrus	47	56	0.35	43	25	-8
Right inferior parietal lobule	40	25	0.35	64	-44	22
Right superior temporal gyrus	21	106	0.38	52	-8	-11
Left middle frontal gyrus	6	31	-0.35	-29	1	58
Left superior parietal lobule	7	32	-0.39	-26	-50	58
Right cerebellum	-	34	-0.33	46	-41	-32
(C) MLAT II						
Left superior frontal gyrus	8	24	0.36	-8	46	52
Left precuneus	31	21	0.34	-8	-62	22
Left angular gyrus	39	57	0.35	-50	-74	37
Right angular gyrus	39	47	0.35	55	-65	40
Left medial frontal gyrus	10	31	-0.35	-17	61	-8
Left middle frontal gyrus	6	47	-0.37	-23	-5	46
Left middle frontal gyrus	9	50	-0.35	-32	28	22
Left superior frontal gyrus	8	46	-0.36	-5	22	49
Left postcentral gyrus	4	23	-0.34	-14	-32	58
Left supramarginal gyrus	40	236	-0.37	-56	-41	46
Left temporal gyrus	41	30	-0.37	-53	-14	13
Left middle occipital gyrus	19	46	-0.37	-26	-80	16

Left putamen	-	71	-0.35	-20	13	1
Left cerebellum	-	34	-0.35	-5	-77	-44
Left cerebellum	-	38	-0.35	-32	-44	-50
Right superior frontal gyrus	6	30	-0.38	13	-2	67
Right middle frontal gyrus	10	21	-0.35	40	46	22
Right postcentral gyrus	1	168	-0.37	55	-20	55
Right caudate	-	174	-0.36	13	13	-2
Right cerebellum	-	29	-0.36	13	-80	-50
Right cerebellum	-	129	-0.35	40	-65	-50
Right cerebellum	-	76	-0.35	34	-53	-26
(D) Forward Digit Span						
Right middle occipital gyrus	19	28	0.34	49	-83	16
Left inferior frontal gyrus	45	104	-0.36	-50	31	4
Left middle frontal gyrus	6	52	-0.38	-38	7	58
Left precentral gyrus	6	125	-0.36	-56	1	37
Left postcentral gyrus	3	179	-0.37	-44	-23	67
Left superior temporal gyrus	22	35	-0.36	-47	-38	4
Left superior parietal lobule	7	140	-0.39	-35	-71	49
Left precuneus	31	30	-0.36	-8	-68	25
Left cuneus	18	78	-0.36	-23	-95	1
Left visual cortex	18	56	-0.36	-17	-89	-11
Supplemental motor area	6	125	-0.37	-5	-8	58
Right middle frontal gyrus	8	37	-0.39	58	13	46
Right precentral gyrus	6	55	-0.36	46	-8	55
Right postcentral gyrus	5	44	-0.36	22	-41	76
Right middle temporal gyrus	21	22	-0.37	49	-29	-2
Right superior temporal gyrus	39	29	-0.37	49	-50	7
Right fusiform gyrus	37	56	-0.35	58	-47	-23
Right superior parietal lobule	7	32	-0.38	34	-62	49
Right superior parietal lobule	7	29	-0.38	22	-65	67
Right precuneus	31	75	-0.36	22	-56	22
Right visual cortex	18	57	-0.37	25	-104	-8
Right cerebellum	-	1519	-0.37	13	-53	-32
Right cerebellum	-	47	-0.38	7	-65	-50
Right cerebellum	-	39	-0.35	40	-53	-29
(E) IRQ - Verbal						
Left superior temporal gyrus	38	20	0.37	-23	10	-23
Left middle temporal gyrus	39	27	0.35	-59	-59	10
Left angular gyrus	39	27	0.36	-38	-65	40
Left postcentral gyrus	40	20	0.37	-50	-23	16
Left precuneus	7	71	0.37	-5	-62	67
Left posterior cingulate	31	33	0.36	-20	-65	19
Left fusiform gyrus	37	43	0.36	-26	-65	-11
Left middle occipital gyrus	19	34	0.38	-29	-89	19

Left thalamus	-	92	0.36	-5	-23	13
Left cerebellum	-	40	0.39	-11	-56	-35
PreSMA	6	44	0.38	-8	-5	61
Right middle frontal gyrus	6	34	0.36	52	13	52
Right precentral gyrus	6	44	0.36	64	-2	19
Right supramarginal gyrus	40	22	0.36	46	-35	49
Right parahippocampal gyrus	36	32	0.38	10	-44	-5
Right thalamus	-	120	0.36	13	-20	10
Right cerebellum	-	43	0.38	1	-80	-23
(F) IRQ - Visual						
Left visual cortex	18	62	0.35	-17	-80	-11
Right visual cortex	18	69	0.35	16	-101	-17
Right visual cortex	18	74	0.38	16	-74	-8
Left middle frontal gyrus	8	22	-0.38	-35	16	43
Left middle frontal gyrus	8	27	-0.35	-17	31	40
Left middle frontal gyrus	11	45	-0.37	-23	40	-8
Left medial frontal gyrus	6	54	-0.36	-5	40	37
Left supramarginal gyrus	40	83	-0.36	-62	-32	49
Left cuneus	17	28	-0.37	-5	-104	-2
Right middle frontal gyrus	9	23	-0.37	40	19	34
Right postcentral gyrus	4	71	-0.35	55	-14	34
Right middle temporal gyrus	21	25	-0.35	61	-41	1
(G) Card Sorting Verbal Bias						
Left middle frontal gyrus	8	123	0.37	-29	40	52
Left inferior frontal gyrus	47	24	0.37	-47	28	-14
Left medial frontal gyrus	6	24	0.34	-14	4	64
Left angular gyrus	39	81	0.36	-53	-68	19
Left inferior parietal lobule	40	2413	0.36	-50	-56	55
Left supramarginal gyrus	40	93	0.36	-50	-32	55
Left inferior temporal gyrus	20	26	0.35	-53	4	-44
Left cerebellum	-	247	0.37	-47	-41	-50
Right middle frontal gyrus	10	28	0.36	34	46	4
Right inferior frontal gyrus	47	24	0.36	55	19	-2
Right insula	13	37	0.36	40	19	1
Right middle temporal gyrus	21	81	0.36	67	7	-17
Right precuneus	19	55	0.35	25	-77	40
Right cerebellum	-	834	0.36	40	-92	-44

Note. Uncorrected threshold: $r > 0.3$ ($p < 0.05$); extent threshold = 20 voxels

Table 14. Comparison of GLM model fit by regressor and ROI

	Python		Scrambled Word Reading		Fixation	
ROI	Beta	SE	Beta	SE	Beta	SE
Left inferior frontal gyrus						
Original Model	0.660	0.021	0.141	0.021	-0.541	0.020
RT Model	0.481	0.014	0.384	0.019	-0.638	0.009
Left parietal						
Original Model	0.642	0.022	0.091	0.022	-0.440	0.020
RT Model	0.407	0.014	0.260	0.019	-0.553	0.009
Right inferior frontal gyrus						
Original Model	0.479	0.022	0.094	0.022	-0.490	0.021
RT Model	0.311	0.015	0.375	0.019	-0.538	0.009
Right parietal						
Original Model	0.397	0.023	-0.073	0.023	-0.524	0.021
RT Model	0.203	0.015	0.082	0.020	-0.532	0.010
preSMA						
Original Model	0.096	0.022	-0.298	0.022	-0.725	0.021
RT Model	0.111	0.015	0.090	0.020	-0.513	0.010

Notes. SE denotes standard error, where a lower SE indicates better model fit

Table 15. Python > Scrambled Word Reading: Original group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
(A) Python > Word Reading						
Left inferior frontal gyrus	9	1160	4.78	-50	10	34
Left superior frontal gyrus	8	239	4.14	-5	22	49
Left inferior frontal gyrus	45	75	4.71	-29	28	4
Left superior parietal lobule	7	4978	5.00	-23	-68	43
Left caudate	-	425	4.02	-14	1	19
Left cerebellum	-	50	4.01	-29	-74	-50
SMA	6	534	4.30	28	10	55
Right inferior frontal gyrus	47	84	4.80	34	28	1
Right cerebellum	-	63	4.63	1	-53	-35
(B) Word Reading > Python						
Left posterior insula	13	592	4.15	-38	-20	1
Left postcentral gyrus	5	82	4.12	-26	-41	58
Left cerebellum	-	42	3.61	-26	-89	-32
Anterior cingulate	32	588	3.77	-2	46	-5
Right posterior insula	13	787	4.21	40	-5	-11
Right inferior parietal lobule	40	72	3.74	34	-41	52
Right precuneus	7	392	3.97	16	-41	49
Right precuneus	31	82	3.82	10	-53	31
Right cuneus	19	516	4.15	7	-89	34
Right middle temporal gyrus	38	84	3.73	43	13	-41
Right middle temporal gyrus	21	72	3.71	64	-14	-11
Right occipital	18	114	3.97	7	-74	1

Note. FDR corrected $p < 0.05$ (t threshold = 3); extent threshold = 30 voxels

Table 16. Python > Scrambled Word Reading: RT group-level GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean t value	Peak MNI coordinates		
				x	*y*	*z*
(A) Python > Word Reading						
Left middle frontal gyrus	6	520	4.27	-26	10	58
Left inferior frontal gyrus	46	57	3.55	-44	43	7
Left middle temporal gyrus	37	241	4.13	-56	-50	-5
Left precuneus	19	901	4.23	-38	-77	40
Left fusiform gyrus	19	293	4.42	-26	-68	-8
Left posterior cingulate	30	64	4.63	-17	-56	19
Right postcentral gyrus	1	31	3.57	67	-23	46
Right premotor cortex	6	51	3.60	22	4	52
Right inferior temporal gyrus	20	95	3.88	58	-56	-11
Right supramarginal gyrus	40	37	3.45	37	-41	43
Right precuneus	19	186	3.81	37	-77	43
Right fusiform gyrus	37	151	4.64	34	-38	-11
Right cingulate gyrus	31	57	4.60	22	-53	22
(B) Word Reading > Python						
Left middle frontal gyrus	9	120	3.80	-29	43	34
Left postcentral gyrus	2	500	3.81	-47	-26	52
Left insula	13	390	4.44	-38	10	13
Left putamen	-	30	3.65	-26	-2	13
Left middle temporal gyrus	22	40	4.87	-62	-29	4
Left superior temporal gyrus	22	43	3.75	-41	-11	-8
Left visual cortex	18	261	4.33	-17	-95	1
Left cerebellum	-	266	3.86	-32	-50	-32
Left cerebellum	-	407	3.94	-5	-74	-26
Right inferior frontal gyrus	47	536	4.51	43	22	1
Right superior frontal gyrus	9	335	3.91	34	49	28
Right precentral gyrus	6	204	3.77	37	-5	55
Right insula	13	306	3.79	46	-23	22
Right thalamus	-	39	3.51	16	-11	4
Right cingulate gyrus	23	1364	4.24	4	-20	31
Right middle temporal gyrus	21	133	3.98	49	-29	-2
Right visual cortex	17	215	4.08	16	-95	-2
Right visual cortex	18	285	3.68	7	-74	16
Right precuneus	7	30	3.43	13	-68	40
Right cerebellum	-	78	3.68	31	-50	-53

Note. FDR corrected $p < 0.05$ (t threshold = 3); extent threshold = 30 voxels

Table 17. Python > Scrambled Word Reading whole-brain correlations: RT GLM model results

Peak cortical region	Brodmann's area	Cluster size	Mean r value	Peak MNI coordinates		
				x	*y*	*z*
(A) Python Declarative Knowledge Test						
Right middle frontal gyrus	10	32	0.37	37	55	7
Right cingulate gyrus	31	22	0.36	7	-32	34
Right superior temporal gyrus	22	25	-0.36	61	-41	13
(B) Nelson Denny Comprehension						
Right inferior parietal lobule	40	30	0.35	49	-50	46
Left precuneus	7	25	-0.35	-2	-53	58
Left visual cortex	18	29	-0.35	-8	-95	-5
Posterior cingulate gyrus	30	29	-0.35	19	-65	13
Right superior frontal gyrus	8	22	-0.36	28	46	49
(C) MLAT II						
Left inferior frontal gyrus	46	186	0.36	-50	31	16
Left middle frontal gyrus	8	198	0.36	-14	34	43
Left superior temporal gyrus	41	30	0.37	-56	-17	7
Left fusiform gyrus	37	26	0.36	-32	-29	-26
Right middle frontal gyrus	9	91	0.36	49	31	28
(D) Forward Digit Span						
Left visual cortex	17	27	0.35	-14	-92	-14
Right inferior temporal gyrus	20	27	0.38	37	-2	-47
Left middle frontal gyrus	6	42	-0.35	-47	10	58
Left superior temporal gyrus	22	20	-0.35	-38	-56	13
Left middle occipital gyrus	19	62	-0.35	-44	-74	-2
Right middle frontal gyrus	10	52	-0.36	49	52	25
(E) IRQ - Verbal						
Left supramarginal gyrus	40	21	-0.36	-62	-47	25
Left middle occipital gyrus	18	22	-0.36	-26	-92	4
Left cuneus	19	50	-0.35	-2	-83	37
Left cerebellum	-	41	-0.37	-26	-77	-32
Right anterior prefrontal cortex	10	29	-0.34	19	-35	-23
Right middle frontal gyrus	10	31	-0.37	1	55	10
Right middle frontal gyrus	8	22	-0.35	25	34	37
Right inferior parietal lobule	40	111	-0.38	52	-44	40
Right superior parietal lobule	7	38	-0.39	28	-65	49
Right inferior temporal gyrus	20	31	-0.38	67	-26	-26
Right precuneus	31	125	-0.36	10	-62	22
Right visual cortex	18	20	-0.35	31	-56	-17
Right visual cortex	18	165	-0.37	10	-80	-14
Right caudate	-	24	-0.36	13	7	16
Right cerebellum	-	48	-0.37	4	-68	-20

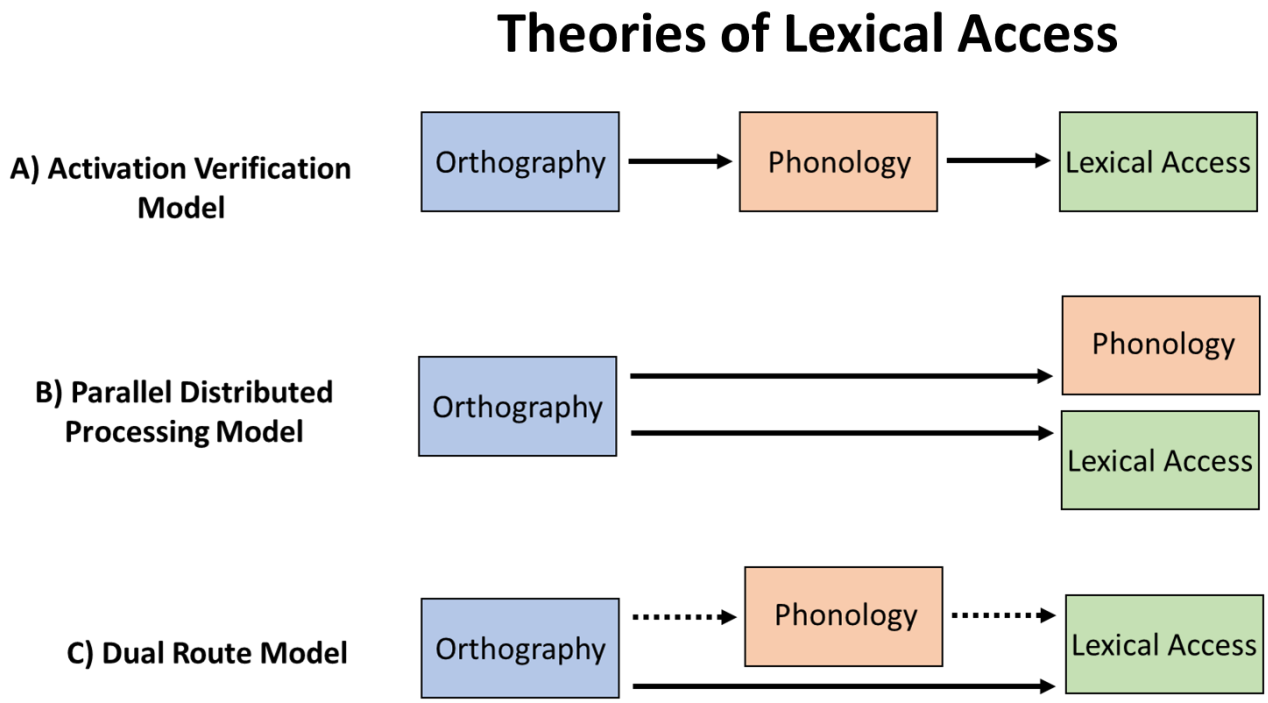
Right cerebellum	-	23	-0.35	25	-89	-44
Right cerebellum	-	44	-0.35	40	-68	-41
(F) IRQ - Visual						
Left medial frontal gyrus	6	24	-0.35	-5	40	34
Left postcentral gyrus	2	50	-0.35	-47	-29	40
Left inferior parietal lobule	40	30	-0.37	-53	-38	40
Left precuneus	31	21	-0.37	-2	-68	25
Left visual cortex	18	31	-0.38	-23	-80	-5
Right inferior parietal lobule	40	30	-0.37	-53	-38	40
(G) Card Sorting Verbal Bias						
Left superior frontal gyrus	10	102	0.37	-20	76	7
Left middle frontal gyrus	8	50	0.37	-38	40	40
Left medial frontal gyrus	9	27	0.36	-8	46	19
Left angular gyrus	39	85	0.37	-53	-53	43
Left putamen	-	103	0.36	-20	19	1
Right superior frontal gyrus	9	54	0.36	19	37	37
Right cerebellum	-	40	0.37	31	-86	-20
Right cerebellum	-	48	0.37	34	-38	-38

Note. Uncorrected threshold: $r > 0.3$ ($p < 0.05$); extent threshold = 20 voxels

Figures

Figure 1

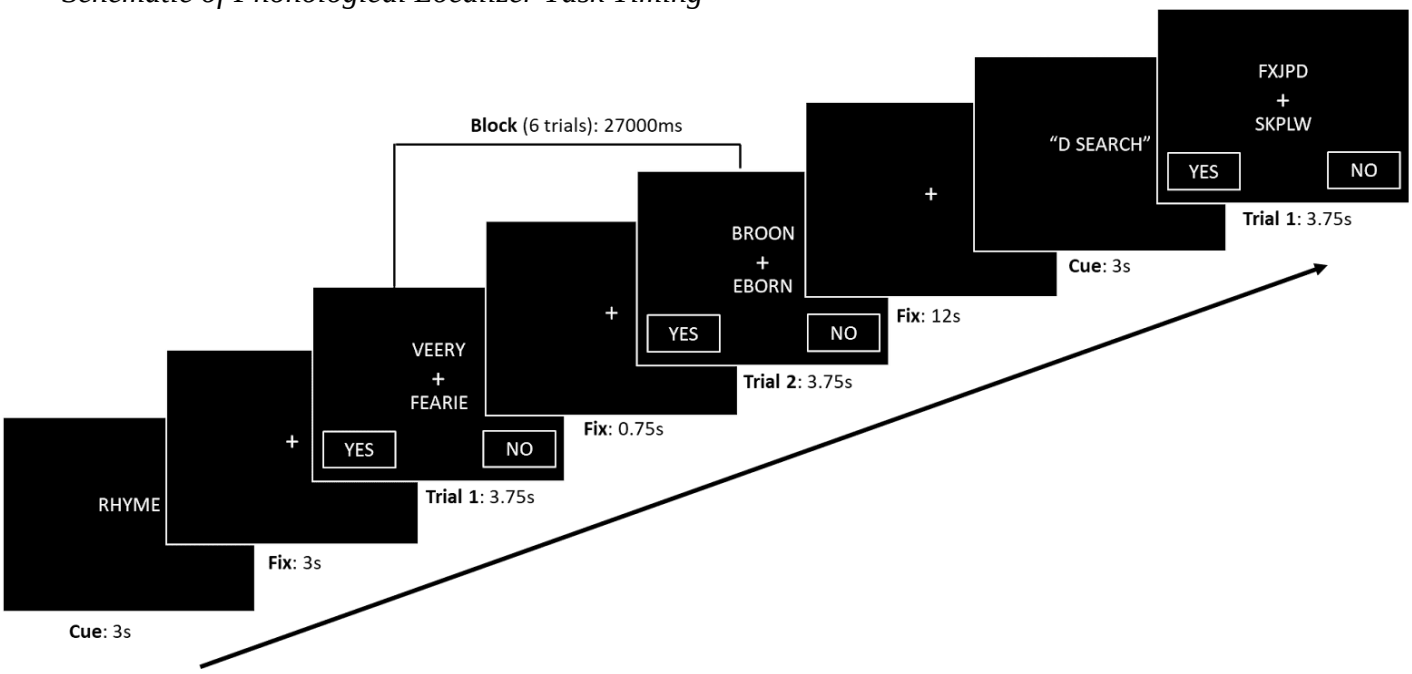
Schematic of Theories of Lexical Access



Note. A) Depicts the Activation Verification Model (Van Orden, 1987) where phonology is a necessary intermediate step that gives rise to lexical access. B) Depicts the Parallel Distributed Processing Model (Sidenberg & McClelland, 1989) where the phonological representation is activated in parallel with the lexical representation. C) Depicts the Dual Route Model (Coltheart, 1980) whereby phonology is activated during contexts where the reader is less-skilled or the words are more unfamiliar (dashed line), in more skilled readers the orthographic representation can directly by converted to a lexical representation (solid line).

Figure 2

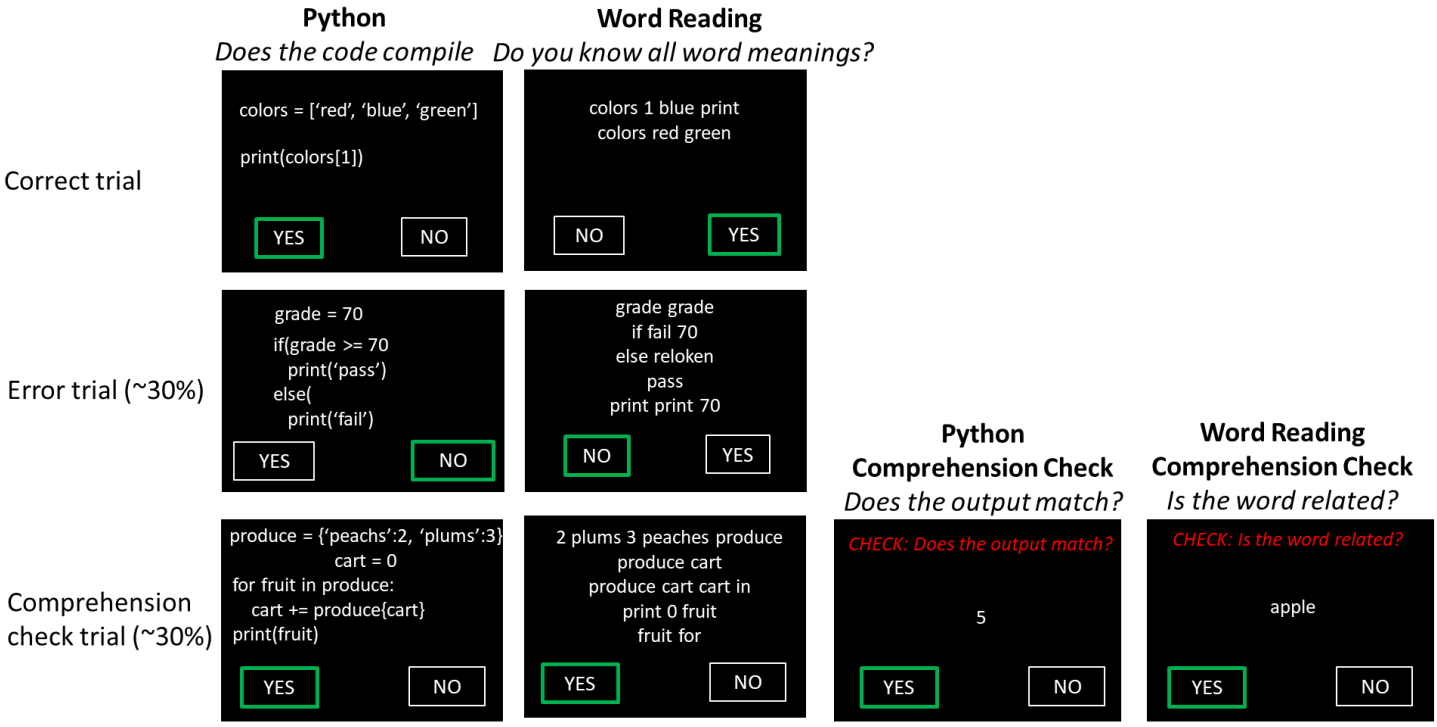
Schematic of Phonological Localizer Task Timing



Note. Schematic depicting the Phonological Localizer MRI task. For both task conditions participants viewed a cue alerting them to upcoming block condition (i.e., "Rhyme" or "D Search"). Then participants completed a block of six trials where they had to determine if the nonwords rhymed (Rhyming trials) or if the letter "D" was present in both of the consonant strings (Letter Search trials). Between each block there was a longer fixation period to allow the hemodynamic response to return to baseline.

Figure 3

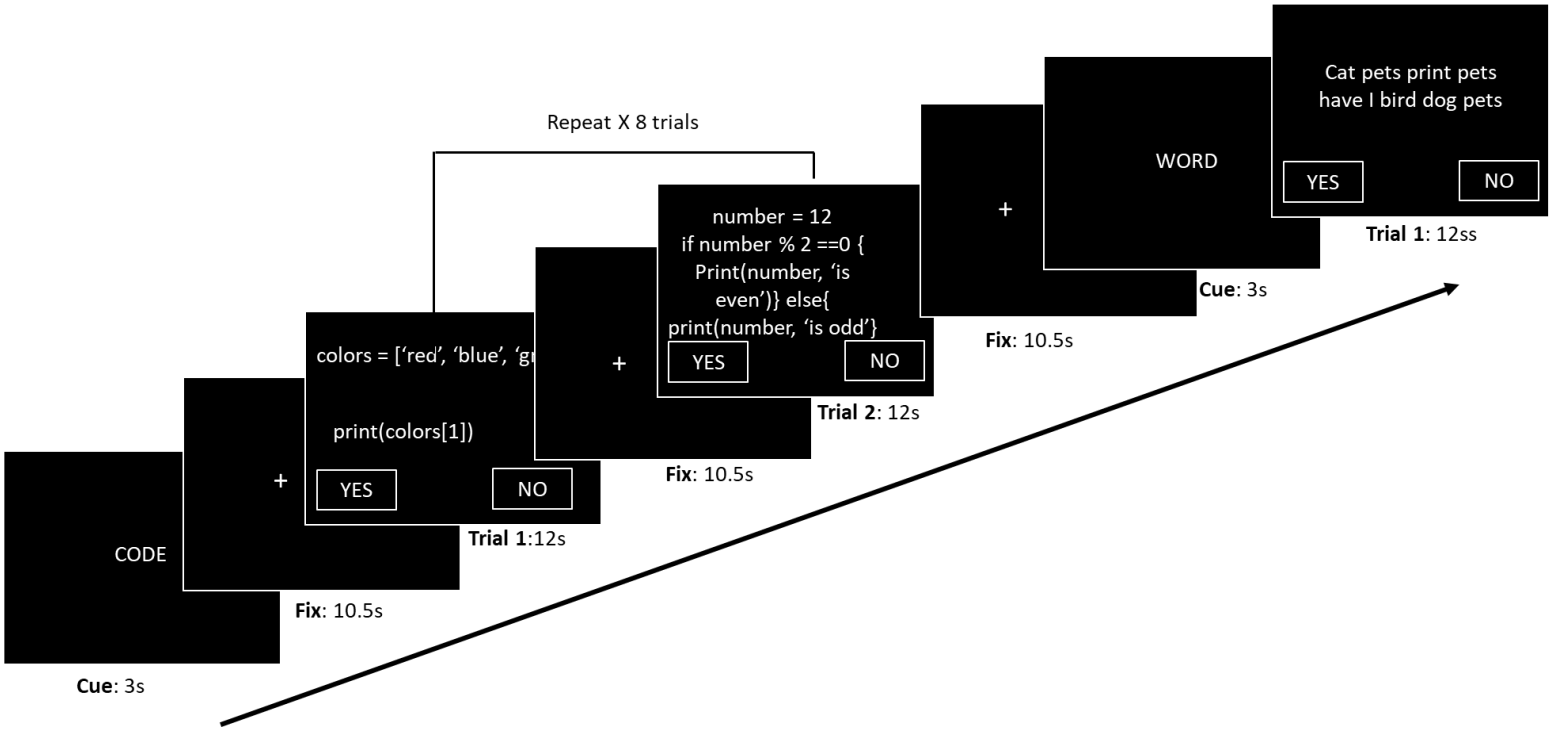
Schematic of Comprehension Task Stimuli



Note. Example stimuli from the Comprehension task. The words from the Python condition were scrambled and used as the stimuli for the Scrambled Word Reading condition of the opposing version. The button laterality was also counterbalanced across versions. Participants were tasked with determining if the code compiled successfully to an output (Python trials) or if the participant knew the meanings of all the words (Scrambled Word Reading trials). On 30% of the trials there was an error (i.e., second row) that required the participants to respond using the "No" button indicating that there was either a syntax error (Python trials) or a nonword (Scrambled Word Reading). On 30% of trials there was a comprehension check probe (i.e., third row) that immediately followed the participant's response where they were asked if a probe Python output matched what they predicted (Python trials) or if the probe word was thematically related to the words they saw previously (Scrambled Word Reading trials). On some trials, comprehension check probes co-occurred with errors, this is not depicted here.

Figure 4

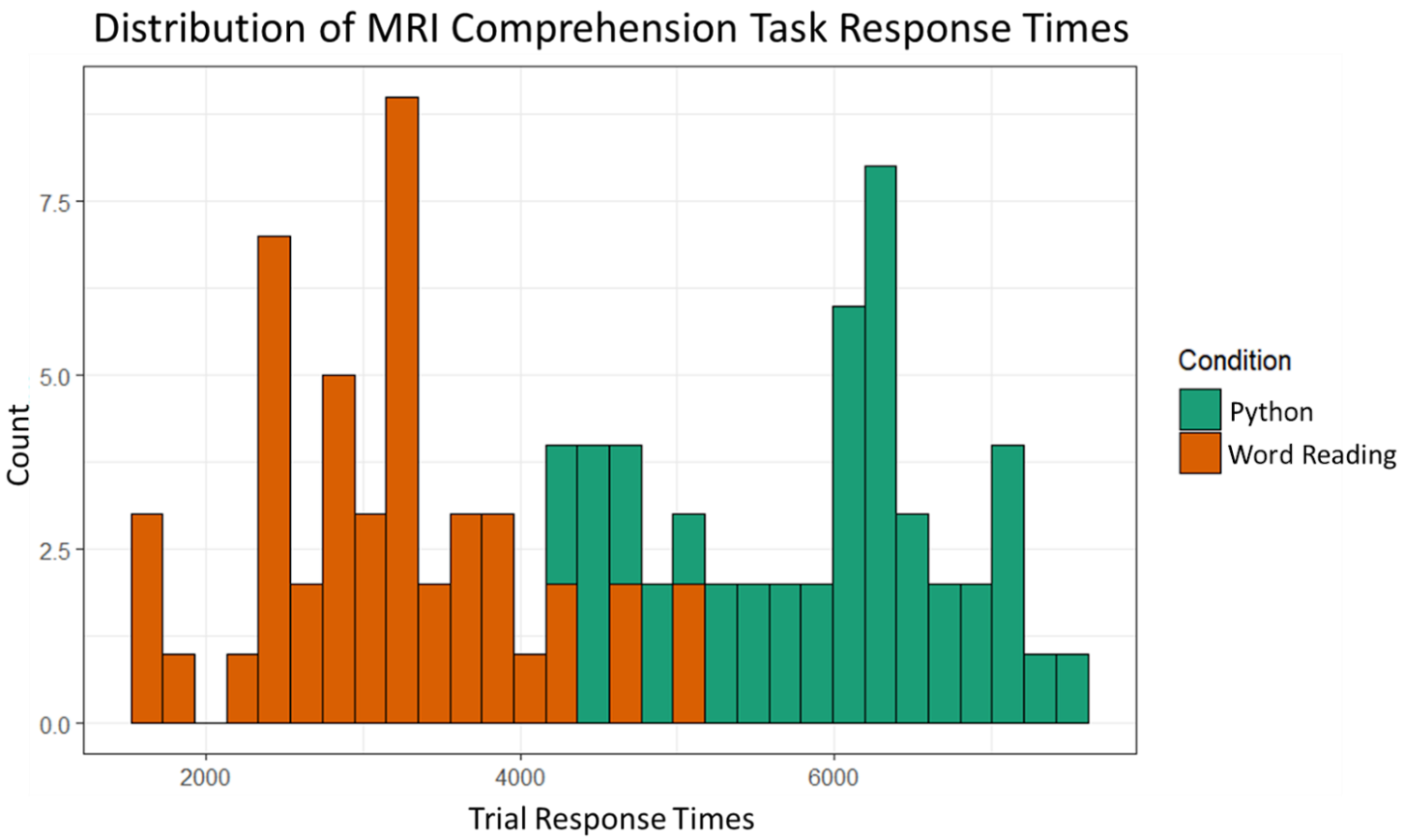
Schematic of Comprehension Task Timing



Note. Schematic depicting timing of the Comprehension task. Both conditions began with a cue alerting the participant what task condition was coming up next (i.e., CODE or WORD). Participants then completed eight task trials with a long fixation between each trial to ensure the hemodynamic response could be modeled at the individual trial level.

Figure 5

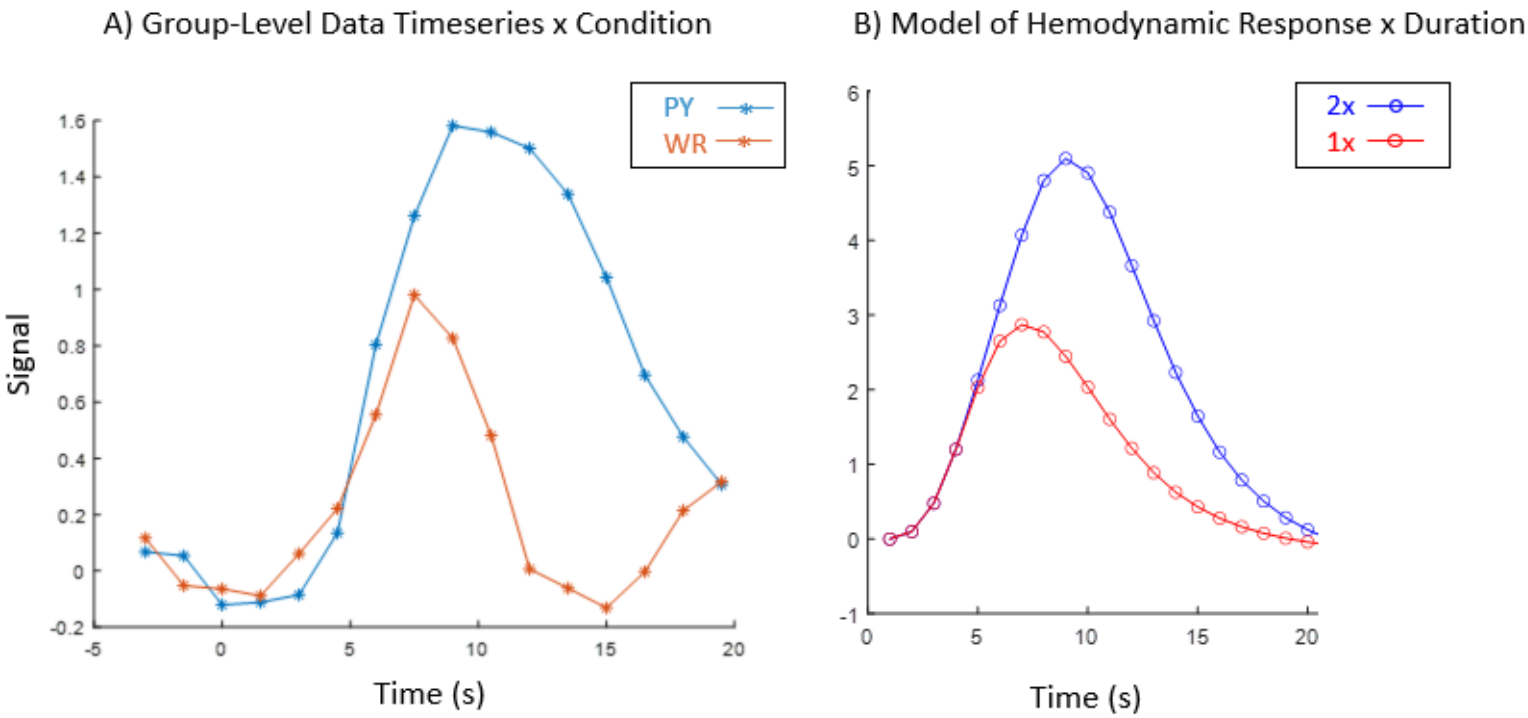
Response Time Differences in Comprehension Test



Note. Histogram of the difference in response times on the Comprehension task between the Python (green) and Scrambled Word Reading (orange) conditions. Response times were significantly longer for Python than for Scrambled Word Reading trials.

Figure 6

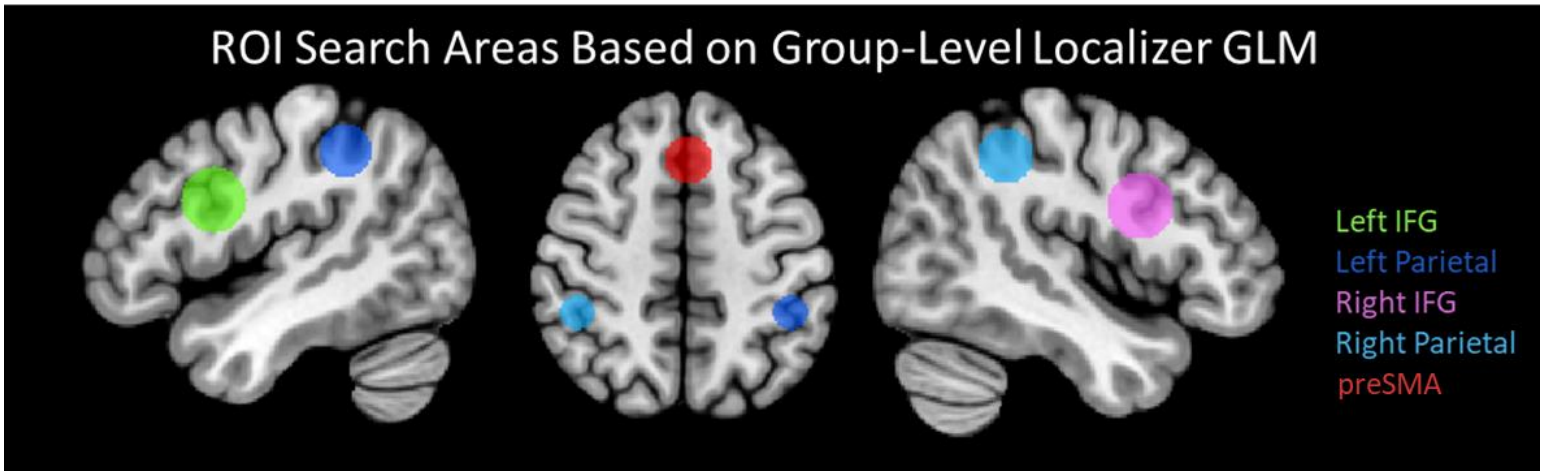
Role of Timing on Modeling the Hemodynamic Response



Note. A) Depicts the condition timecourses from the Comprehension task for Python (blue) and Scrambled Word Reading (orange), these timecourses were averaged across subjects and across the five ROIs used in the present study. B) Depicts a theoretical model of how increasing the duration of time the brain is working for leads to a corresponding amplitude increase in the hemodynamic response. The red line depicts the theoretical hemodynamic response for a task lasting x amount of time. When this amount of time is doubled to 2x, the hemodynamic response increases in amplitude (blue line). This theoretical model looks very similar to the observed timecourse data and motivated the inclusion of the GLM based on response times (RT model).

Figure 7

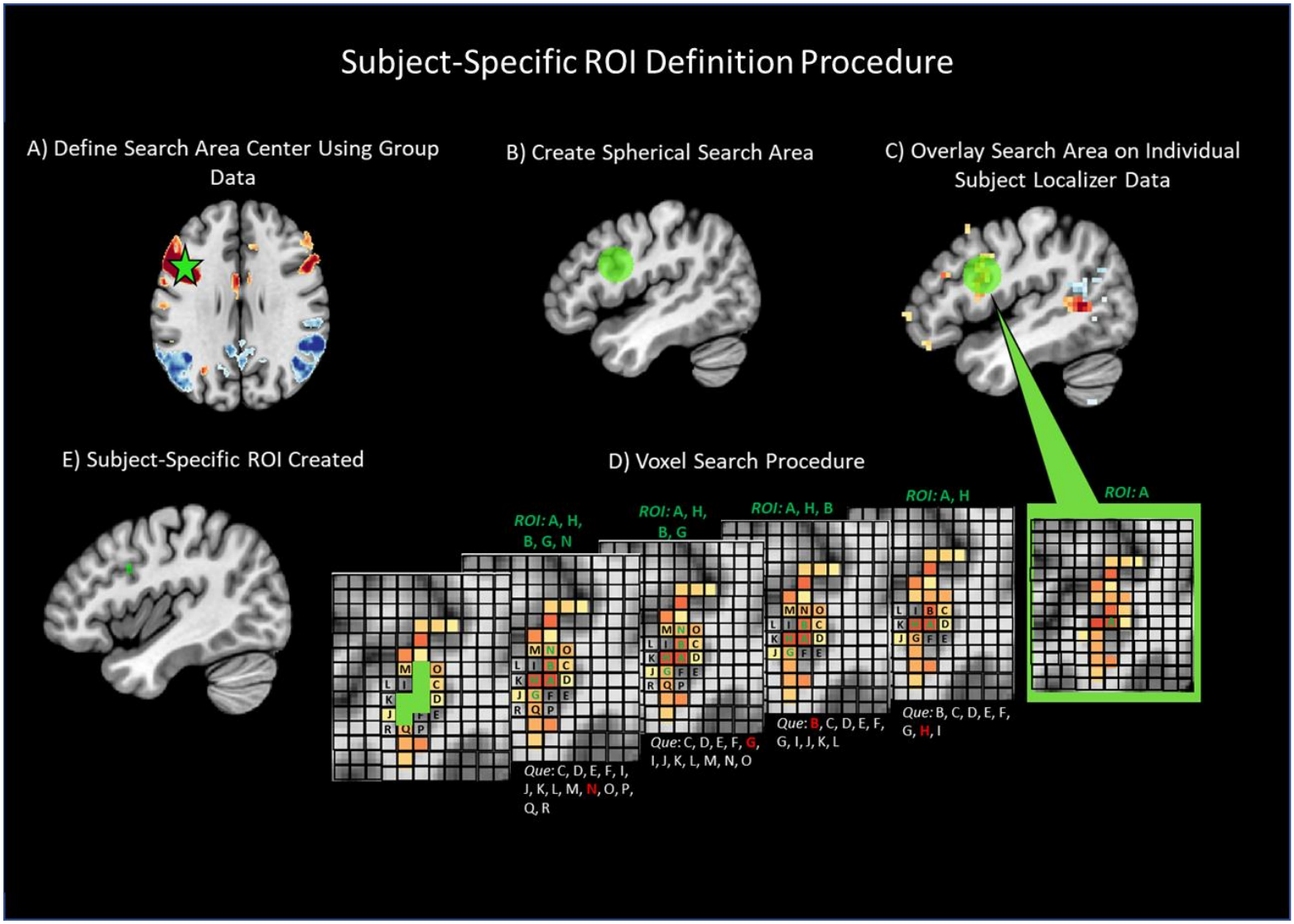
Phonological Localizer Search Areas Used for ROI Selection



Note. Spherical search areas used in the subject-specific ROI selection procedure. These areas were theoretically motivated and the specific center coordinate for each search areas was determined using the peak voxel in the group-level result from the Rhyming > Letter Search contrast for the Phonological Localizer task. Exact coordinates and sizes are included in Table 2.

Figure 8

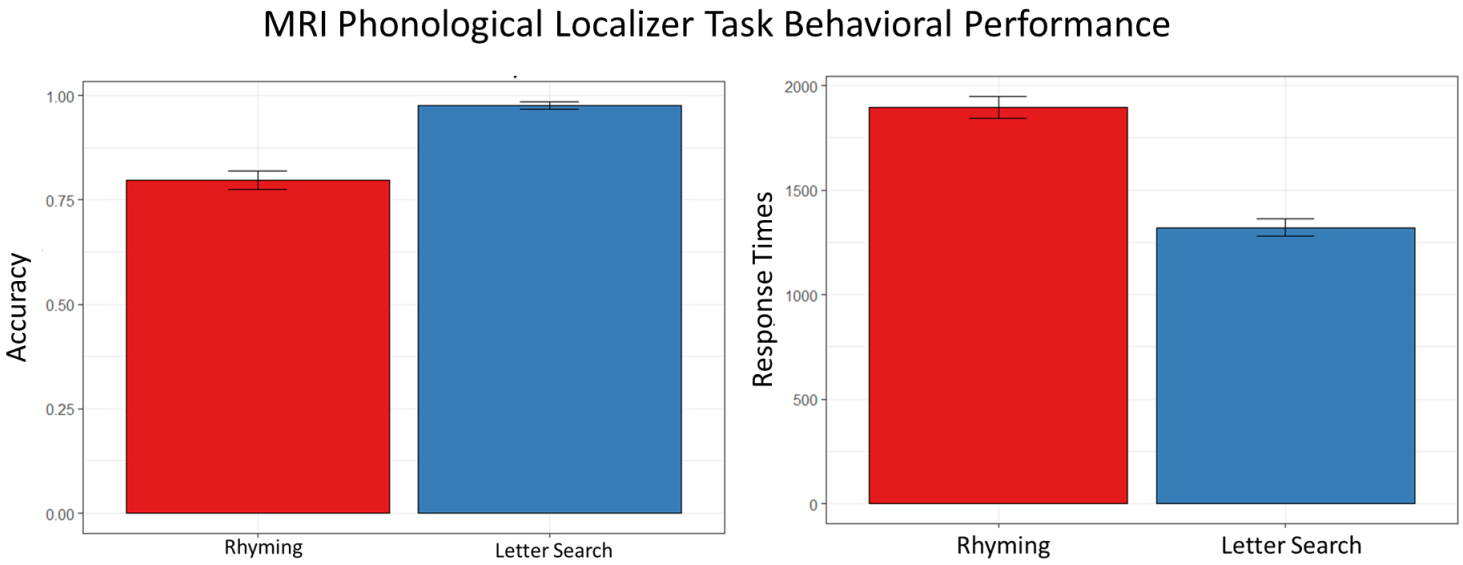
Subject Specific ROI Definition Procedure



Note. A) The theoretical search areas were based on prior literature and then the exact coordinates of the search areas were defined using the peak activation in each area at the group level. B) Spherical search areas were then created around these group-level activation peaks. C) For each participant the peak activation was found within each search area – this is depicted as Voxel A – and this voxel is added to the subject-specific ROI. D) From the peak voxel (Voxel A) all surrounding contiguous voxels are considered and the next most active voxel is added to the ROI (Voxel H), the remaining voxels are added to a que. As the search continues the voxels in the que are considered as well as new voxels contiguous to the voxels selected for inclusion in the ROI. This procedure continues until a size limit (5-voxels) has been reached. E) The resulting subject-specific ROI is used to isolate the phonological system in the Comprehension Task.

Figure 9

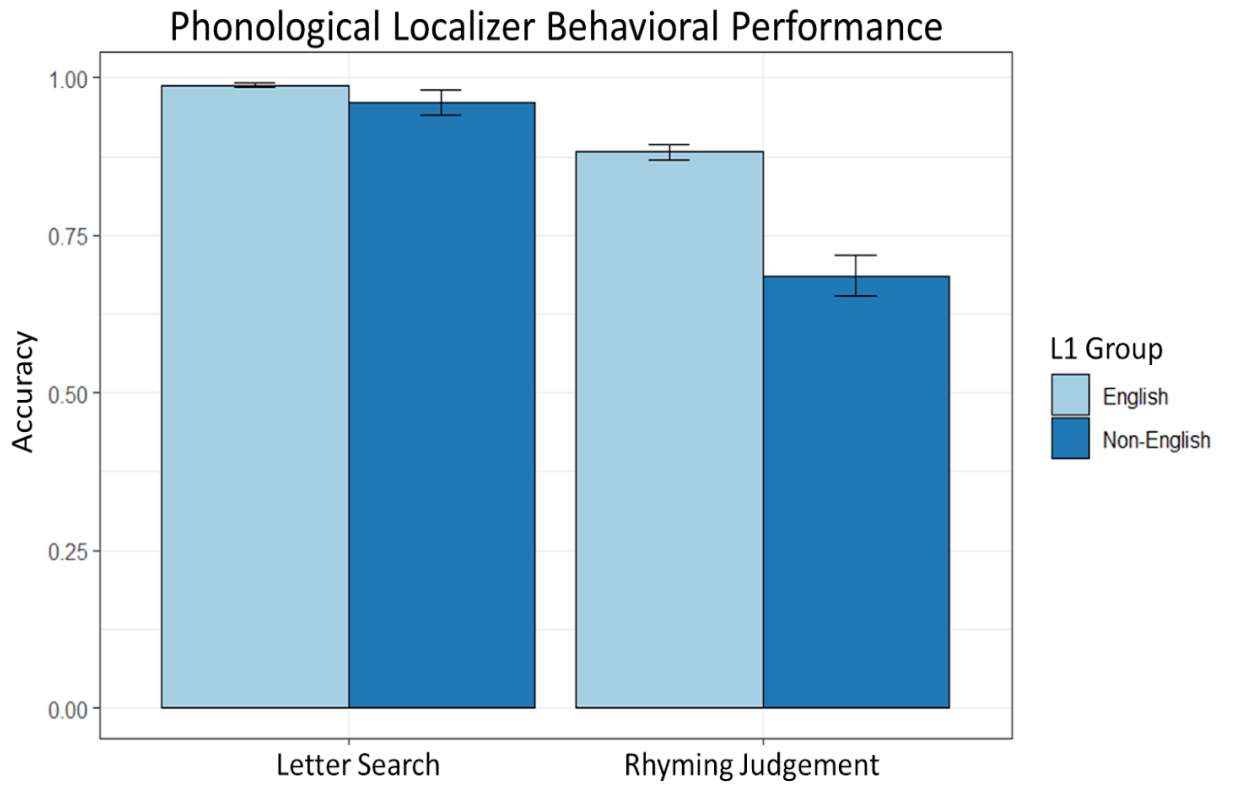
Behavioral Performance on the Phonological Localizer



Note. Behavioral accuracy and response time data from the Phonological Localizer task. Participants had higher accuracy and faster response times on the Letter Search (blue) trials than on the Rhyming (red) trials. Error bars denote standard error.

Figure 10

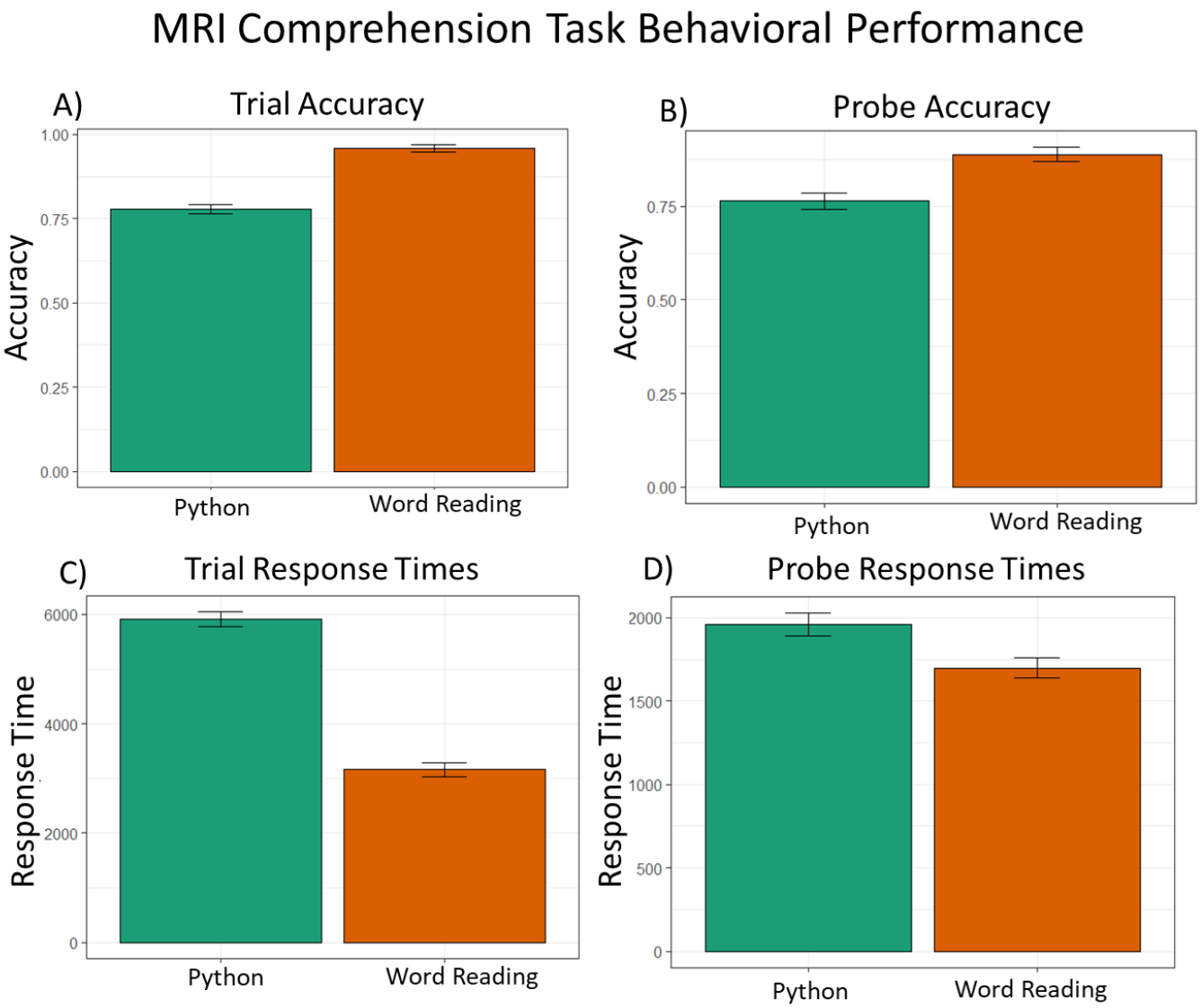
Phonological Localizer Differences Based on Language Background



Note. Behavioral accuracy data from the Phonological Localizer task. Both language groups had higher accuracy on the Letter Search trials than on the Rhyming trials. However, the magnitude of this difference varied between participants with English as the first language (light blue bars) and individuals with a non-English first language (dark blue bars), such that participants with non-English L1s had a greater difference in accuracy between conditions. Error bars denote standard error.

Figure 11

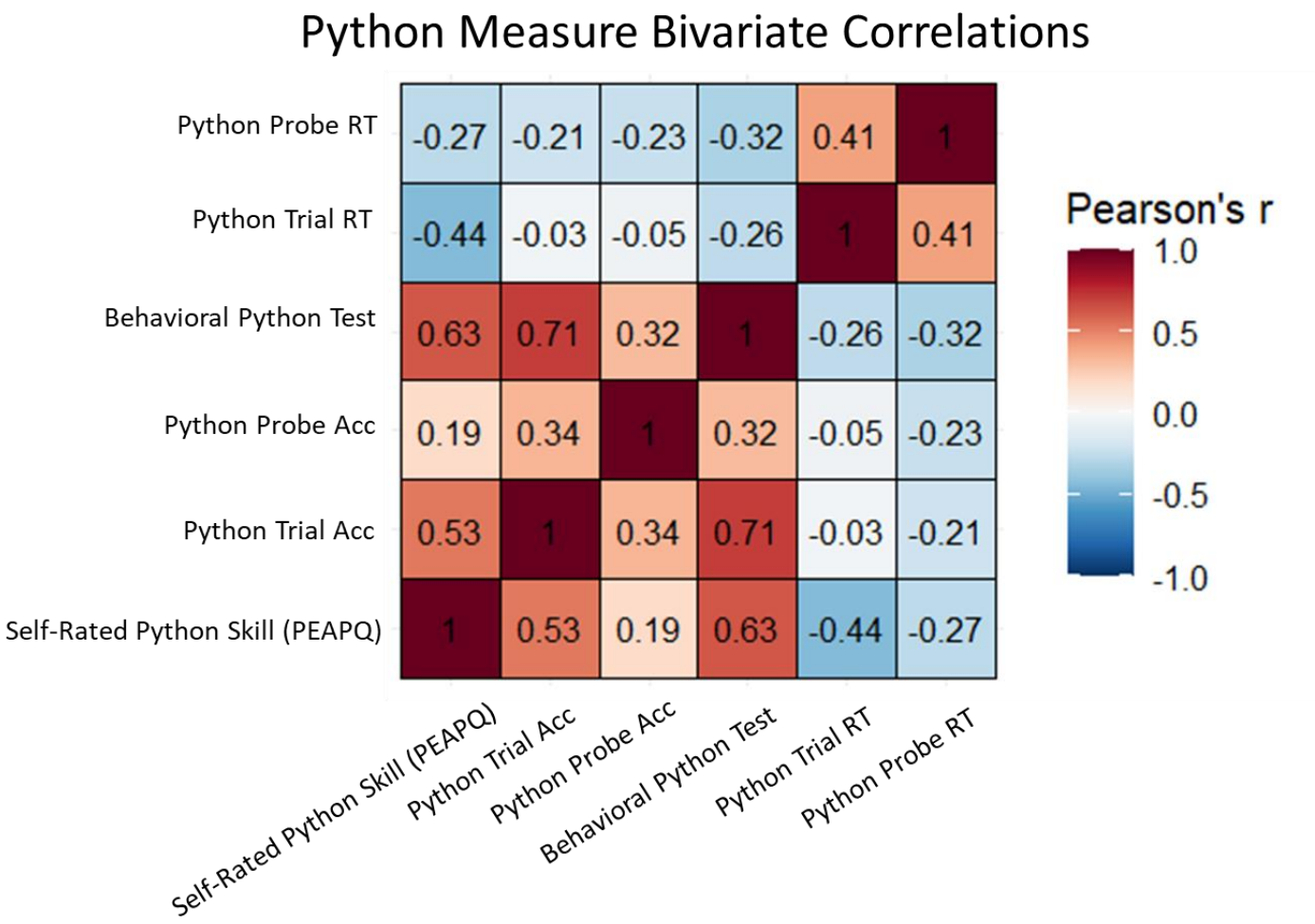
Behavioral Performance on the Comprehension Task



Note. Behavioral performance data from the Comprehension task. For all plots performance on the Python condition is shown in green and performance on the Scrambled Word Reading condition is shown in orange: A) depicts total trial accuracy, B) depicts comprehension check probe accuracies, C) depicts response times on the initial decision as to whether the code compiled to an output (Python) or whether the participant knew the meanings of all words in the word list (Scrambled Word Reading), and D) depicts response times on the comprehension check probes. In all cases, the Python condition was more difficult than the Scrambled Word Reading condition leading to lower accuracies and longer response times for Python than for Scrambled Word Reading. Error bars denote standard error.

Figure 12

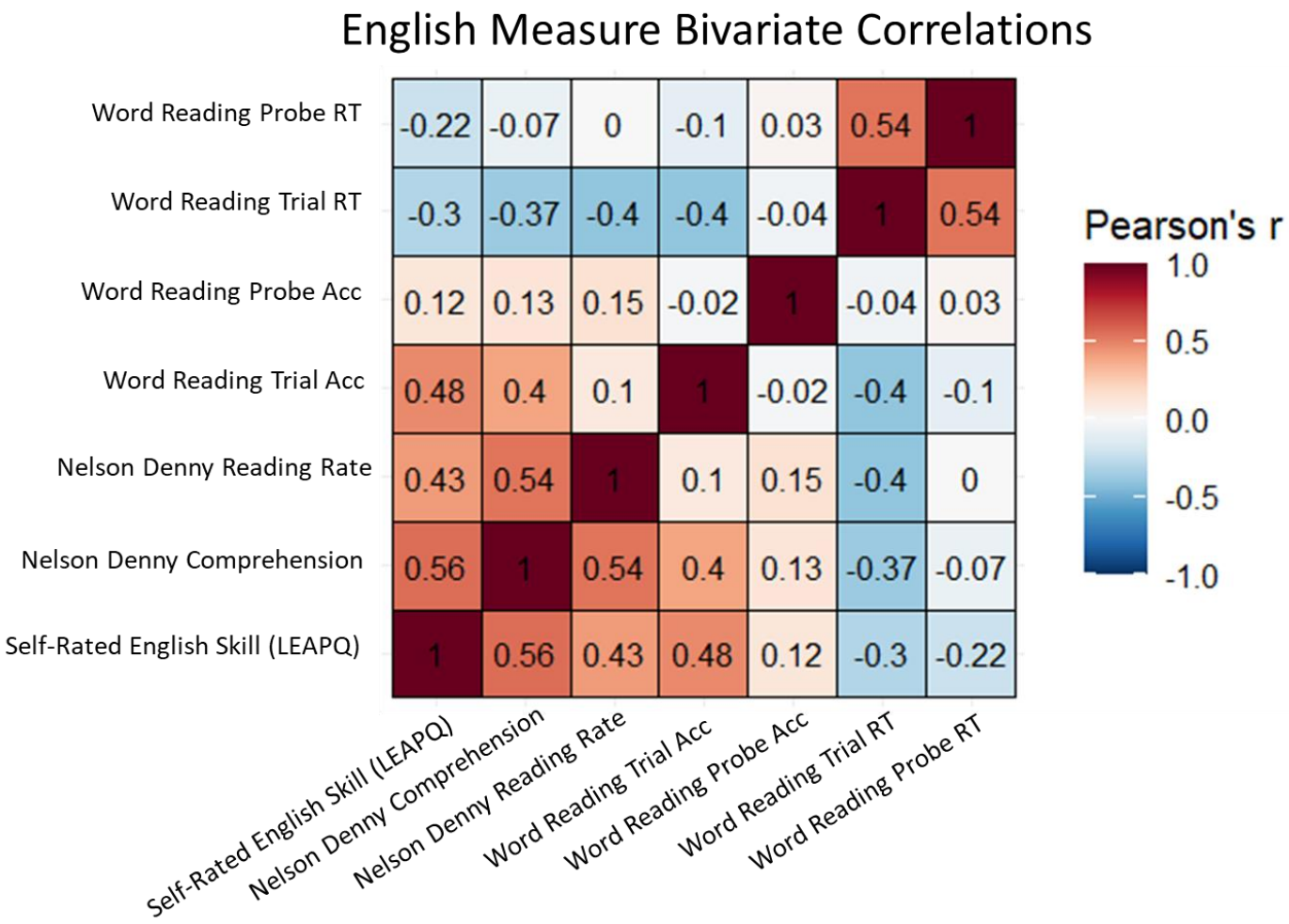
Bivariate Correlation Matrix of Python Measures



Note. Matrix of bivariate correlations between behavioral measures indexing Python skill. Heatmap denotes the correlation coefficient strength as measured by Pearson's *r*. By and large, most programming measures were correlated with one another demonstrating convergent validity of the Python related behavioral measures. A full correlation matrix including all behavioral measures of interest is provided in Table 4.

Figure 13

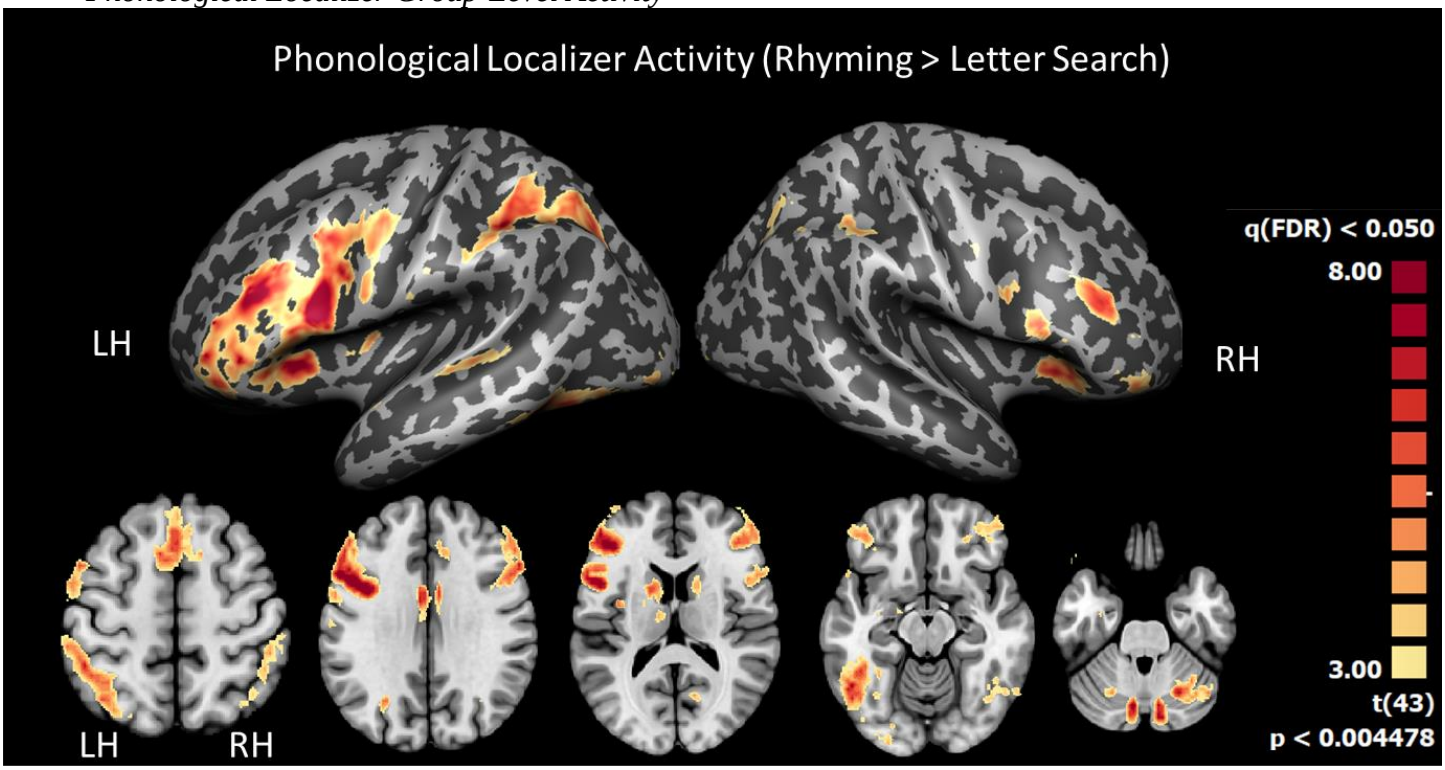
Bivariate Correlation Matrix of English Measures



Note. Matrix of bivariate correlations between behavioral measures indexing English skill. Heatmap denotes the correlation coefficient strength as measured by Pearson's *r*. By and large, most measures of English skill were correlated with one another demonstrating convergent validity of the English related behavioral measures. A full correlation matrix including all behavioral measures of interest is provided in Table 4.

Figure 14

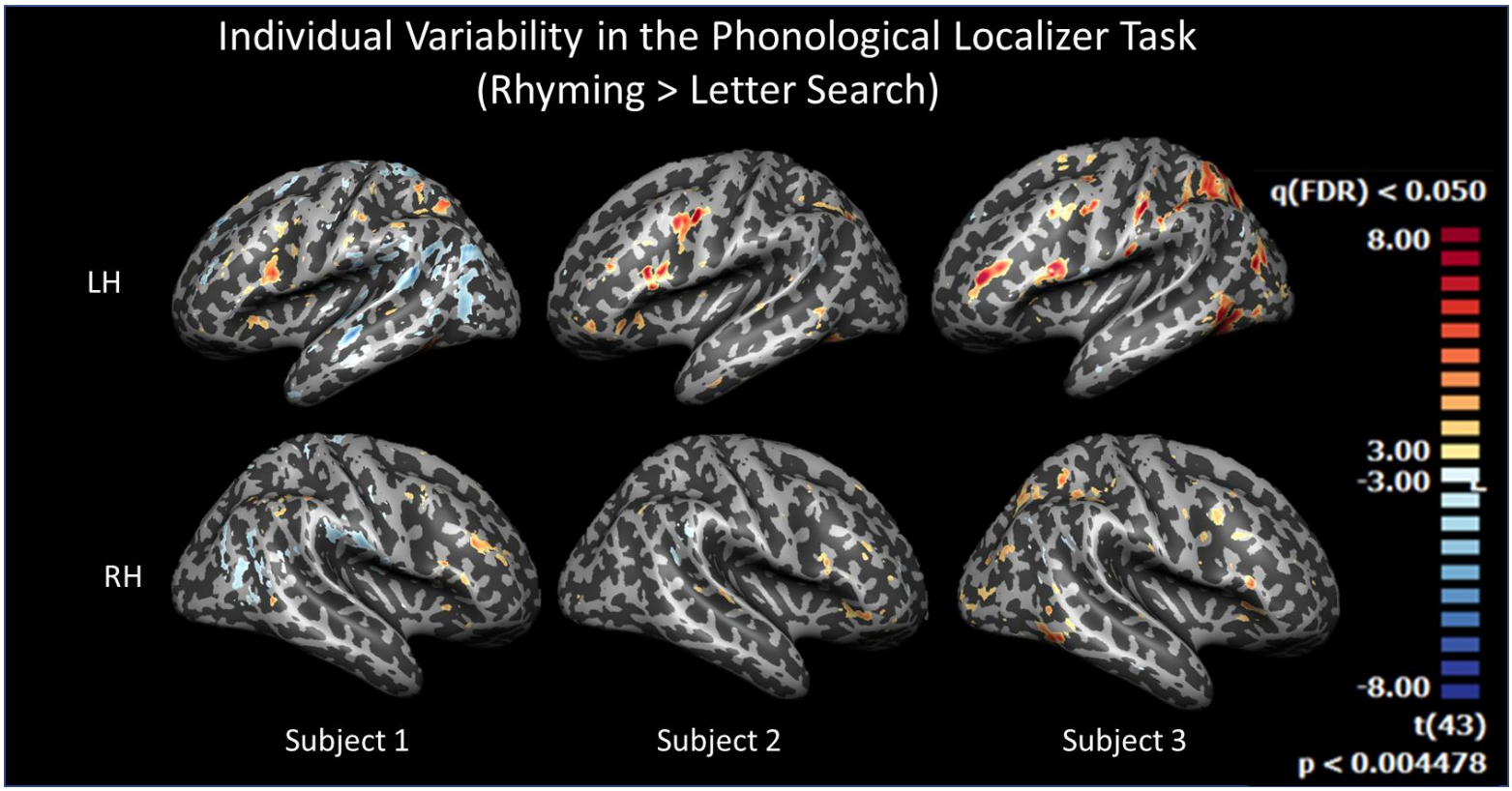
Phonological Localizer Group-Level Activity



Note. Group-level whole-brain General Linear Model (GLM) results for the Phonological Localizer task. Activity depicted is the difference between the Rhyming > Letter Search contrast. Warm colors represent regions where Rhyming > Letter Search and cool colors represent regions where Letter Search > Rhyming. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 5.

Figure 15

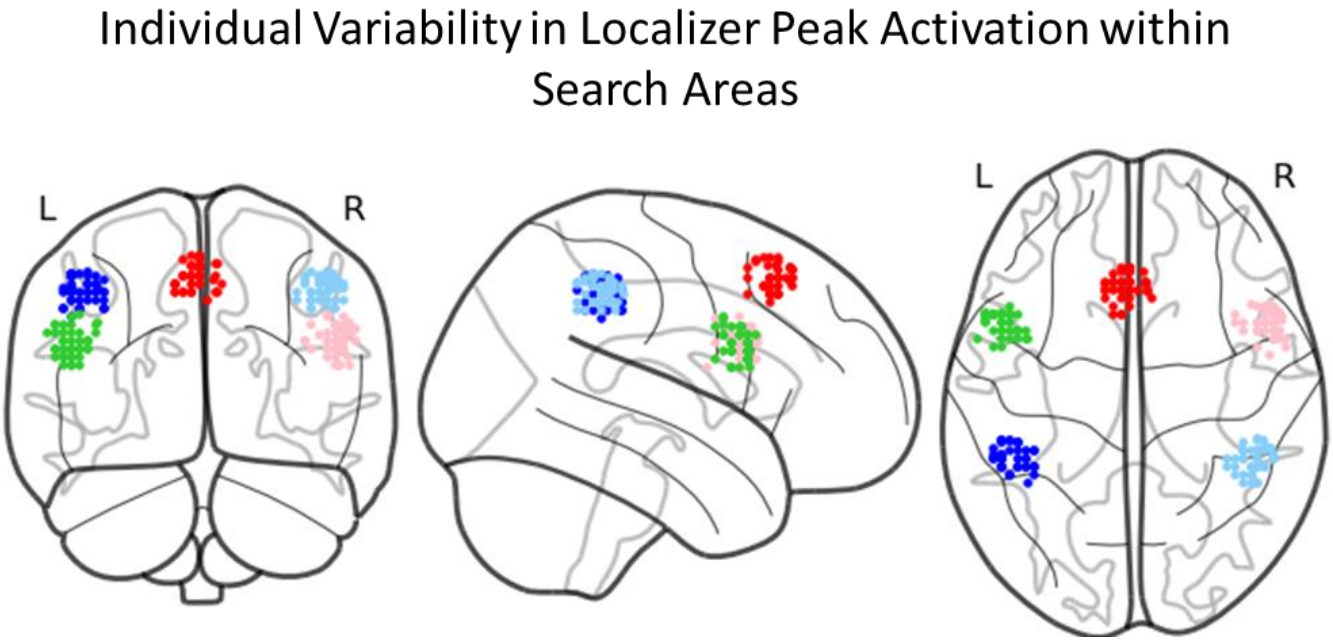
Individual Variability in Phonological Localizer Activity



Note. First-level individual subject data from the Phonological Localizer task is displayed for three participants. All data depicted the difference between the Rhyming > Letter Search contrast. Warm colors represent regions where Rhyming > Letter Search and cool colors represent regions where Letter Search > Rhyming. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. Most subjects, including the three depicted here, showed activation in the left inferior frontal gyrus (top row). However, recruitment of areas in the right hemisphere (bottom row) were much more variable across subjects.

Figure 16

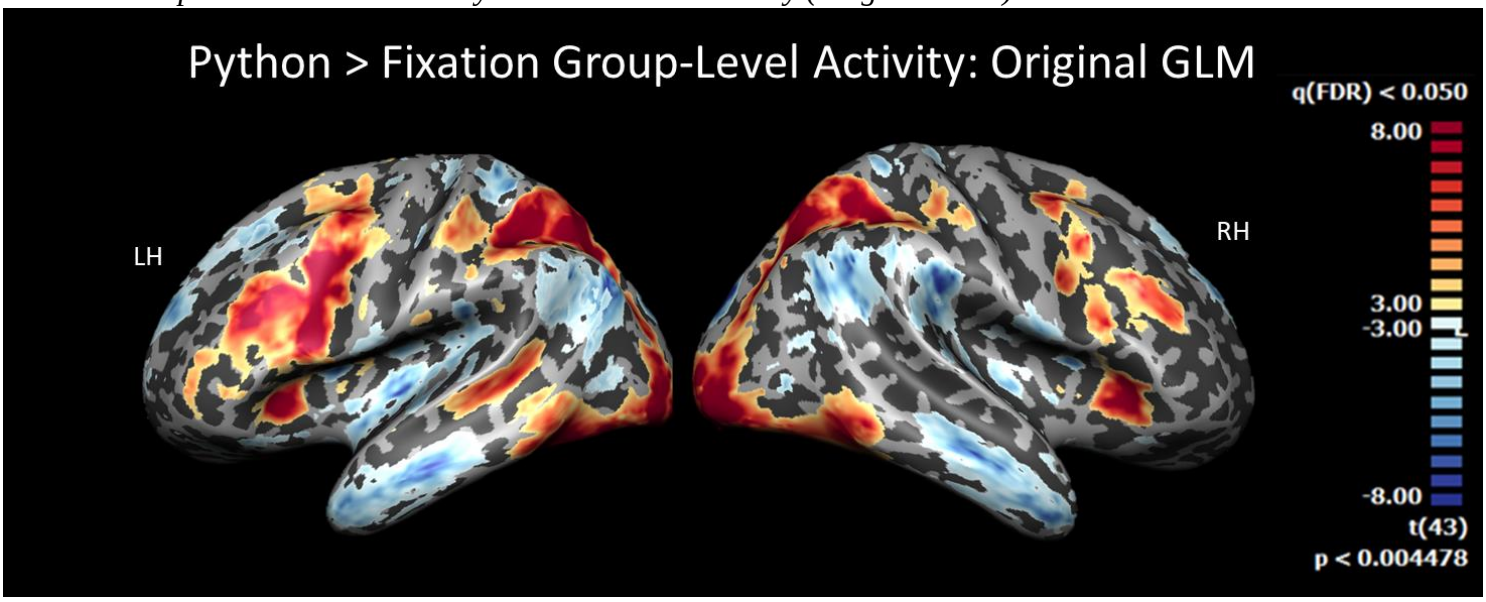
Individual Variability in Phonological Localizer Peak Activation within Search Areas



Note. Glass brain plot depicting each participants' peak activation within the search area for the Rhyming > Letter Search contrast of the Phonological Localizer task. Within each search area, each dot represents an individual subject's peak activation. The search areas are color coded as follows: left inferior frontal gyrus (green), left parietal (dark blue), right inferior frontal gyrus (pink), right parietal (light blue), and preSMA (red). The spread of peak activation within each search area highlights that there was considerable variability across subjects in terms of which voxels were mostly implicated in phonological processing.

Figure 17

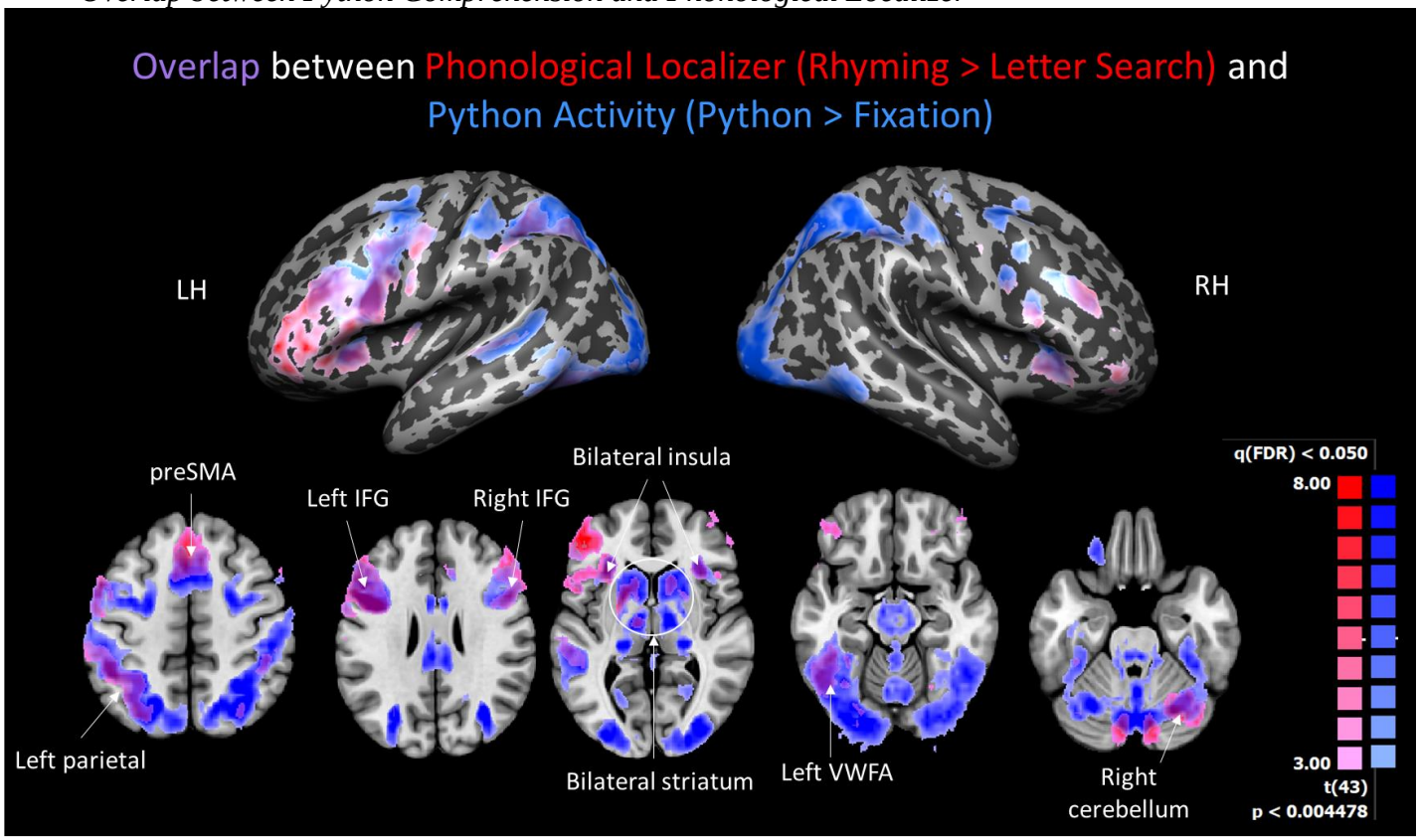
Group-Level Whole-Brain Python > Fixation Activity (Original GLM)



Note. Group-level whole-brain results using the Original GLM for the Python > Fixation contrast. Warm colors represent regions where Python > Fixation and cool colors represent regions where Fixation > Python. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 6.

Figure 18

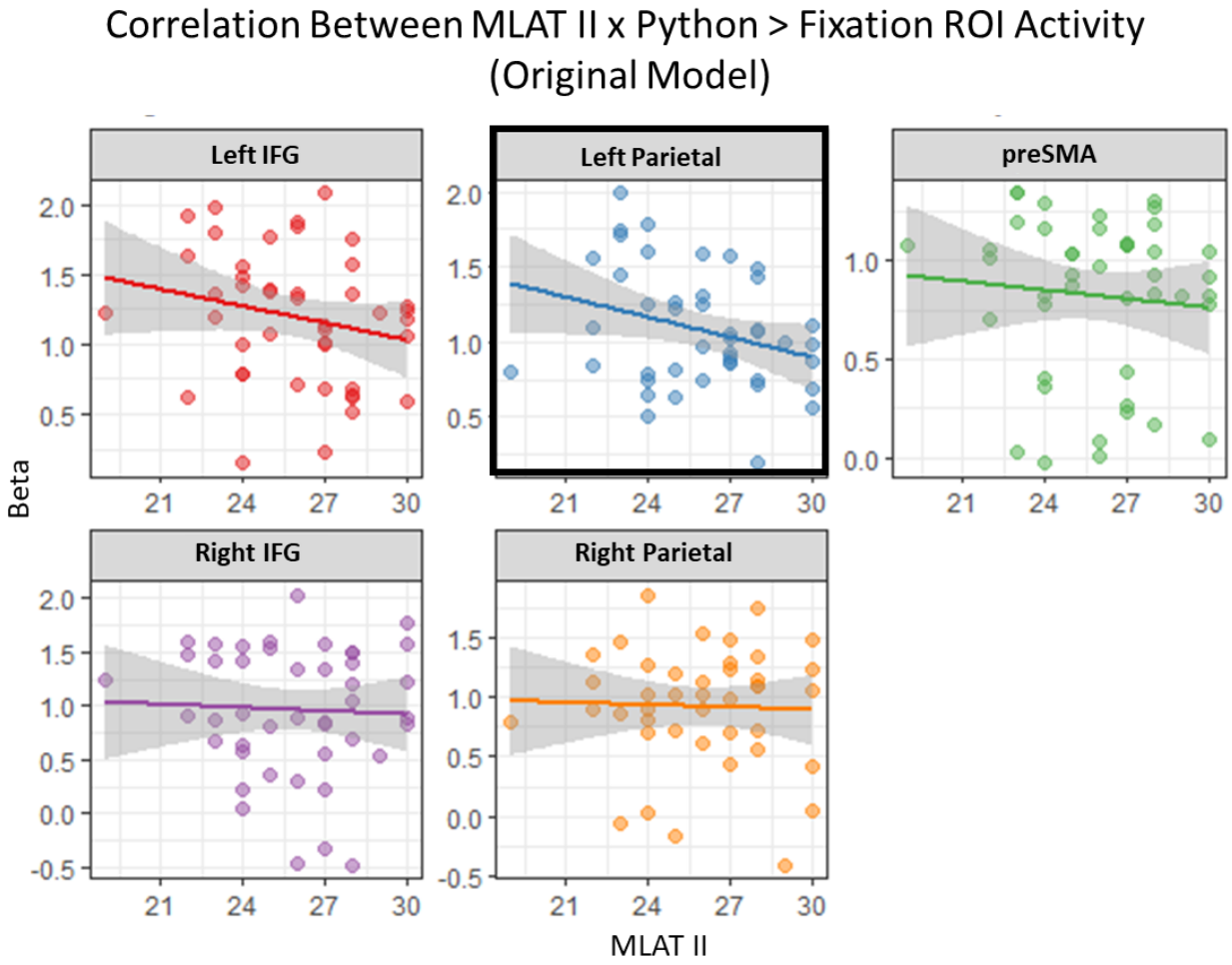
Overlap between Python Comprehension and Phonological Localizer



Note. Overlap between whole-brain group-level activity for the Phonological Localizer task (red) and the Python > Fixation contrast for the Comprehension task (blue) under the Original GLM. Both statistical maps are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. Overlap between the maps is depicted in purple.

Figure 19

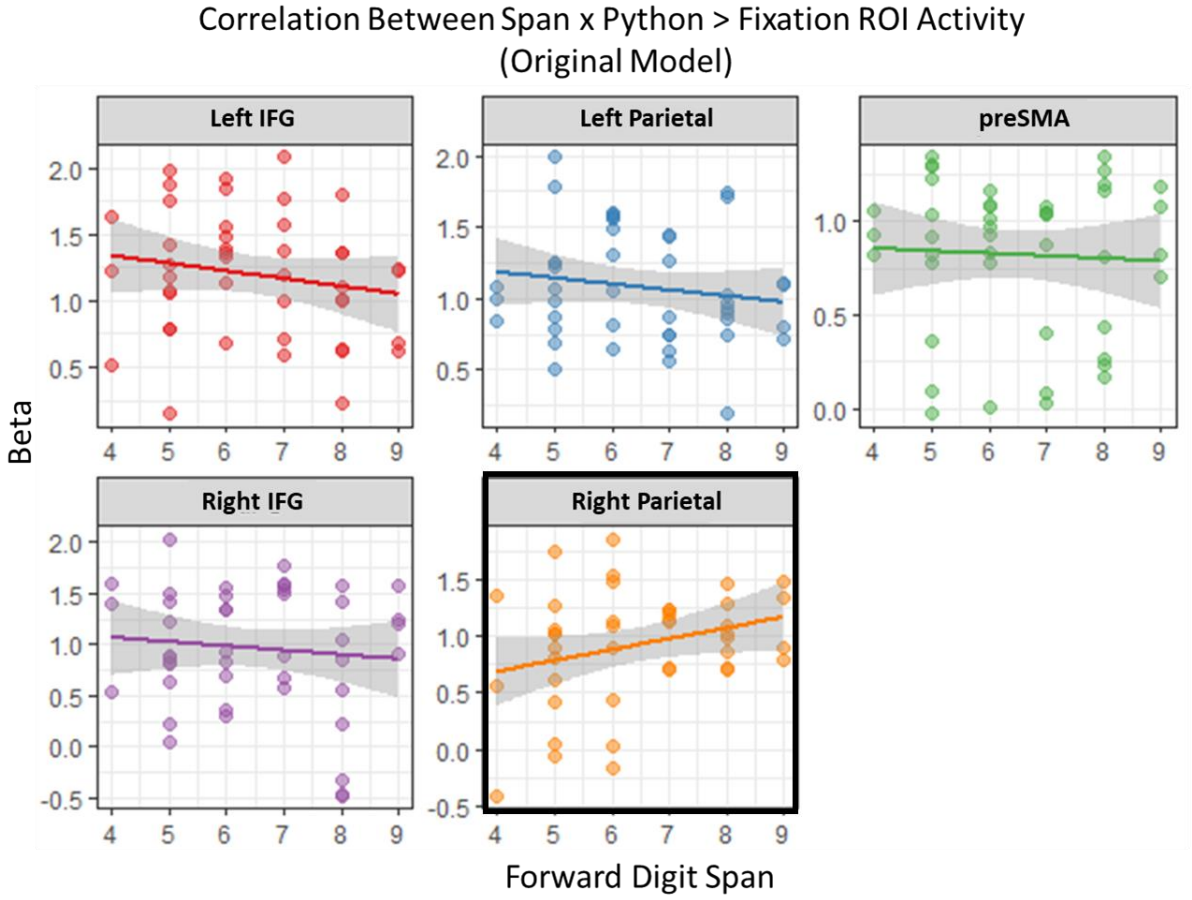
Scatterplots Depicting Correlation Between MLAT II x Python > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and MLAT II under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7A.

Figure 20

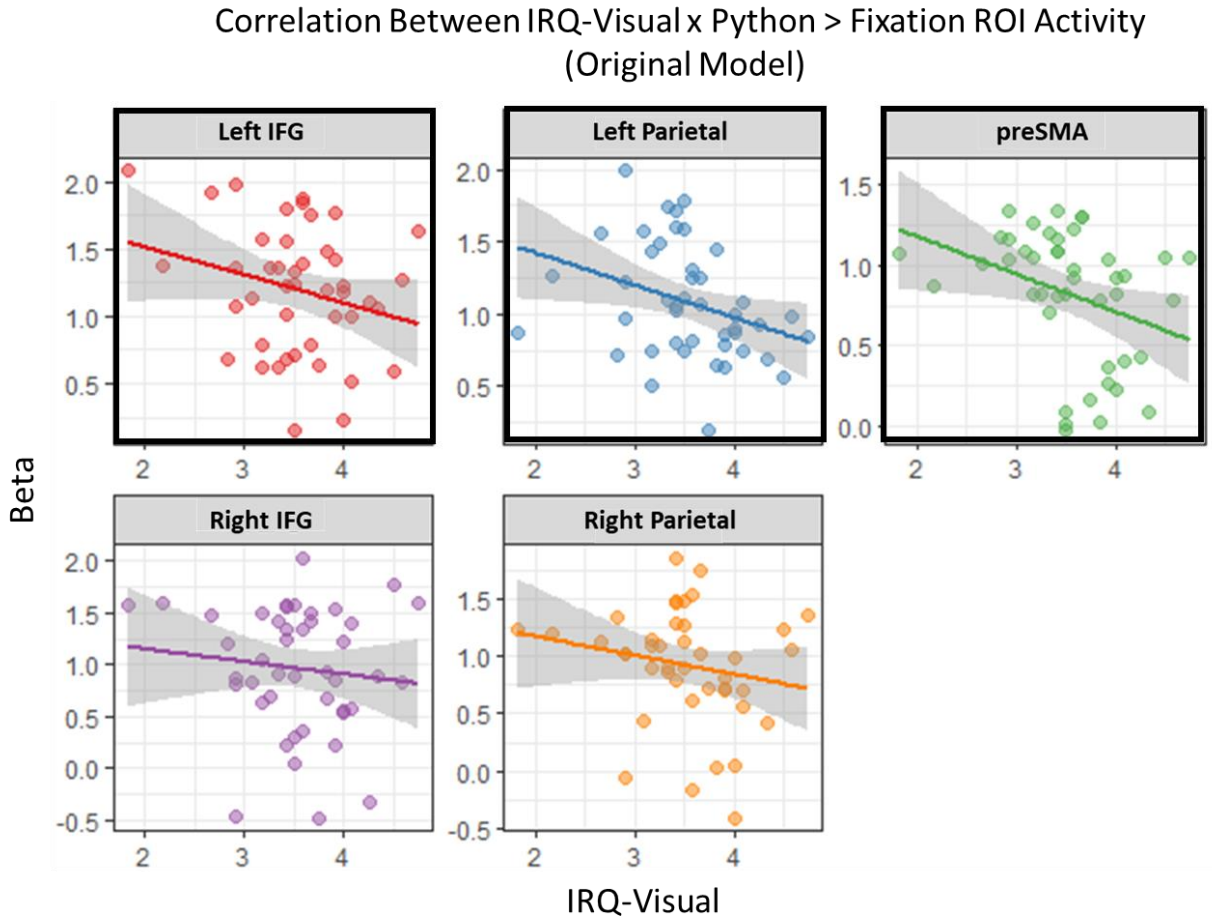
Scatterplots Depicting Correlation Between Span x Python > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and Forward Digit Span under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7A.

Figure 21

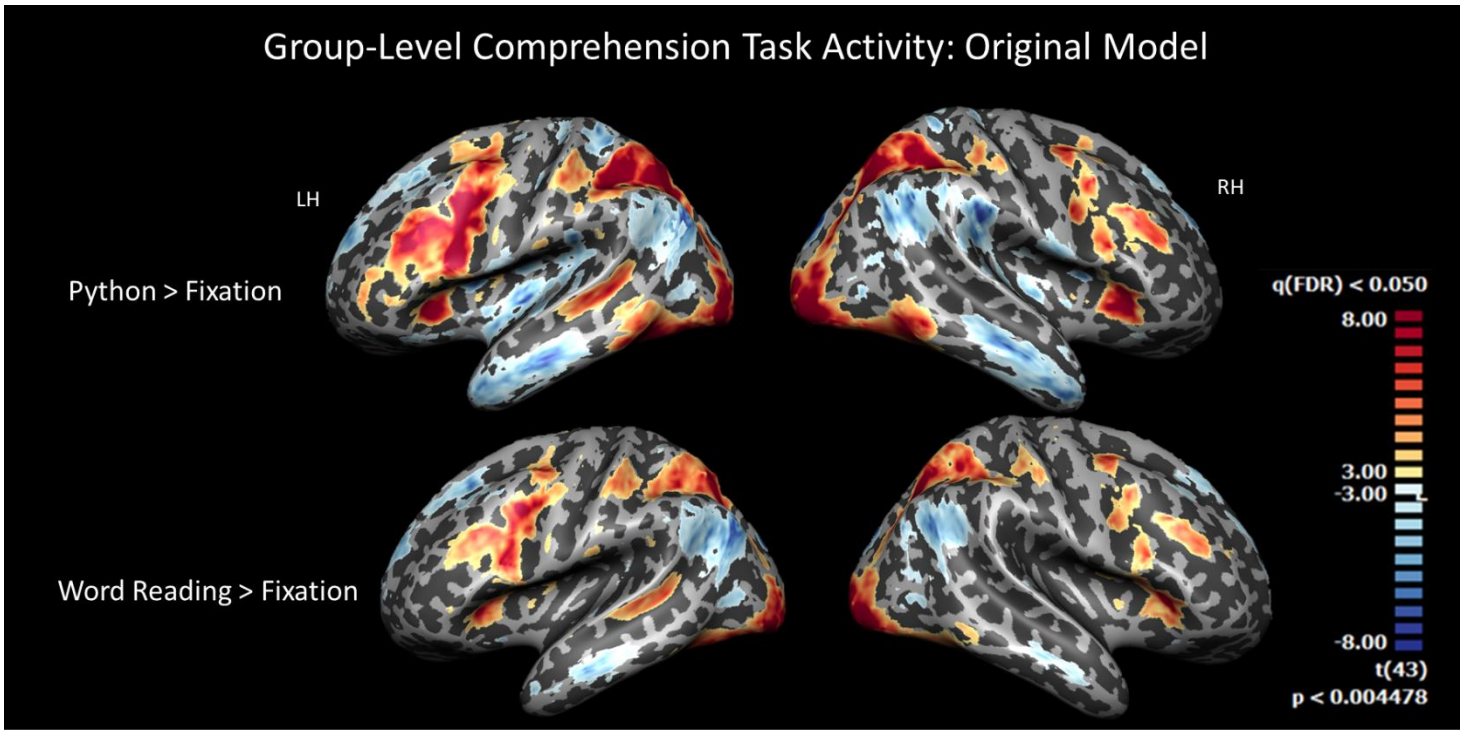
Scatterplots Depicting Correlation Between IRQ-Visual x Python > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and IRQ-Visual under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7A.

Figure 22

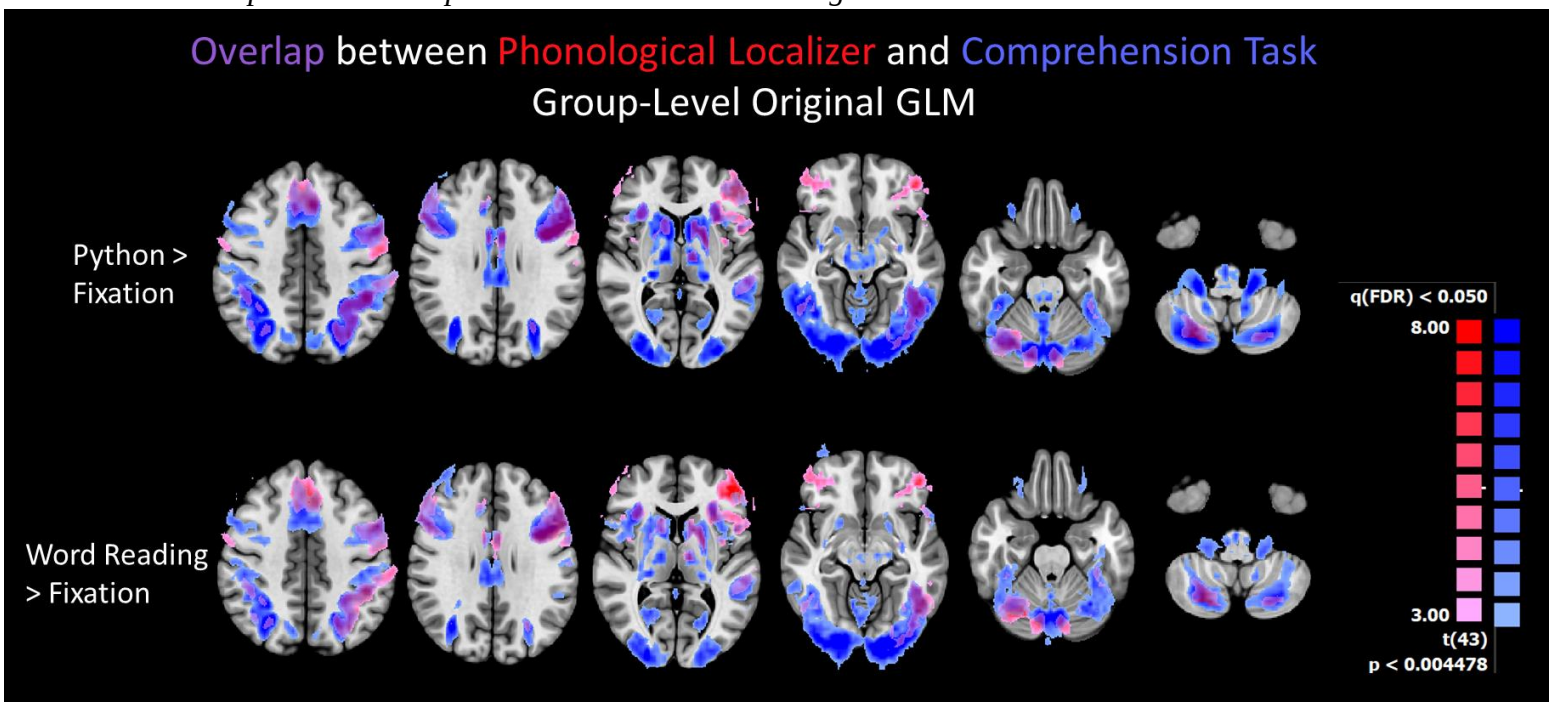
Group-Level Whole-Brain Activity (Original GLM)



Note. Group-level whole-brain results using the Original GLM for the Python > Fixation (top row) and Scrambled Word Reading (bottom row) contrasts. Warm colors represent regions where Task Condition > Fixation and cool colors represent regions where Fixation > Task Condition. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 6 for the Python > Fixation contrast and in Table 9 for the Scrambled Word Reading > Fixation contrast.

Figure 23

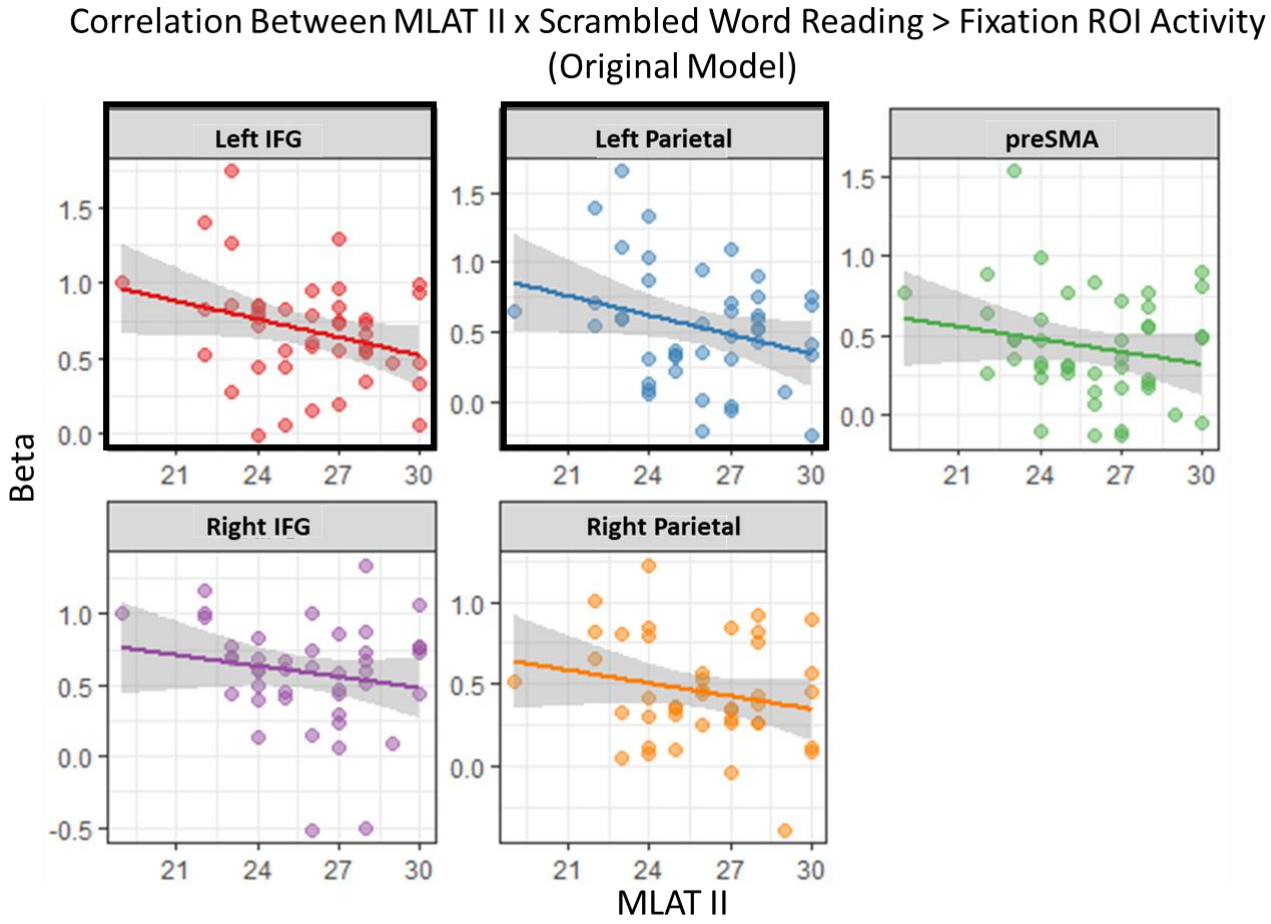
Overlap between Comprehension Task and Phonological Localizer Task



Note. Overlap between whole-brain group-level activity for the Phonological Localizer task (red) and the Comprehension task (blue) under the Original GLM. Phonological Localizer overlap with the Python > Fixation contrast is depicted in the top row and overlap the with Scrambled Word Reading contrast is depicted in the bottom row. Both statistical maps are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. Overlap between the maps is depicted in purple.

Figure 24

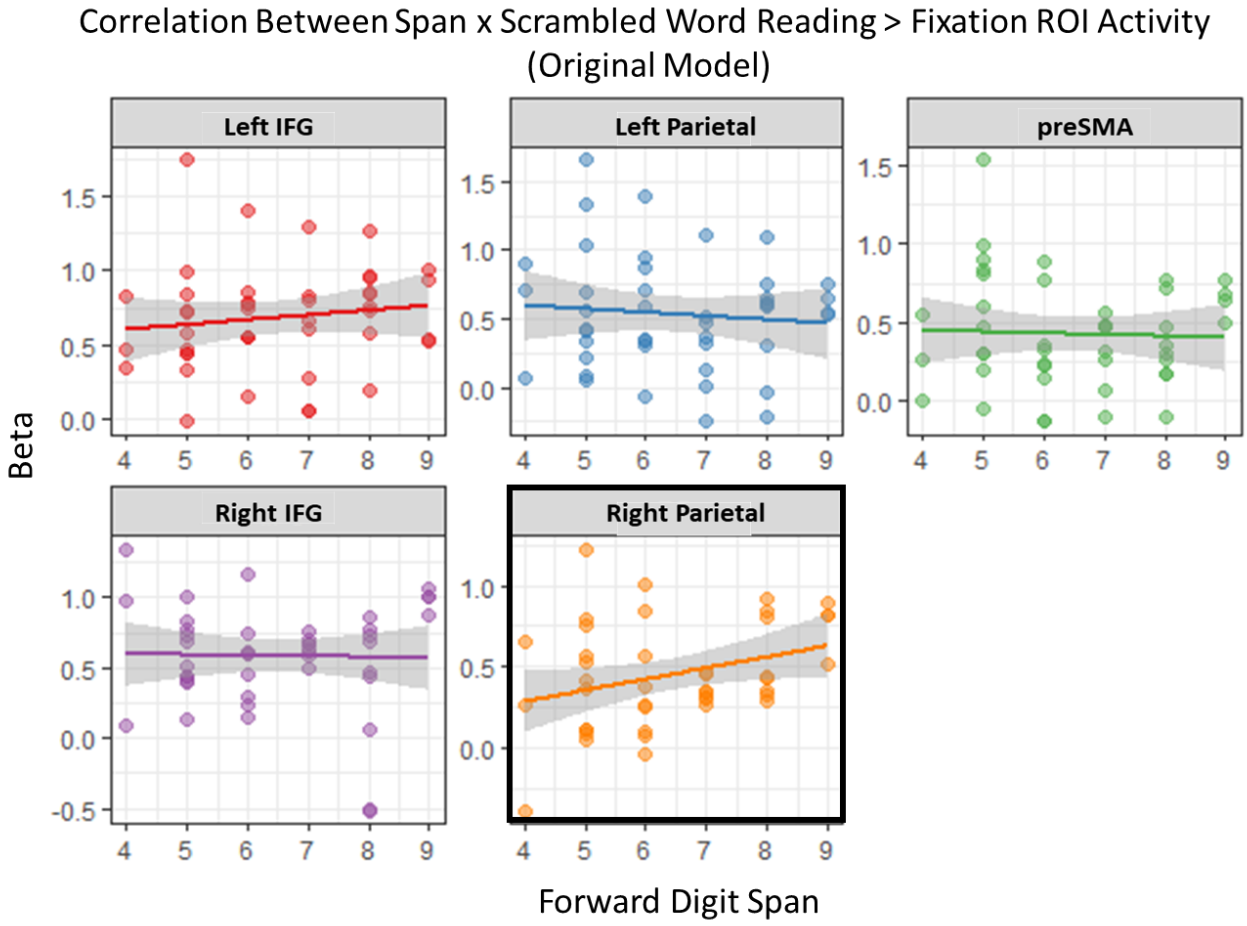
Scatterplots Depicting Correlation Between MLAT II x Word Reading > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Scrambled Word Reading > Fixation in the phonological ROIs and MLAT II under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7B.

Figure 25

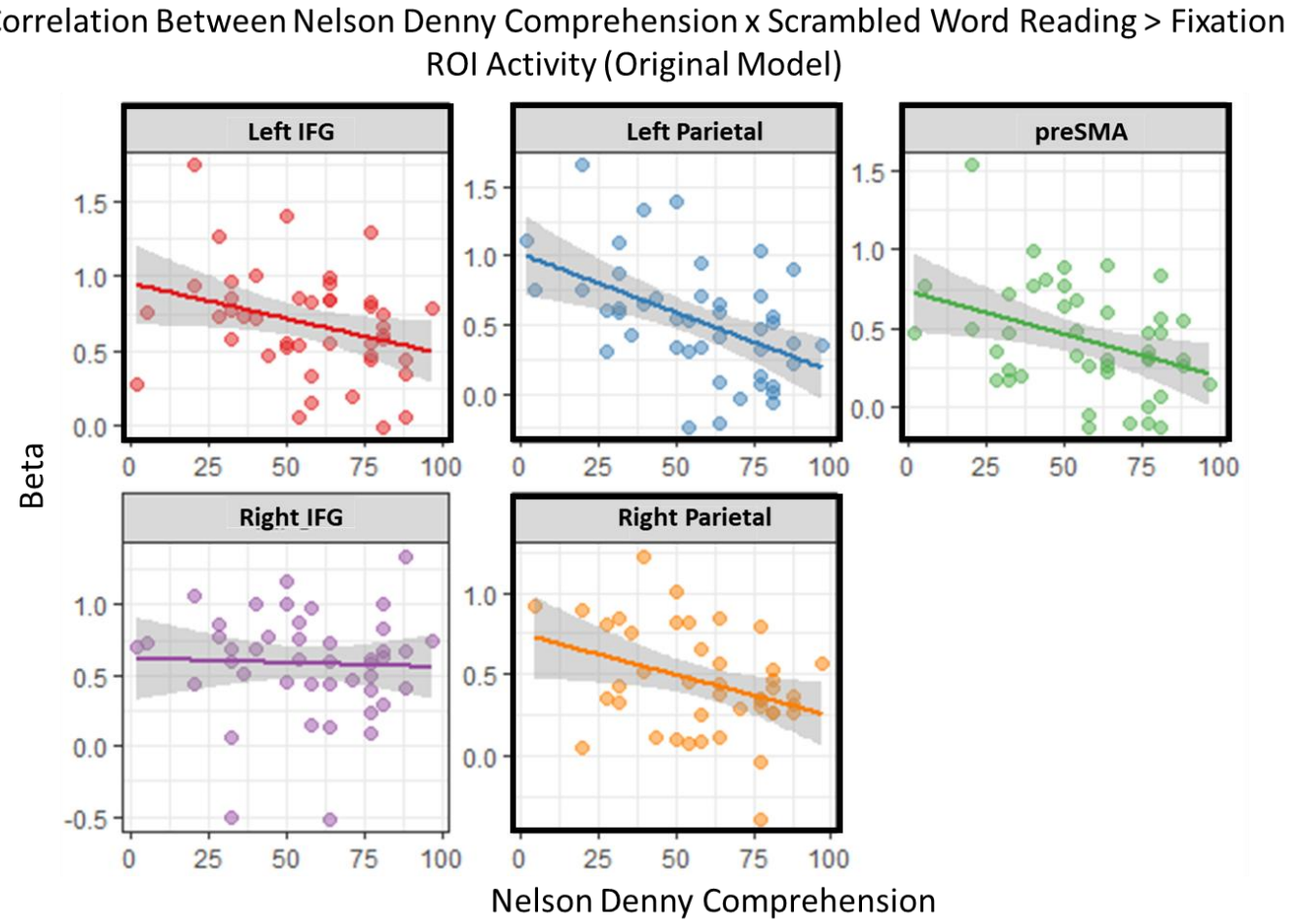
Scatterplots Depicting Correlation Between Span x Word Reading > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Scrambled Word Reading > Fixation in the phonological ROIs and Forward Digit Span under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7B.

Figure 26

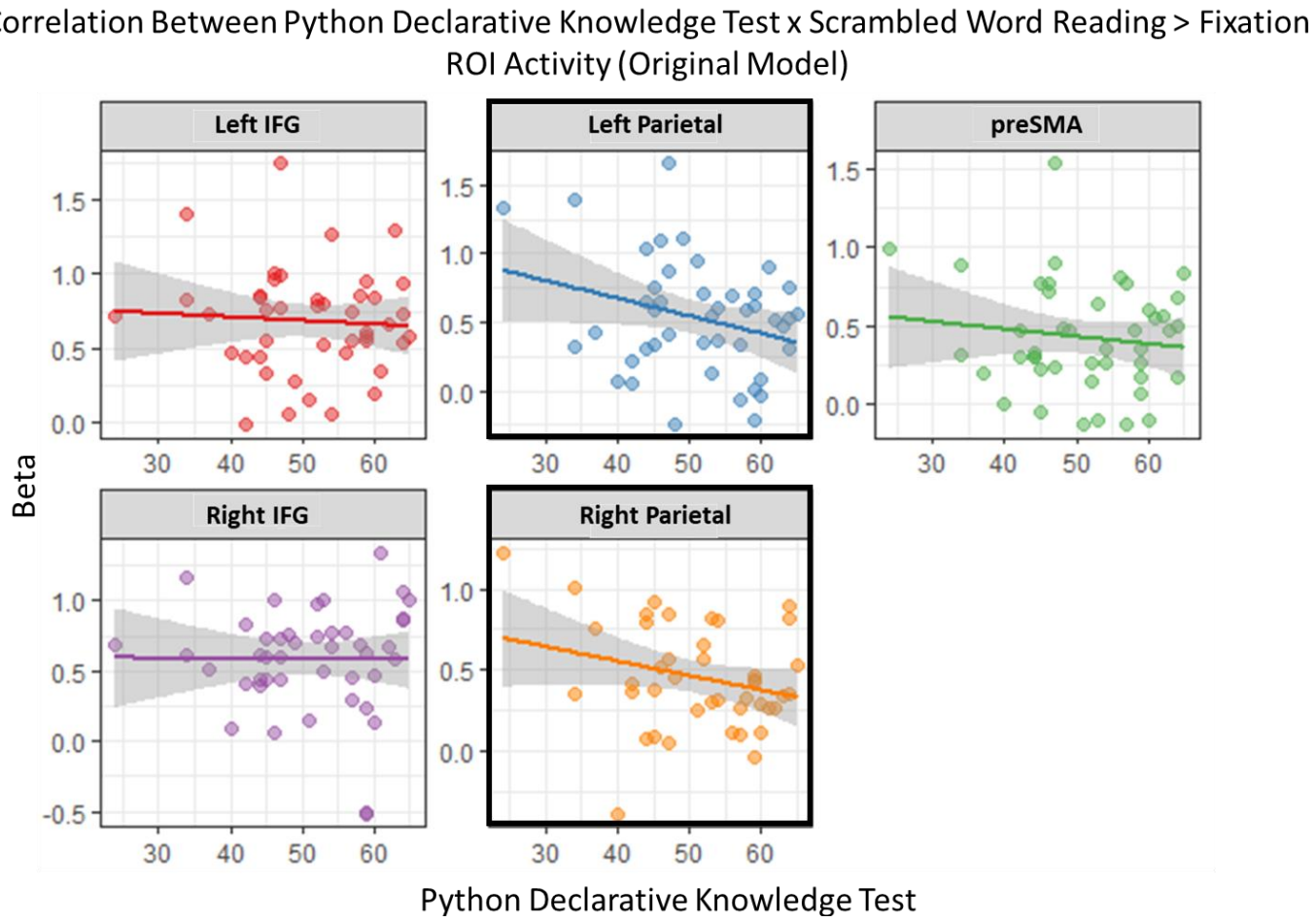
Scatterplots Depicting Correlation Between Nelson Denny Comprehension x Word Reading > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Scrambled Word Reading > Fixation in the phonological ROIs and Nelson Denny Reading Comprehension under the Original Model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7B.

Figure 27

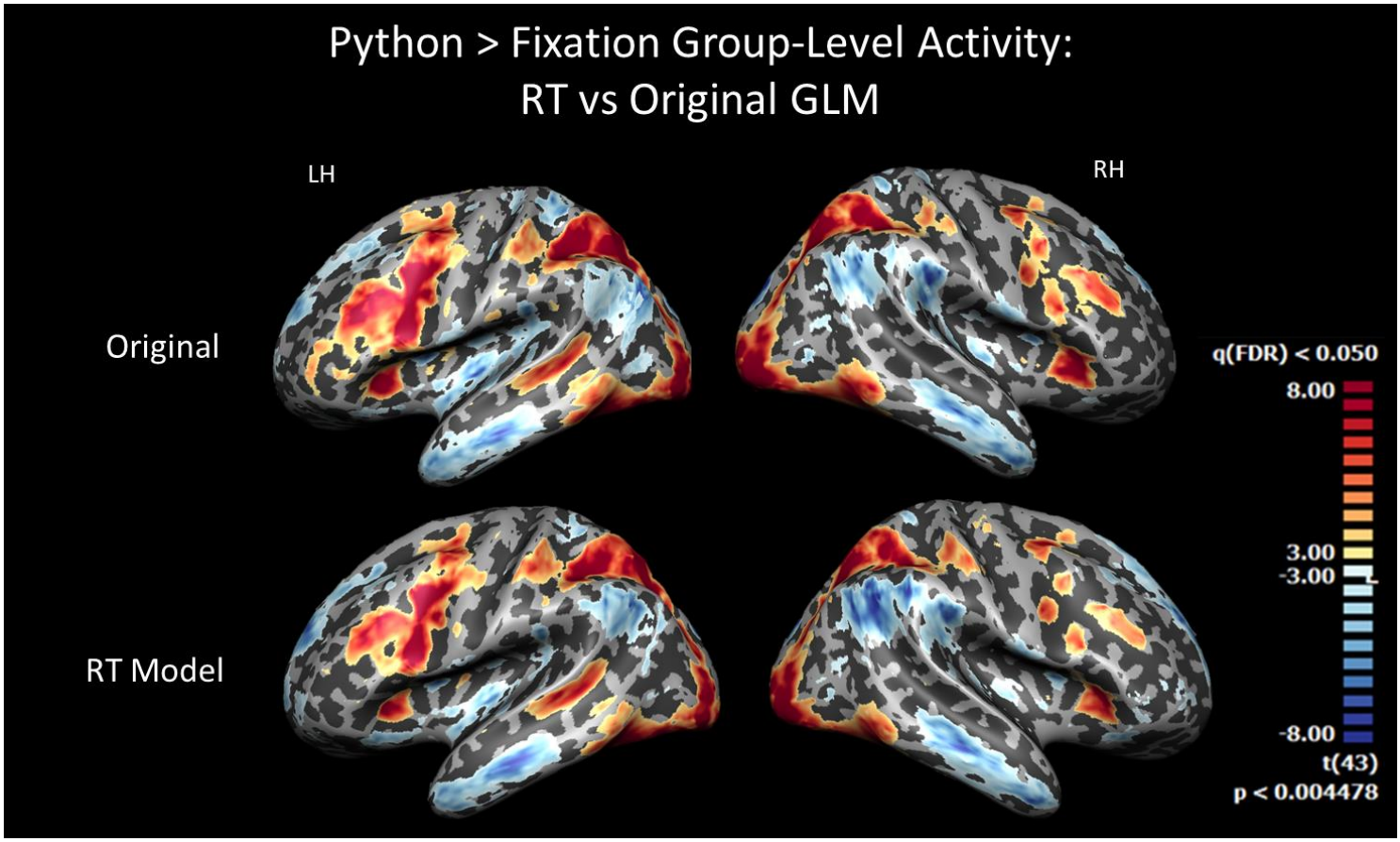
Scatterplots Depicting Correlation Between Python Declarative Knowledge Test x Word Reading > Fixation ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Scrambled Word Reading > Fixation in the phonological ROIs and Python Declarative Knowledge Test under the Original Model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7B.

Figure 28

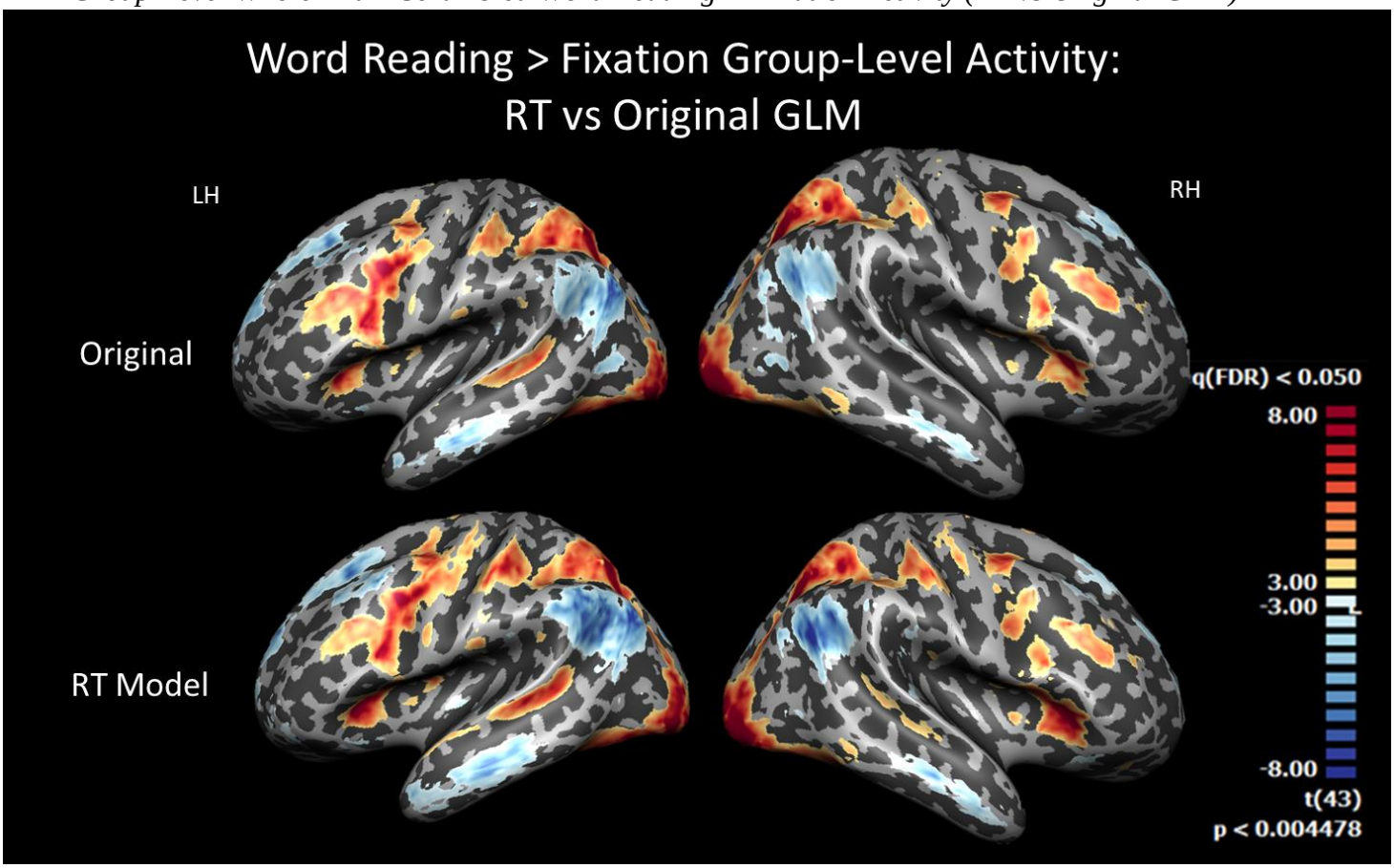
Group-Level Whole-Brain Python > Fixation Activity (RT vs Original GLM)



Note. Group-level whole-brain results GLM for the Python > Fixation. The Original model results are depicted in the top row and the RT model results are depicted in the bottom row. Warm colors represent regions where Python > Fixation and cool colors represent regions where Fixation > Python. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 6 for the Original model and in Table 10 for the RT model.

Figure 29

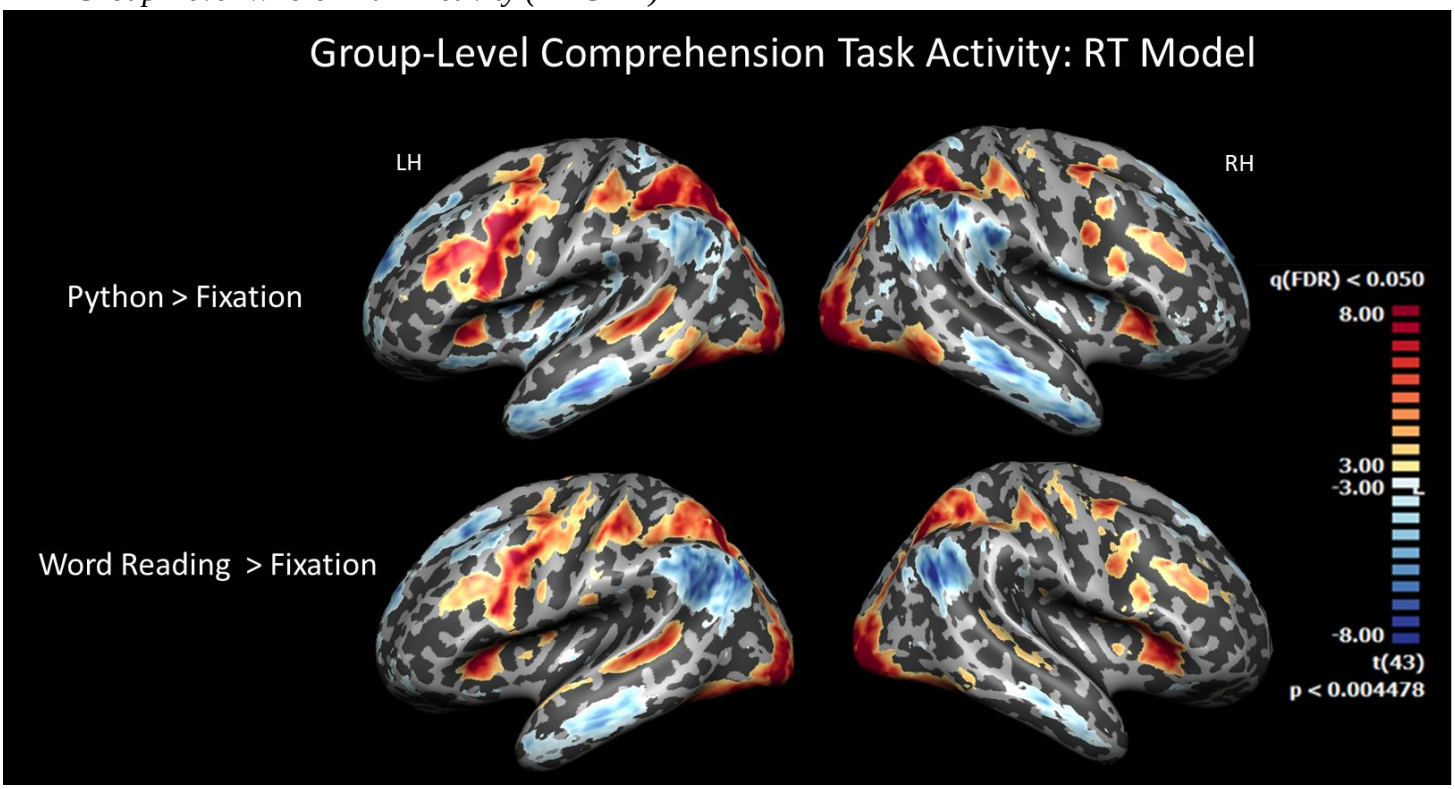
Group-Level Whole-Brain Scrambled Word Reading > Fixation Activity (RT vs Original GLM)



Note. Group-level whole-brain results GLM for the Scrambled Word Reading > Fixation. The Original model results are depicted in the top row and the RT model results are depicted in the bottom row. Warm colors represent regions where Scrambled Word Reading > Fixation and cool colors represent regions where Fixation > Scrambled Word Reading. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 9 for the Original model and in Table 11 for the RT model.

Figure 30

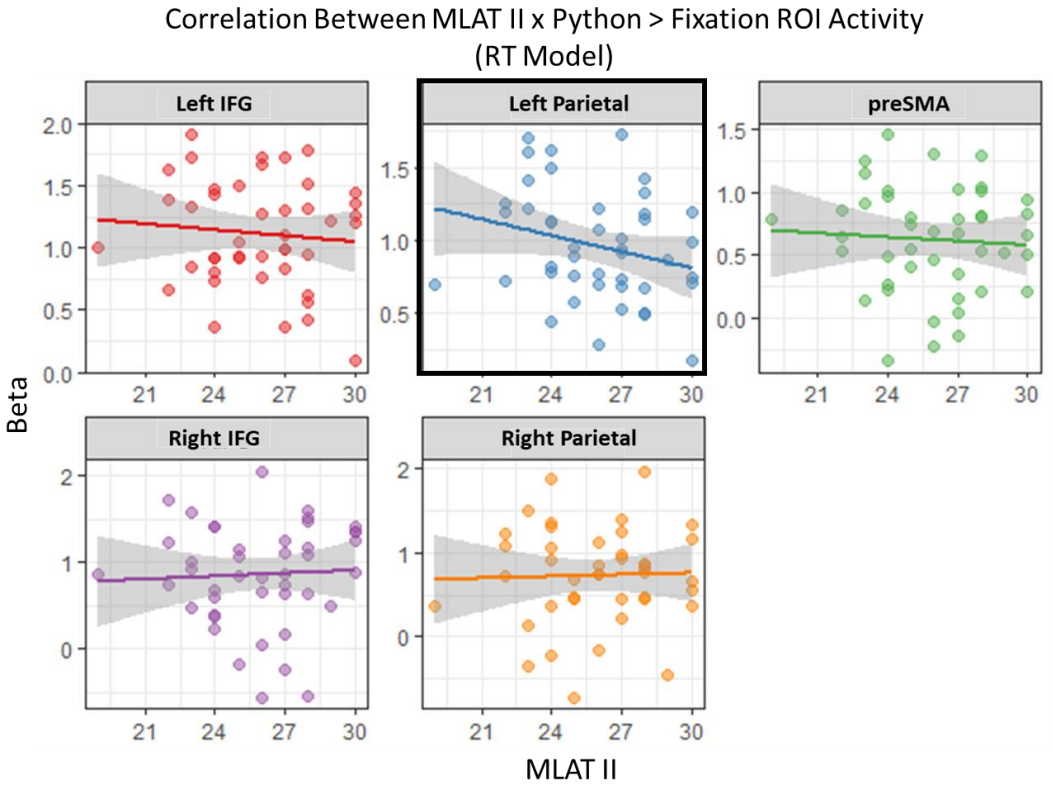
Group-Level Whole-Brain Activity (RT GLM)



Note. Group-level whole-brain results using the RT GLM for the Python > Fixation (top row) and Scrambled Word Reading (bottom row) contrasts. Warm colors represent regions where Task Condition > Fixation and cool colors represent regions where Fixation > Task Condition. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 10 for the Python > Fixation contrast and in Table 11 for the Scrambled Word Reading > Fixation contrast.

Figure 31

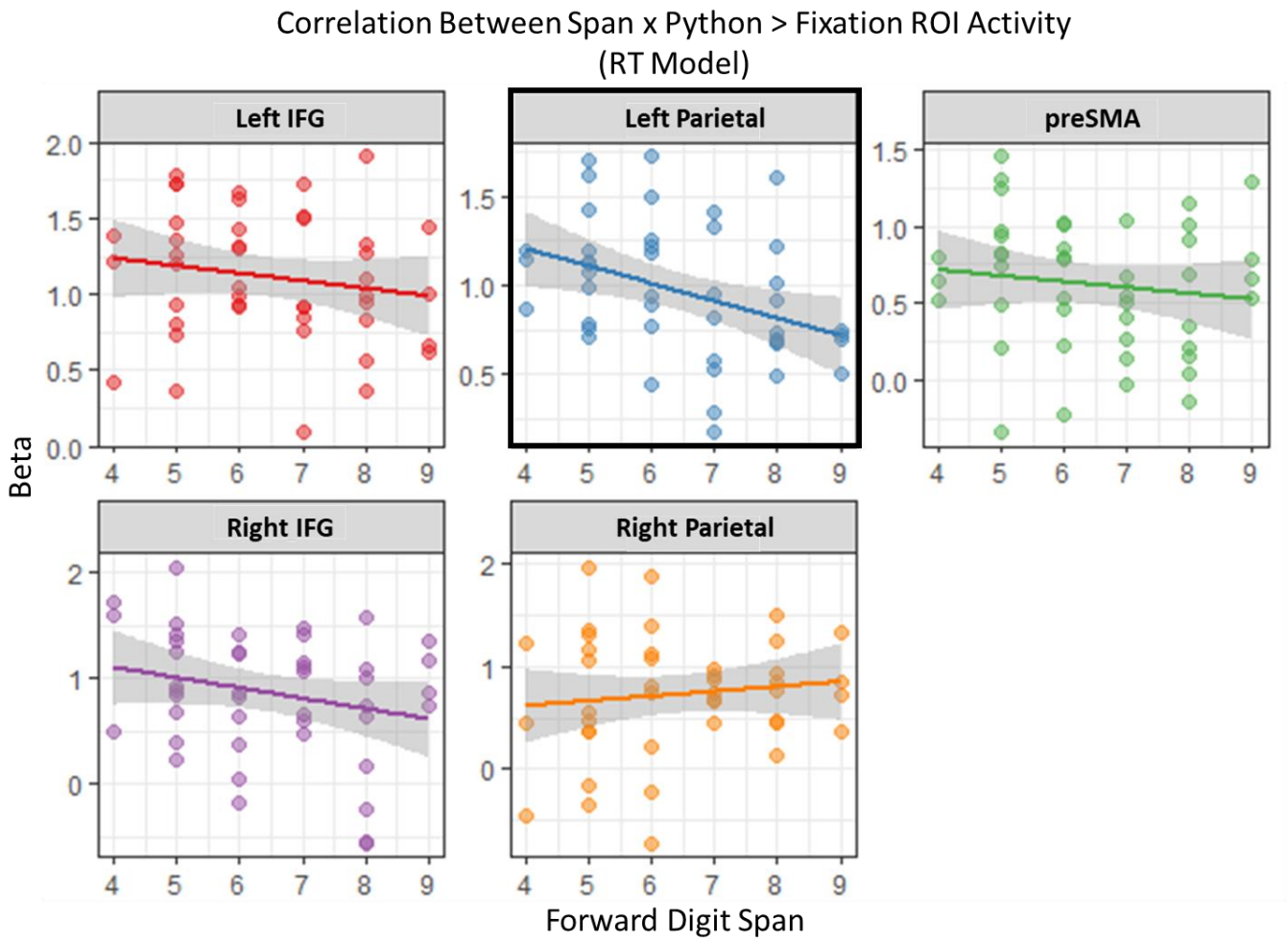
Scatterplots Depicting Correlation Between MLAT II x Python > Fixation ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and MLAT II under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12A.

Figure 32

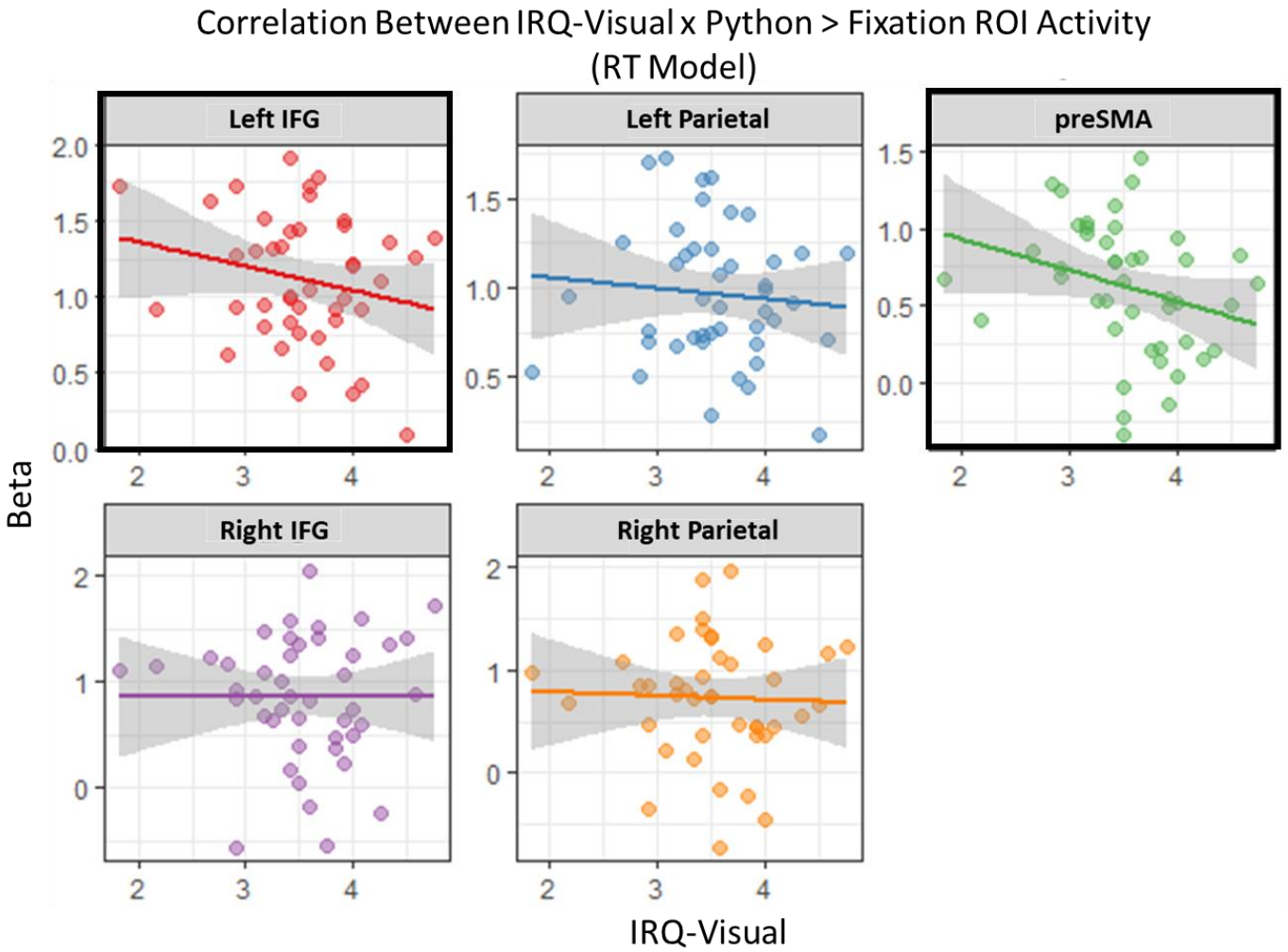
Scatterplots Depicting Correlation Between Span x Python > Fixation ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and Forward Digit Span under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12A.

Figure 33

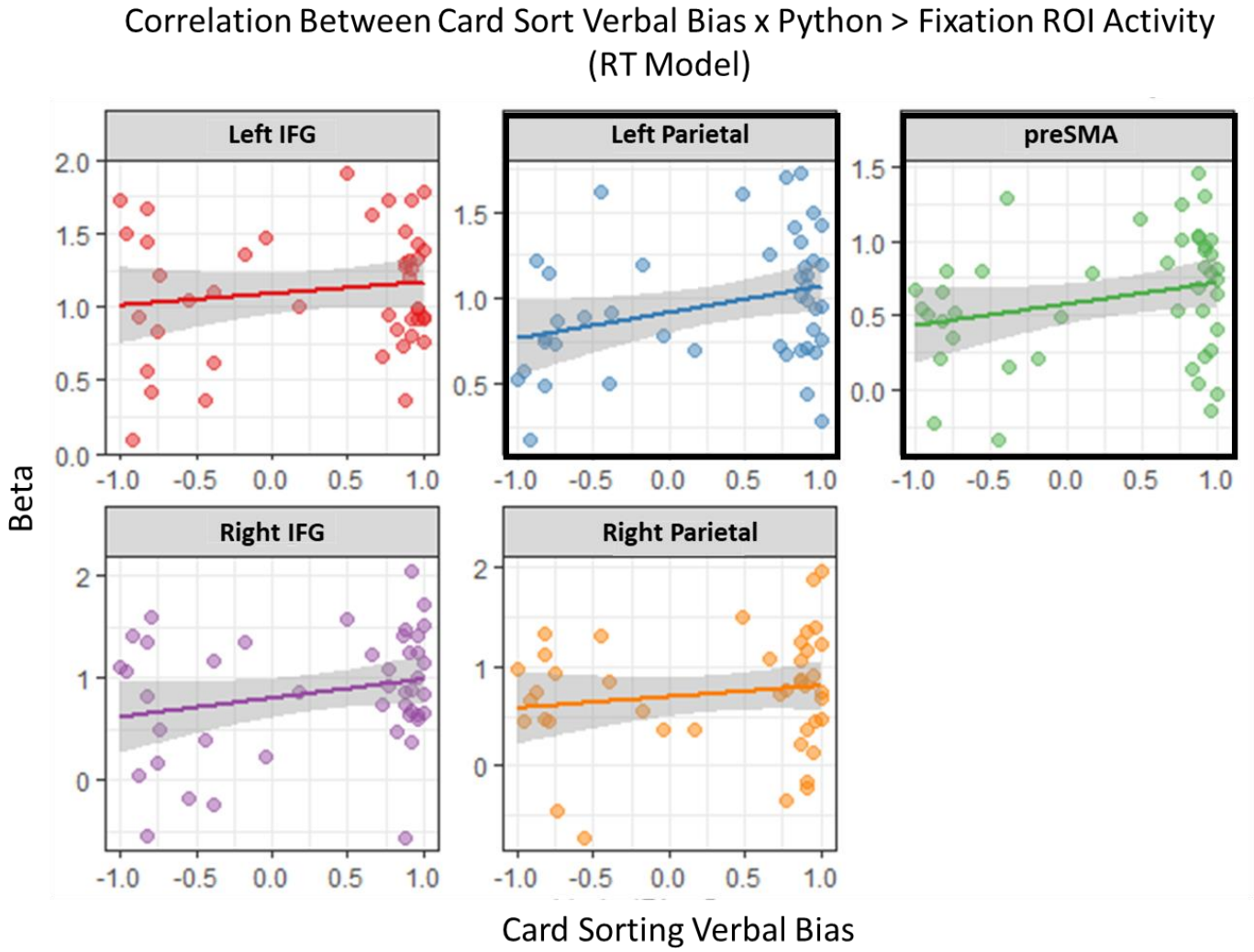
Scatterplots Depicting Correlation Between IRQ-Visual x Python > Fixation ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and IRQ-Visual under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12A.

Figure 34

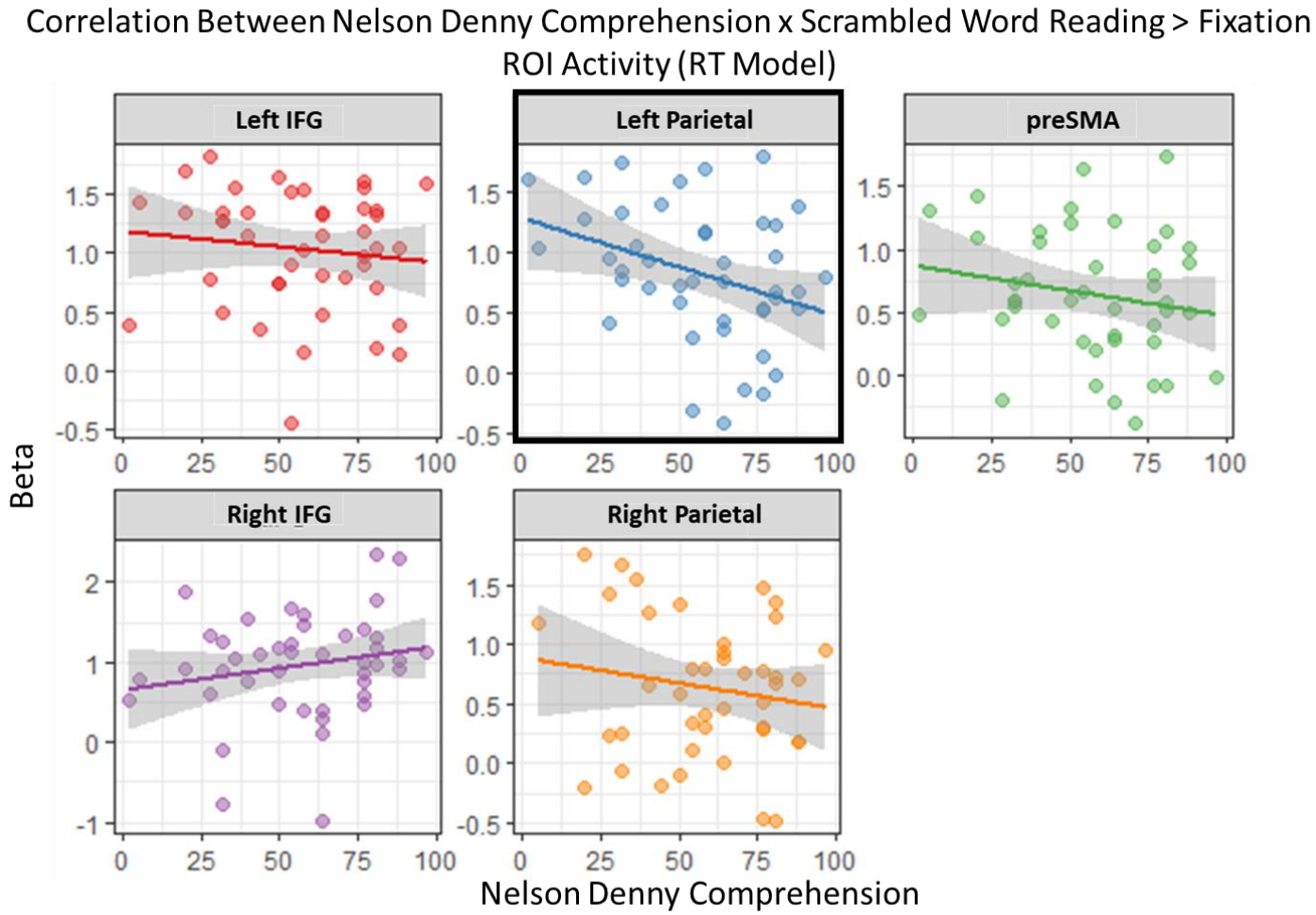
Scatterplots Depicting Correlation Between Card Sorting Verbal Bias x Python > Fixation ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Python > Fixation in the phonological ROIs and Verbal Bias on the Card Sorting task under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12A.

Figure 35

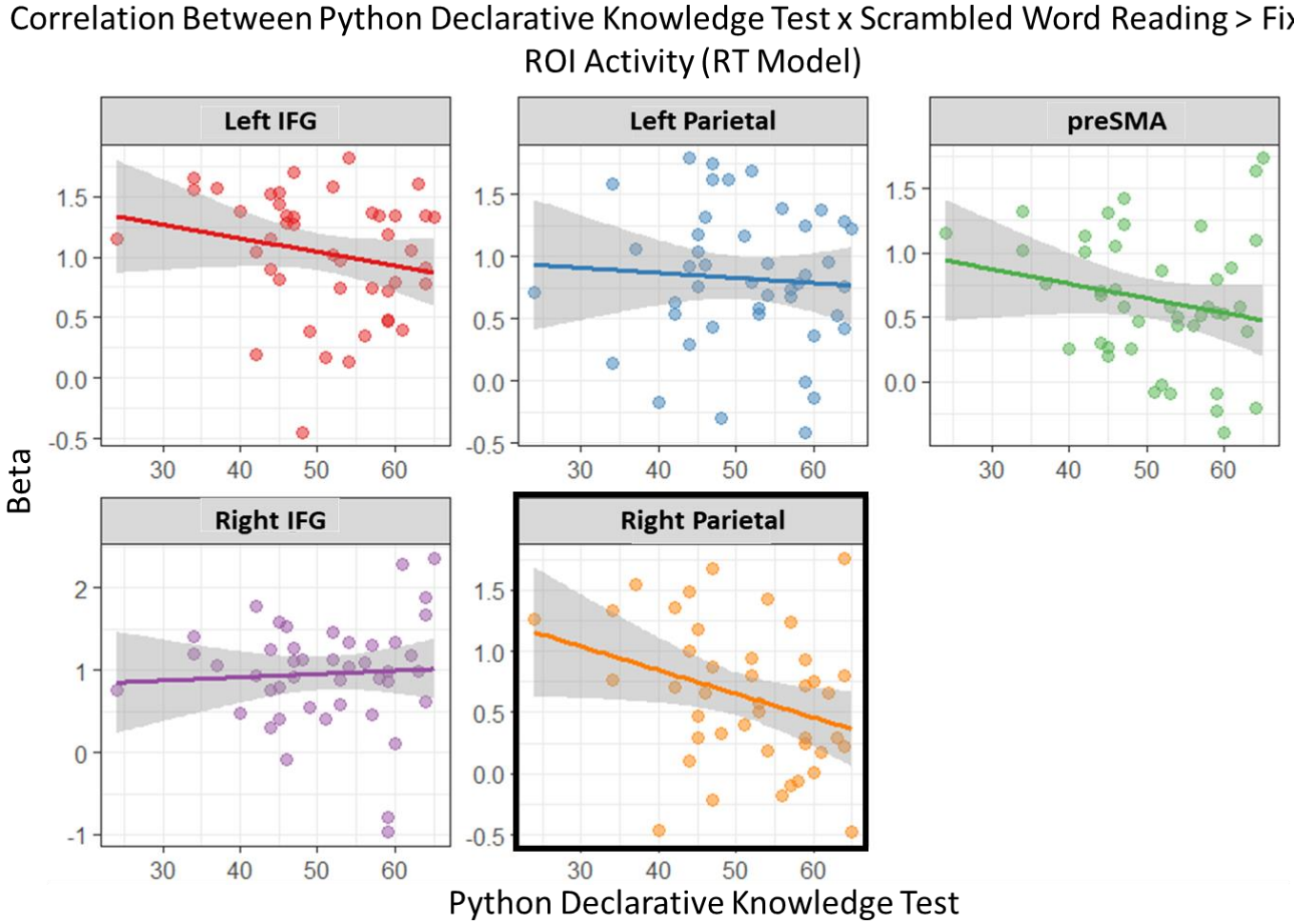
Scatterplots Depicting Correlation Between Nelson Denny Comprehension x Word Reading > Fixation ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Scrambled Word Reading > Fixation in the phonological ROIs and Nelson Denny Comprehension under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12B.

Figure 36

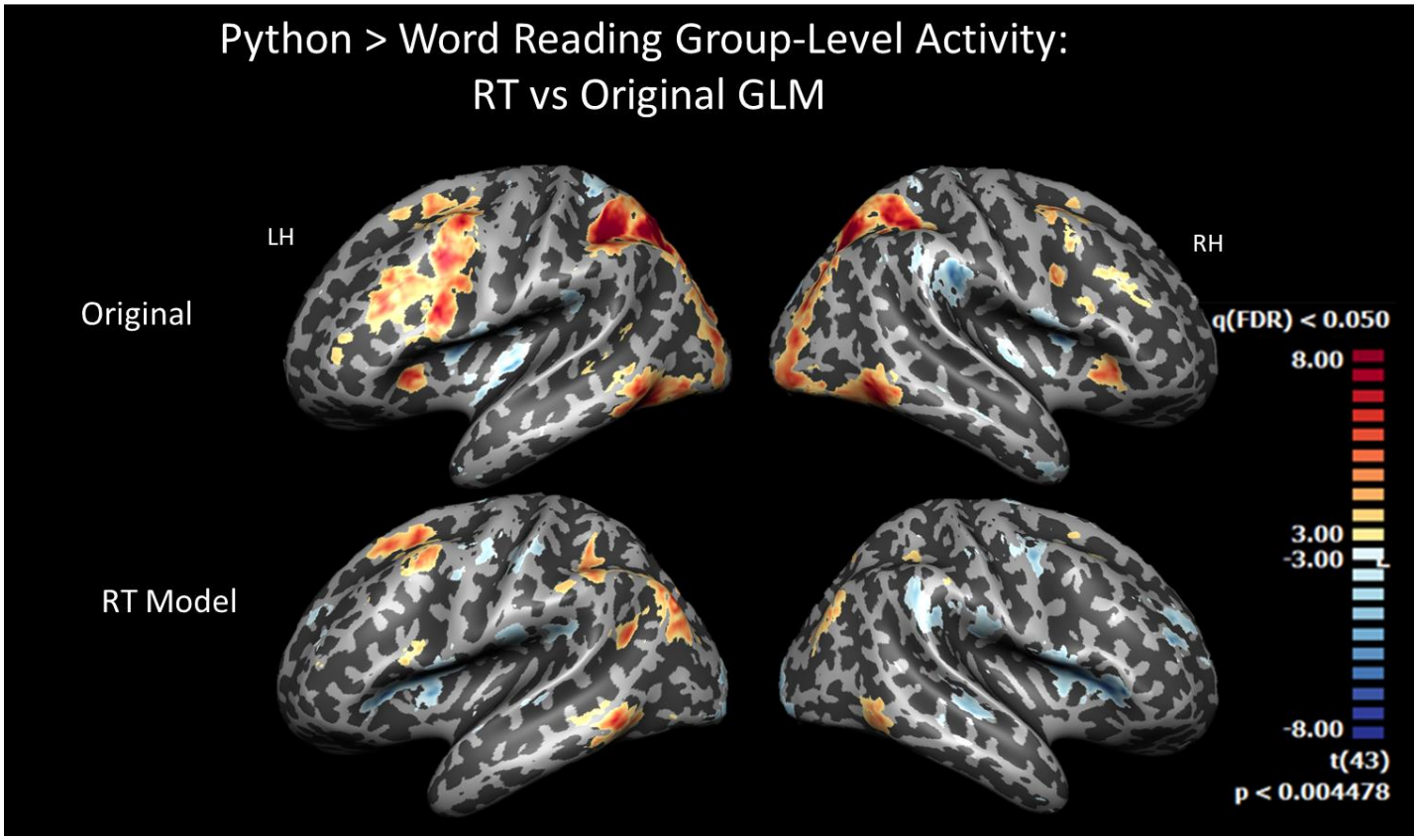
Scatterplots Depicting Correlation Between Python Declarative Knowledge Test x Word Reading > Fixation ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Scrambled Word Reading > Fixation in the phonological ROIs and Python Declarative Knowledge test under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12B.

Figure 37

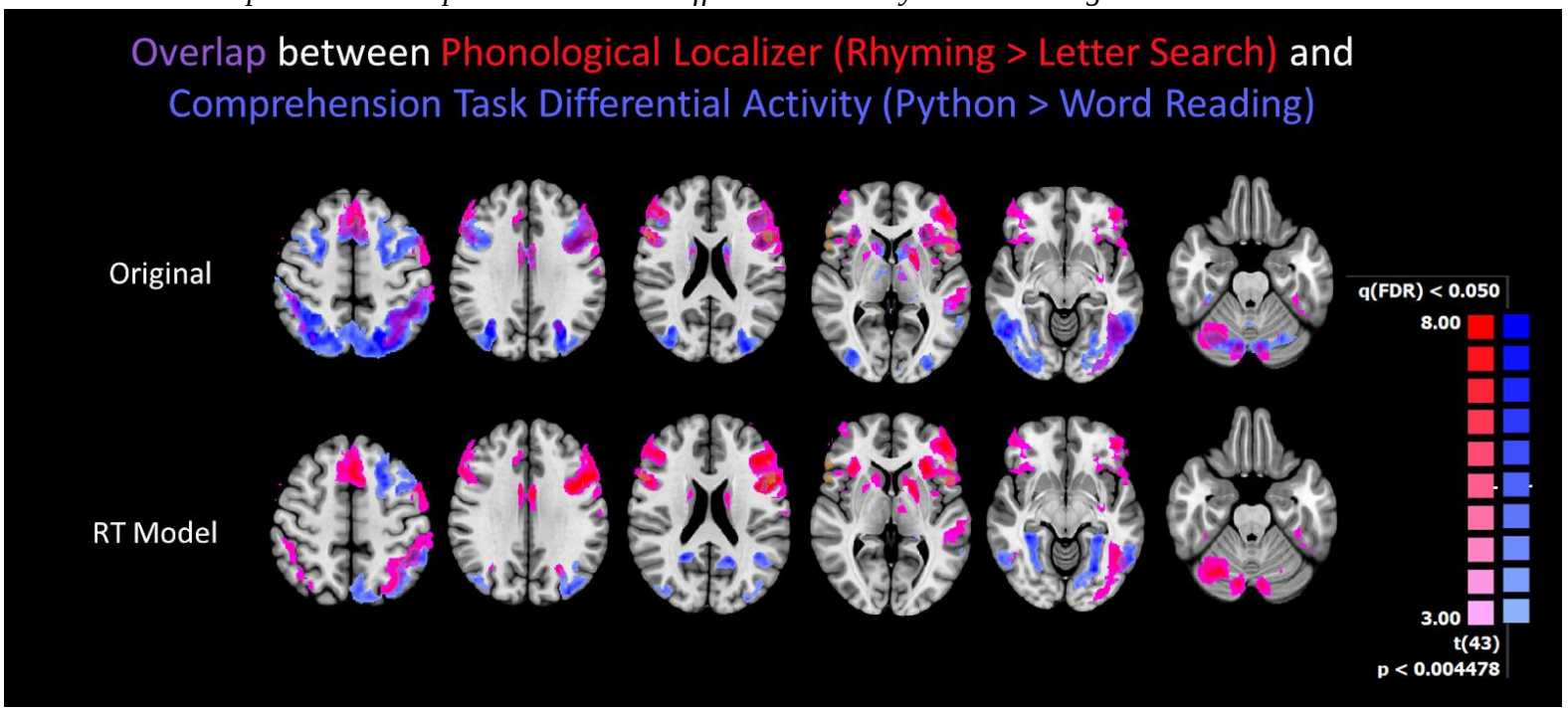
Group-Level Whole-Brain Python > Word Reading Activity (RT vs Original GLM)



Note. Group-level whole-brain results GLM for the Python > Scrambled Word Reading contrast. The Original model results are depicted in the top row and the RT model results are depicted in the bottom row. Warm colors represent regions where Python > Scrambled Word Reading and cool colors represent regions where Scrambled Word Reading > Python. All results are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. A full list of clusters significant above this threshold is provided in Table 15 for the Original model and in Table 16 for the RT model.

Figure 38

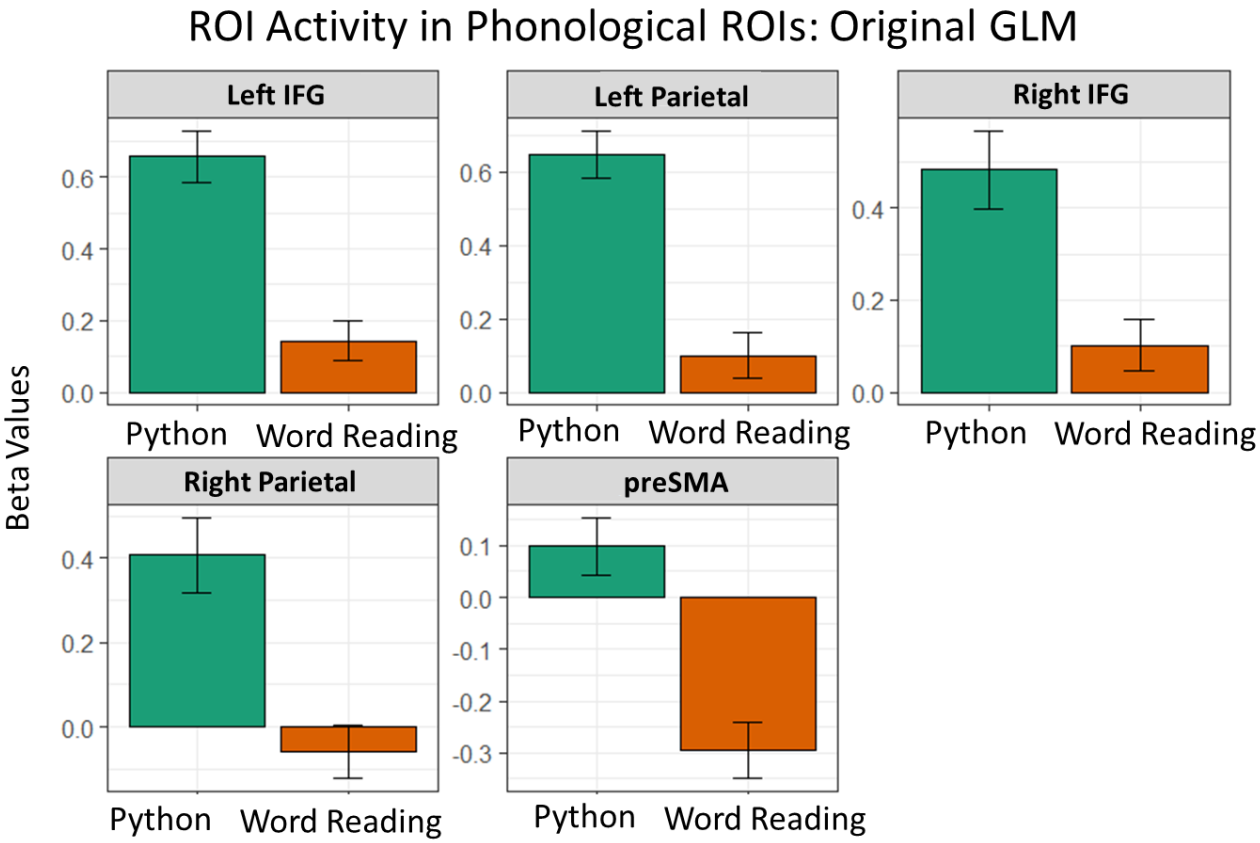
Overlap between Comprehension Task Differential Activity and Phonological Localizer Task



Note. Overlap between whole-brain group-level activity for the Phonological Localizer task (red) and the differential Python > Scrambled Word Reading activity from the Comprehension task (blue). Overlap under the Original model is depicted in the top row and overlap under the RT model is depicted in the bottom row. Both statistical maps are FDR corrected $p < 0.05$ with a t threshold = 3; extent threshold = 30 voxels. Overlap between the maps is depicted in purple.

Figure 39

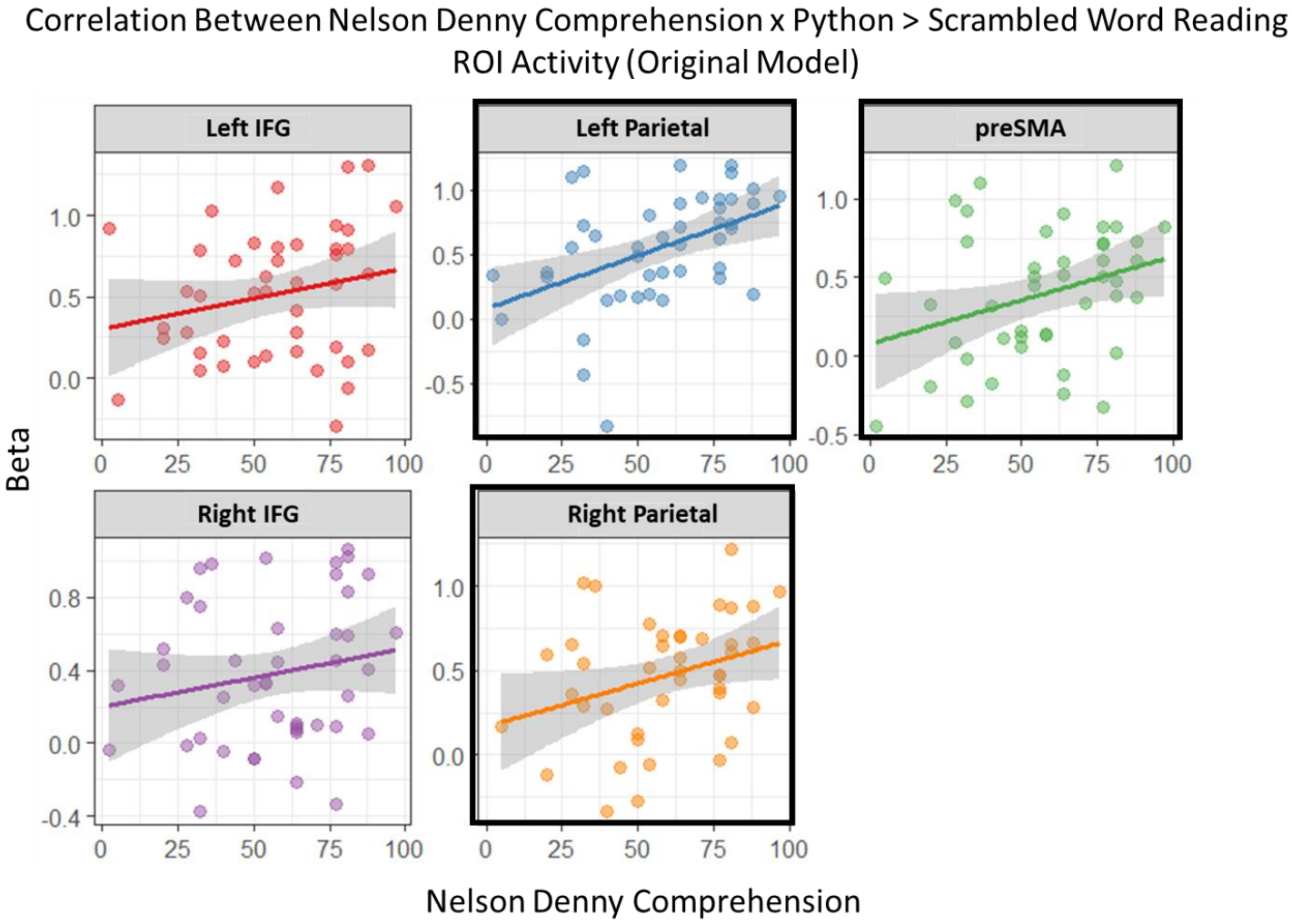
Group-level ROI Activity for Python and Scrambled Word Reading Conditions (Original Model)



Note. Bar plots depicting the average beta value for the Python (green) and Scrambled Word Reading (orange) conditions in the phonologically localized ROIs under the Original model. The beta values for Python were significantly larger than for Scrambled Word Reading in all ROIs. Error bars denote standard error.

Figure 40

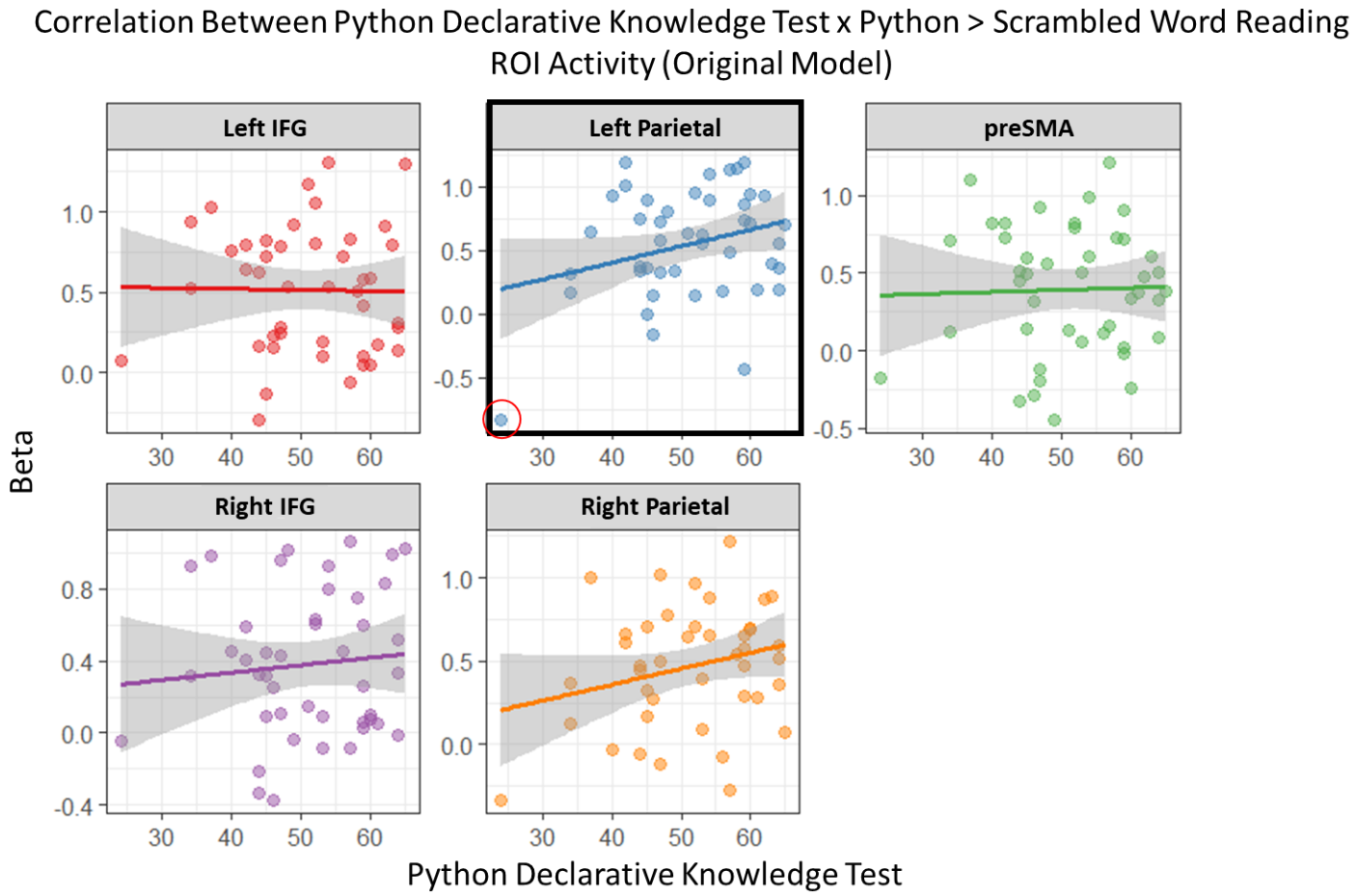
Scatterplots Depicting Correlation Between Nelson Denny Comprehension x Python >Word Reading ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Python > Scrambled Word Reading in the phonological ROIs and Nelson Denny Comprehension under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7C.

Figure 41

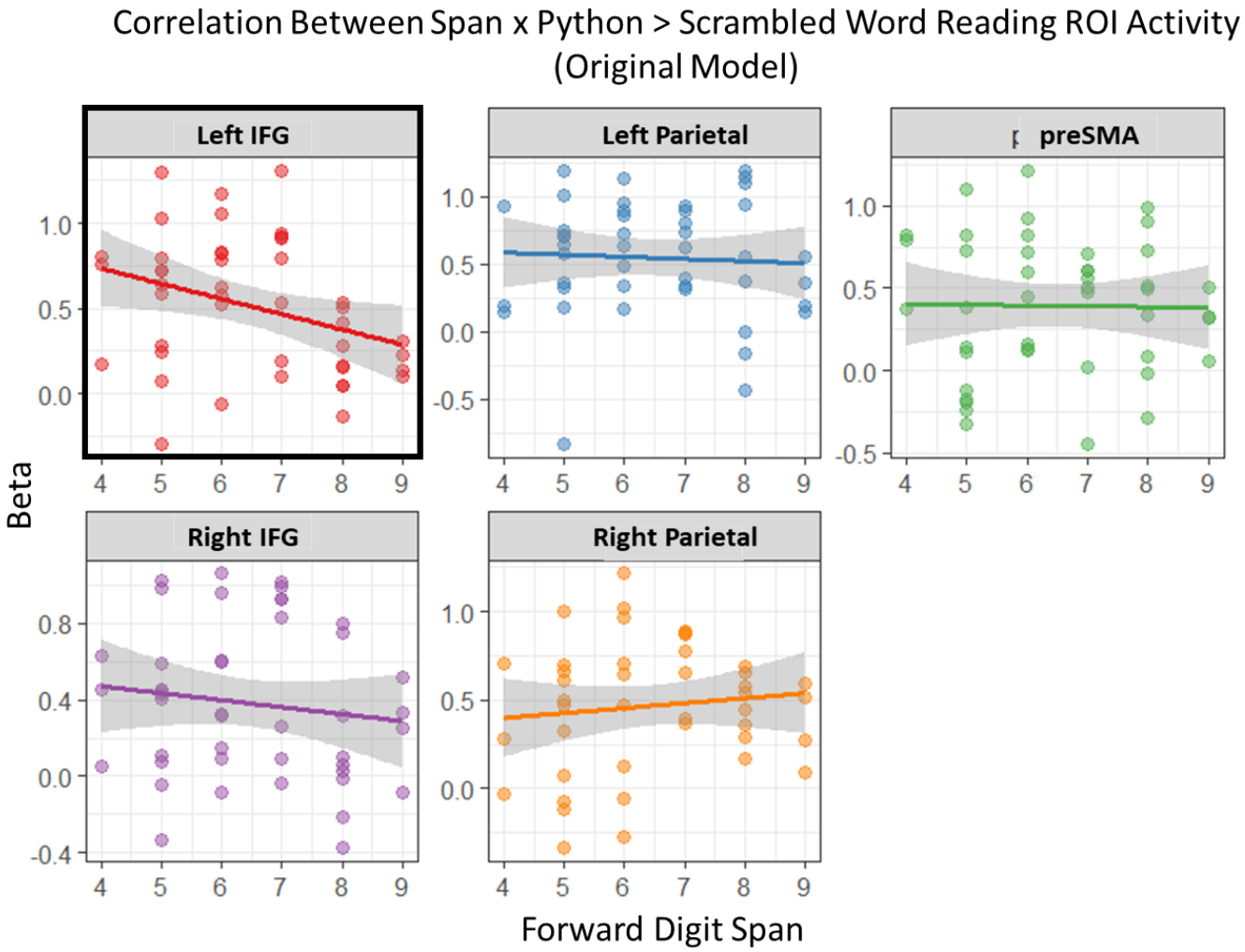
Scatterplots Depicting Correlation Between Python Declarative Knowledge Test x Python >Word Reading ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Python > Scrambled Word Reading in the phonological ROIs and Python Declarative Knowledge test under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7C. Note that this trend is driven by an outlier circled in red.

Figure 42

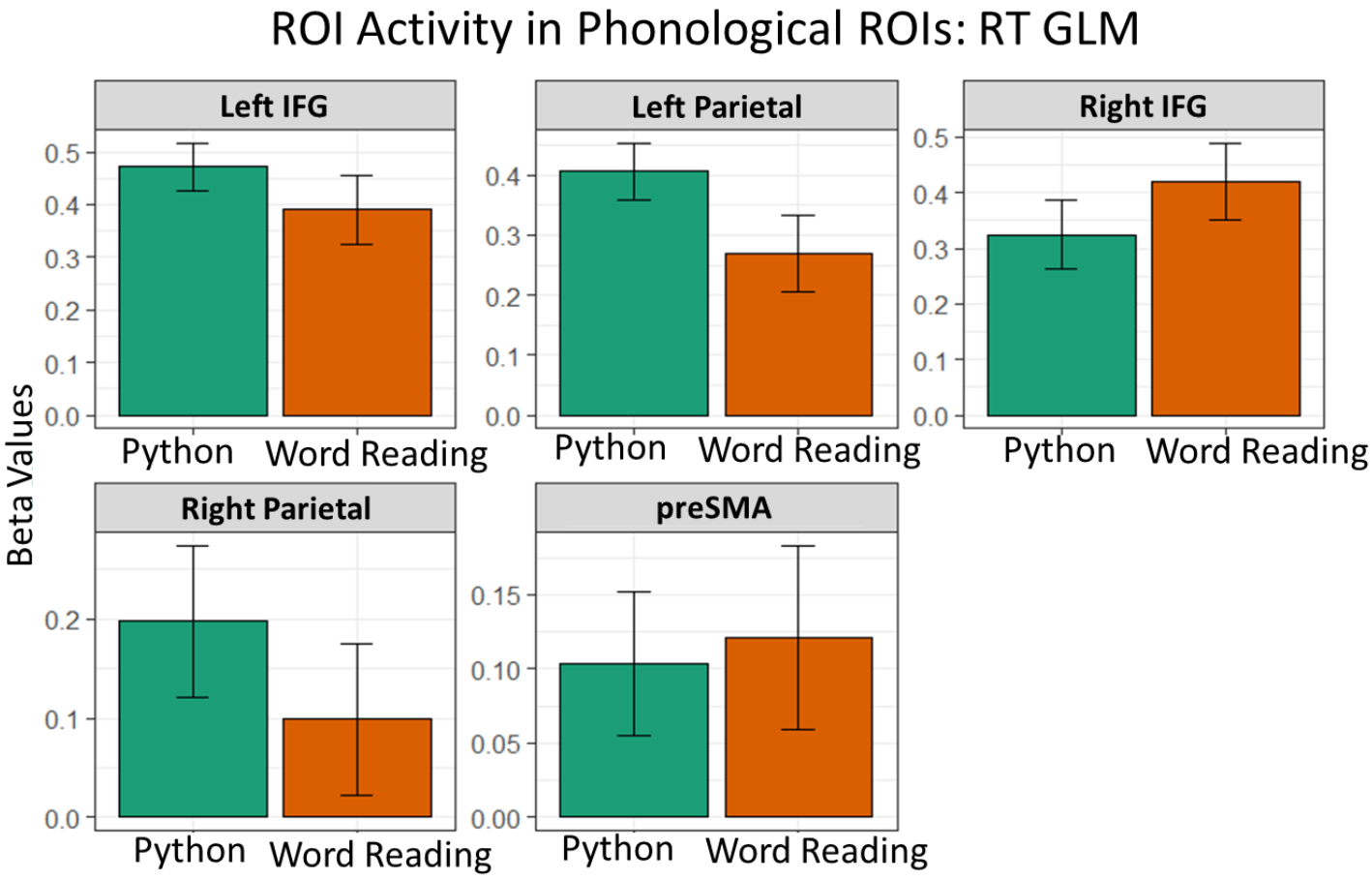
Scatterplots Depicting Correlation Between Span x Python >Word Reading ROI Activity (Original Model)



Note. Scatterplots depicting correlation between Python > Scrambled Word Reading in the phonological ROIs and Forward Digit Span under the Original model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 7C.

Figure 43

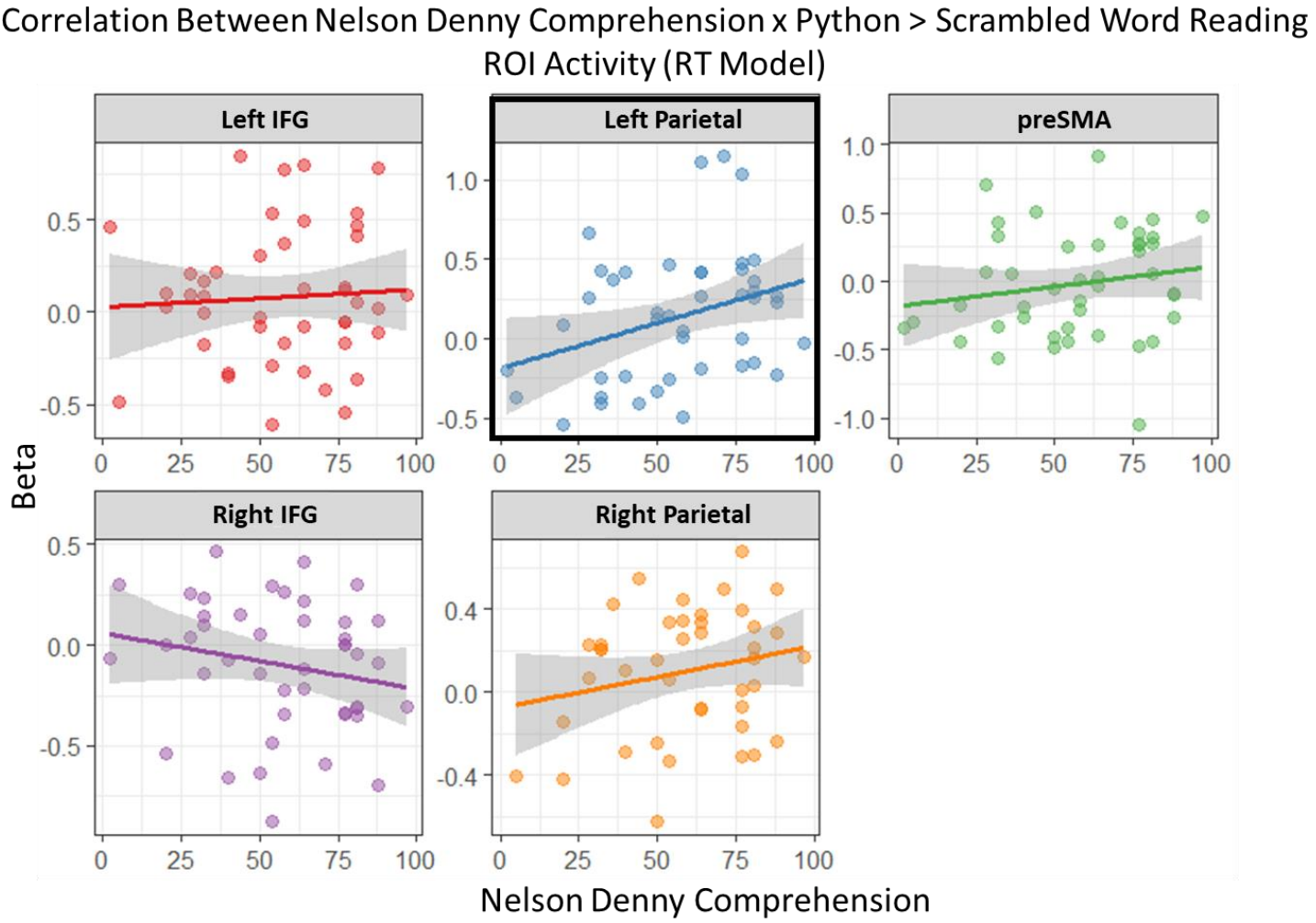
Group-level ROI Activity for Python and Scrambled Word Reading Conditions (RT Model)



Note. Bar plots depicting the average beta value for the Python (green) and Scrambled Word Reading (orange) conditions in the phonologically localized ROIs under the RT model. The beta values for Python were significantly larger than for Scrambled Word Reading in the left parietal and right parietal ROIs. The difference between conditions in the left IFG, right IFG, and preSMA ROIs did not reach significance under the RT model. Error bars denote standard error.

Figure 44

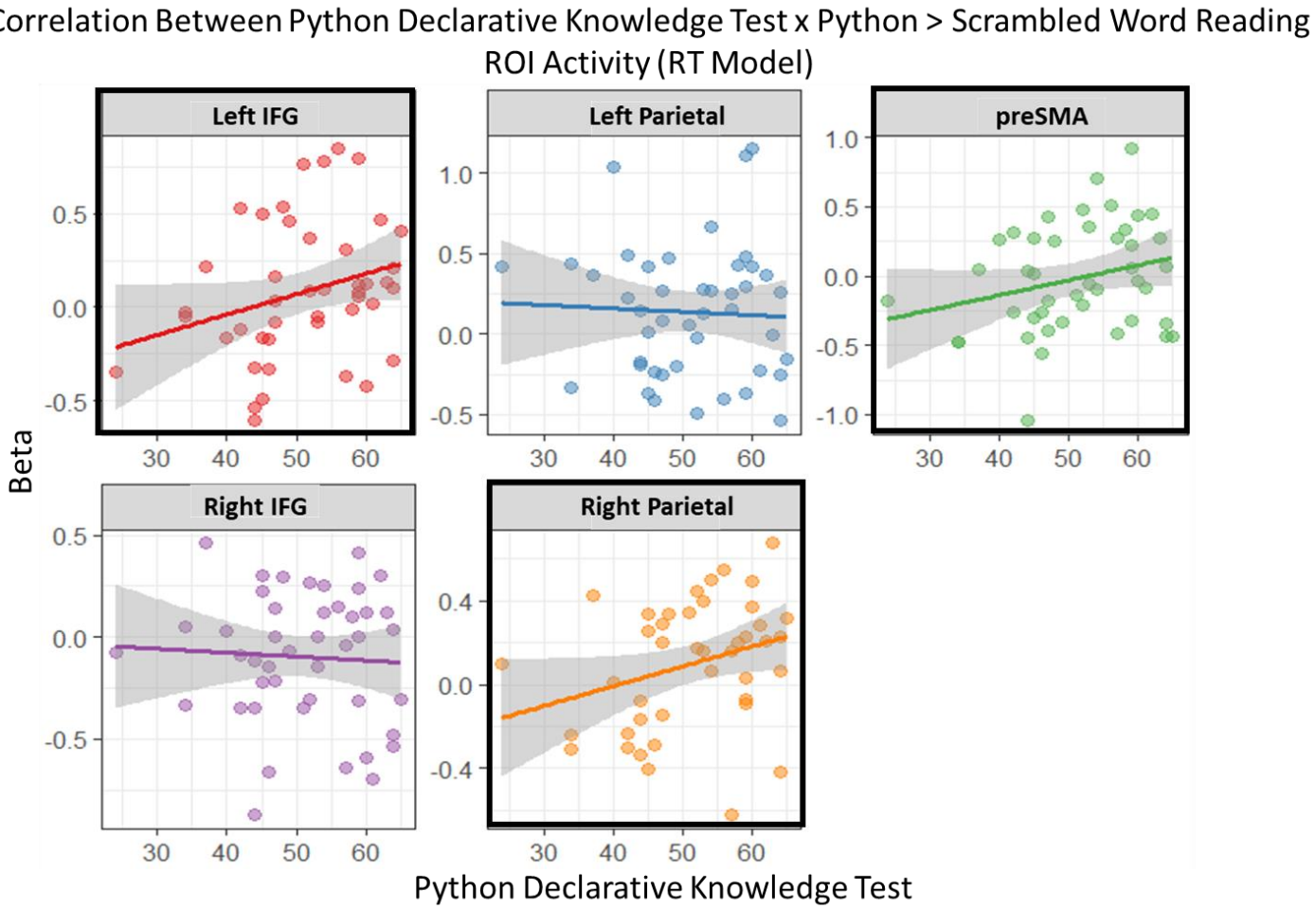
Scatterplots Depicting Correlation Between Nelson Denny Comprehension x Python >Word Reading ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Python > Scrambled Word Reading in the phonological ROIs and Nelson Denny Comprehension under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12C.

Figure 45

Scatterplots Depicting Correlation Between Python Declarative Knowledge Test x Python >Word Reading ROI Activity (RT Model)



Note. Scatterplots depicting correlation between Python > Scrambled Word Reading in the phonological ROIs and Python Declarative Knowledge test under the RT model. The black boxes indicate correlations that are significant or marginally significant for correlation coefficients see Table 12C.

Appendices

Appendix A. Modified Internal Response Questionnaire (IRQ)

Question	Visual / Verbal Loading
1. I often enjoy the use of mental pictures to reminisce	Visual
2. I think about problems in my mind in the form of a conversation with myself	Verbal
3. I can close my eyes and easily picture a scene that I have experienced	Visual
4. My mental images are very vivid and photographic	Visual
5. When I write code, I talk to myself about what I will write next	Verbal*
6. If I am walking somewhere by myself, I often have a silent conversation with myself	Verbal
7. If I am walking somewhere by myself, I frequently think of conversations that I've recently had	Verbal
8. The old saying "A picture is worth a thousand words" is certainly true for me	Visual
9. My inner speech helps my imagination	Verbal
10. When I debug code, I can visualize what the corrected code should look like	Visual*
11. When I think about someone I know well, I instantly see their face in my mind	Visual
12. I tend to think things through verbally when I am relaxing	Verbal
13. When thinking about a social problem, I often talk it through in my head	Verbal
14. I like to give myself some downtime to talk through thoughts in my mind	Verbal
15. I often use mental images or pictures to help me remember things	Visual
16. My memories are mainly visual in nature	Visual
17. When I write code I see in my "mind's eye" what the code will look like	Visual*
18. I hear words in my "mind's ear" when I think	Verbal
19. When traveling to get somewhere I tend to think more visually than verbally	Visual
20. I rarely vocalize thoughts in my mind	Verbal (reverse scored)
21. If I talk to myself in my head, it is usually accompanied by visual imagery	Visual
22. I often talk to myself internally while watching TV	Verbal
23. When I debug code, narrating the code to myself helps me to find errors	Verbal*
24. My memories often involve conversations I've had	Verbal
25. If I imagine my memories visually, they are often more static than moving	Visual
26. When I read, I tend to hear a voice in my "mind's ear"	Verbal

Note. *Denotes questions added to the IRQ, all other questions are from Roebuck & Lupyan, 2020

Appendix B. Word properties of Comprehension Task Stimuli

Occurrences	Word	Length	HAL Frequency (log)
1	adult	5	10.136
1	Amy	3	8.995
1	apartment	9	9.101
1	apple	5	11.095
1	apples	6	8.29
1	are	3	14.909
1	arms	4	10.395
1	bacon	5	8.054
1	ballet	6	7.42
1	banana	6	7.965
1	basketball	10	9.193
1	Billy	5	9.281
1	bird	4	9.856
1	blue	4	11.396
1	boots	5	9.28
1	bread	5	9.112
1	broom	5	6.397
1	bus	3	10.485
1	butterfly	9	7.494
1	carrot	6	7.269
1	cat	3	10.562
1	cheese	6	9.067
1	chicken	7	9.348
1	child	5	11.078
1	chocolate	9	9.161
1	circle	6	9.898
1	Claire	6	7.734
1	cookie	6	7.991
1	couch	5	8.511

1	crackers	8	7.194
1	cucumber	8	6.537
1	cupcake	7	4.369
1	daisy	5	7.193
1	denied	6	9.272
1	do	2	14.123
1	doctor	6	10.371
1	dog	3	10.974
1	dormitory	9	5.635
1	dove	4	7.906
1	dress	5	9.424
1	drums	5	8.717
1	elm	3	8.423
1	fail	4	9.835
1	fall	4	10.738
1	fit	3	10.661
1	flour	5	8.439
1	flute	5	8.158
1	fork	4	9.328
1	from	4	14.455
1	George	6	10.74
1	granted	7	9.81
1	grass	5	8.903
1	green	5	11.416
1	gymnastics	10	6.452
1	hammer	6	8.812
1	harry	5	9.408
1	head	4	11.6
1	hockey	6	9.409
1	house	5	11.554
1	it	2	15.365
1	Jeff	4	10.642
1	Jose	4	9.964
1	knife	5	8.871
1	lamp	4	8.803
1	Larry	5	10.218

1	legs	4	10.183
1	lemon	5	8.298
1	length	6	10.835
1	lettuce	7	6.877
1	lily	4	7.88
1	mango	5	6.765
1	many	4	13.071
1	milk	4	9.629
1	mop	3	6.583
1	muffin	6	6.899
1	Nancy	5	9.219
1	not	3	14.773
1	oak	3	8.772
1	octopus	7	6.802
1	only	4	13.627
1	orange	6	9.509
1	out	3	13.902
1	pass	4	10.791
1	pear	4	6.878
1	piano	5	9.135
1	pilot	5	9.682
1	pine	4	8.977
1	plate	5	9.251
1	red	3	11.55
1	rice	4	9.713
1	robin	5	9.457
1	rose	4	9.869
1	sandals	7	7.598
1	sandwich	8	7.753
1	says	4	12.203
1	scissors	8	7.222
1	shark	5	8.088
1	shoe	4	8.558
1	shorts	6	8.618
1	sneakers	8	6.765
1	soccer	6	8.923

1	soda	4	8.122
1	sold	4	12.154
1	sponge	6	7.376
1	spoon	5	7.872
1	spring	6	10.172
1	square	6	9.852
1	steak	5	7.32
1	summer	6	10.448
1	sunglasses	10	8.14
1	Susan	5	9.697
1	swan	4	7.397
1	swimsuit	8	6.567
1	teacher	7	9.738
1	too	3	12.87
1	towel	5	8.131
1	train	5	10.059
1	triangle	8	8.601
1	trumpet	7	8.876
1	vanilla	7	8.136
1	whale	5	7.728
1	who	3	13.745
1	you	3	15.431
2	access	6	12.057
2	activities	10	10.397
2	append	6	7.858
2	bakery	6	6.299
2	body	4	11.663
2	cakes	5	7.2
2	calories	8	7.864
2	can	3	14.301
2	cleaning	8	9.165
2	clock	5	10.106
2	clothes	7	9.571
2	club	4	11.055
2	coffee	6	9.819
2	colors	6	10.041

2	field	5	11.289
2	fish	4	10.329
2	food	4	10.997
2	forest	6	9.747
2	furniture	9	8.456
2	grade	5	9.605
2	grape	5	7.401
2	have	4	14.915
2	hot	3	10.85
2	housing	7	9.077
2	instrument	10	9.224
2	instruments	11	9.517
2	job	3	11.933
2	juice	5	8.943
2	long	4	12.521
2	Mary	4	10.133
2	meal	4	8.784
2	most	4	13.113
2	my	2	14.386
2	occupations	11	6.97
2	ocean	5	9.31
2	offers	6	10.883
2	option	6	10.819
2	order	5	12.174
2	pantry	6	5.361
2	pants	5	9.223
2	pastry	6	6.491
2	points	6	11.468
2	produce	7	10.451
2	protein	7	9.65
2	range	5	10.971
2	shapes	6	8.07
2	shirt	5	9.006
2	short	5	11.467
2	soup	4	8.707
2	sports	6	10.167

2	student	7	10.941
2	students	8	10.929
2	supermarket	11	7.421
2	team	4	11.595
2	temperature	11	9.565
2	tools	5	11.013
2	upper	5	10.195
2	vase	4	6.292
2	vegetables	10	8.387
2	water	5	11.571
2	while	5	12.558
3	Alice	5	8.65
3	amount	6	11.033
3	beach	5	10.191
3	birds	5	9.581
3	breakfast	9	8.939
3	buy	3	11.778
3	car	3	11.365
3	clothing	8	9.215
3	dogs	4	10.393
3	eggs	4	9.014
3	flavor	6	8.542
3	flower	6	8.746
3	fruit	5	9.281
3	fruits	6	8.384
3	hello	5	10.637
3	hour	4	10.754
3	index	5	10.387
3	item	4	10.468
3	kitchen	7	9.182
3	Max	3	10.122
3	meat	4	9.606
3	menu	4	10.26
3	oldest	6	8.44
3	part	4	12.425
3	pets	4	8.49

3	plant	5	10.136
3	score	5	10.055
3	shoes	5	9.356
3	the	3	16.955
4	beverages	9	6.918
4	digit	5	8.113
4	doubled	7	8.06
4	drink	5	9.897
4	friends	7	11.256
4	greeting	8	7.567
4	initials	8	7.347
4	is	2	15.683
4	lengths	7	8.147
4	more	4	13.889
4	snack	5	7.42
4	squares	7	7.891
4	total	5	11.162
4	transportation	14	9.188
4	tree	4	10.212
4	urgent	6	8.246
4	word	4	11.836
5	age	3	11.11
5	number	6	12.553
5	tool	4	10.413
6	I	1	15.79
6	sum	3	9.523
7	value	5	11.325
8	cats	4	9.785
8	values	6	10.435
9	else	4	12.081
12	name	4	12.604
12	numbers	7	11.203
19	if	2	14.653
30	for	3	15.417
37	in	2	15.897

72	print	5	11.032