Fall 12-21-2018

Non-symbolic exact quantity representation in a language-impaired population

John Verbos

NON-SYMBOLIC EXACT QUANTITY REPRESENTATION IN A LANGUAGE-IMPAIRED POPULATION

A Dissertation

Submitted to the McAnulty College and Graduate School of Liberal Arts

Duquesne University

In partial fulfillment of the requirements for

the degree of Doctor of Philosophy

By

John Verbos

December 2018

NON-SYMBOLIC EXACT QUANTITY REPRESENTATION IN A

LANGUAGE-IMPAIRED POPULATION

By

John Verbos

Approved August 29, 2018

Alexander Kranjec, Ph.D.
Assistant Professor of Psychology
(Committee Co-Chair)

Sarah E. Wallace Ph.D., CCC-SLP
Associate Professor of Speech-Language Pathology
(Committee Co-Chair)

Elizabeth Fein, Ph.D.
Assistant Professor of Psychology
(Committee Member)

James Swindal, Ph.D.
Dean, McAnulty College and Graduate School of Liberal Arts
Professor of Philosophy

Leswin Laubscher, Ph.D.
Chair, Psychology Department
Associate Professor of Psychology

ABSTRACT

NON-SYMBOLIC EXACT QUANTITY REPRESENTATION IN A LANGUAGE-IMPAIRED POPULATION

By

John Verbos

December 2018

Dissertation supervised by Alexander Kranjec, Ph.D., Sarah E. Wallace Ph.D., CCC-SLP, and Elizabeth Fein, Ph.D.

The linguistic relativity hypothesis argues that language influences non-linguistic cognition. One version of the hypothesis suggests that language is a set of tools or technologies that variously enhance or dampen an individual's capacity to perceive and operate upon the world in certain ways. A domain in which this may be tested is number, where it is hypothesized that counting language allows us to bridge our innate capacities for recognizing small exact quantities (subitizing) and approximating quantities larger than three or four (analog magnitude estimation). To test this, previous studies have asked adult participants who have limited or no access to counting language to represent non-symbolic exact quantities—that is, for participants to create an array of objects equal in number to a target array of objects presented to the participant. In these studies, both English-speakers whose access to number language was artificially compromised by verbal interference and the Pirahã—an Amazonian tribe whose language does not contain exact number words—appeared to rely on analog magnitude estimation for representing non-symbolic exact quantities greater than three. This suggests that the ability to consistently and accurately recognize and represent non-symbolic exact quantities is impaired by having limited or no access to counting language. Here, sixteen participants with left-hemisphere damage from stroke and resulting aphasia performed the same five non-verbal, non-symbolic matching tasks from these previous studies. It was expected that coefficients of variation for particular tasks, and correlations between target magnitude with both respect to both error rate and error size across tasks, would suggest use of analog magnitude estimation by these verbally impaired participants. Participants also completed three additional number tasks (number elicitation, confrontation naming with Arabic numerals, and a count list recitation task) and a subset of participants completed nonverbal semantic processing and short-term memory tasks (*Pyramids and Palm Trees* and a verbal semantic category probe) to better understand errors on nonverbal matching tasks. Results indicated that for people with aphasia, non-symbolic exact quantity representation was more difficult than for people without aphasia, except when target quantities were presented in subitizable groups. Overall, participants made more frequent and larger errors when representing larger quantities and struggled when the target was not visible. Participants who had difficulty with tasks where the target was visible during response also had difficulty with tasks where the target was not visible during response. However, another group of participants only had difficulty with tasks where the target was not visible during response. Additionally, participants who had difficulty with non-verbal aphasia assessment subtests were more likely to err on non-symbolic exact quantity representation tasks where the target *was* visible during response, while participants who had difficulty with aphasia assessment subtests that required verbal responses were more likely to err on non-symbolic exact quantity representation tasks where the target was *not* visible during response. These results, alongside correlations with aphasia assessment battery performance, suggest that (1) accuracy on non-symbolic exact quantity matching tasks where the target is visible on response rely more heavily on visuospatial abilities than on language or memory; (2) tasks involving subitizing small exact quantities do not appear to require the same visuospatial capacities; and (3) non-symbolic exact quantity matching tasks where the target is not visible on response rely upon language and memory abilities—especially the capacity for verbal counting. Taken together, these findings reinforce the notion that verbal counting facilitates the consistent and accurate recognition and representation of exact quantities larger than three or four by bridging innate human capacities for subitizing and analog magnitude estimation. Overall, the present results further inform our understanding of tasks previously used to understand the relationship between language and number in a culture lacking words for number concepts.

Keywords: linguistic relativity, number, aphasia

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction and Background

When considering the relationship between language and cognition, positions fall along a range: at one end of this spectrum, there is belief that language and cognition are independent domains, where language has no influence on how we perceive, act and reason; and, at the other end of the spectrum sits linguistic determinism, the idea that language determines what it is possible for humans to know and think. There is a broad middle ground between these two ends of the spectrum, and in that middle ground, language and cognition interact, and the grammatical structures and lexicons of a language influence thought to a greater or lesser degree. This is the territory of "linguistic relativity." While the idea that the language you speak has an influence on what you think about, perceive, and behave has a long history in Western philosophy, the clearest articulation of a strong version of this position was articulated in Whorf (1956):

> We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscope flux of impressions which has to be organized by our minds—and this means largely by the linguistic systems of our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way—an agreement that holds throughout our speech communities and is codified in the patterns of our language. (p. 213)

The rationale behind linguistic relativity (also known as the "Whorfian hypothesis") can be stated succinctly—working from the premise that language influences thought, the observation that languages differ in the way that they break up the experienced world into units of meaning leads one immediately to the conclusion that people who speak different languages will break up the world in different ways and perhaps even perceive the world differently (Gentner & Goldin-Meadow, 2003). The radical piece of this logic is in the direction of influence: for strong versions of linguistic relativity, words are not just names for pre-existing concepts; conceptual thought is influenced by the way a particular language is structured, and what they have words for and what they do not. The semantics and syntax of a language will influence the thought of speakers of that language, whereas speakers of another language may think differently as a result of their language having a different vocabulary and syntax. In other words, the world may be conceptually "cut up" in distinct ways for speakers of different languages. When a language is transmitted from one generation of speakers to the next, so are the idiosyncratic conceptual ways of "cutting up" the world that come with speaking it.

Many intuitive examples of this phenomenon exist. Everett (2013) cites the fact that Brazilian soccer players not only have more words for certain ways of dribbling the ball than American soccer players, but that they also exhibit a greater variety of dribbling behaviors as players and, as spectators, are attuned to the dribbling behaviors of players in a way that Americans are not. While this example does not neatly resolve the question of the direction of influence, it does highlight potential correlations between language difference, conceptual difference, behavioral difference, and perceptual difference—that is to say, it highlights the potential influence of language on non-linguistic thought. And this lights the way toward the difficulties and possibilities of investigating the reality of linguistic relativity: by focusing on differences in behavior in populations with different languages, it might be possible to identify in what ways and to what degree language influences those behaviors, so long as it is possible to identify that language is, in fact, influencing non-linguistic thought and action.

Despite the theoretical interest in the topic, until recently, experimental studies that demonstrate the influence of language on thought and behavior have been limited. Anthropologist John Lucy (1997) outlines three ways in which this research has been and might be conducted. Given the three terms involved in the hypothesis—*language* influences *thought* (as shown in behavior) about *the world*—each of Lucy's (1997) approaches begins with one term and proceeds toward the others. Structural-centered approaches work from *language* differences and attempt to isolate how these differences influence thought; domain-centered approaches begin with phenomenal aspects of *the world* and look at how different languages encode that phenomenon; and behavior-centered approaches start from human behaviors as expressions of *thought* and try to locate a linguistic source for that behavior.

Meanwhile, psychologists Gentner and Goldin-Meadow (2003) break up approaches to the research in linguistic relativity along more theoretical lines, identifying "language as lens," "language as tool-kit," and "language as category maker" as three ways that language could be seen to influence thought. A "language as lens" approach to the topic is focused on identifying how an individual's language influences their *perceptions* about the world, identifying ways in which language might shape what a person experiences as real. Although closely related, a "language as category maker" approach takes the influence of language on cognition to be essentially a question of dividing the world differently based on language, that language influences a person's *conceptions* about the world, rather than altering the nature of what is perceived. Finally, a "language as tool-kit" approach sees language as set of tools or technologies that enhances or dampens an individual's capacity to operate upon the world in certain ways.

While each approach has its advantages and disadvantages, all attempt to investigate the question of language's influence on thought, and thus fall under the heading of researching the linguistic relativity hypothesis. Everett (2013) compiles a diverse array of recent research in fundamental domains such as space, time, quantity, gender, and color that utilize different theoretical and methodological approaches but draw similarly positive conclusions about the reality of linguistic relativity. According to Frank, Fedorenko, Lai, Saxe, and Gibson (2012), recent studies have found "meaningful cognitive differences" (p. 75) between those who have words for particular concepts and those who don't, and that those cognitive differences exist both in people whose language doesn't have words for those particular concepts (i.e., across cultures) and when individuals haven't yet learned their language's words for those concepts (i.e., developmentally). At the same time, these differences can evaporate when you interfere with the ability of those who have language for these concepts to access the words for the concepts. In other words, language facilitates a different way of organizing the world.

One concrete example of this is in the arena of color discrimination. In Russian, for instance, the color spectrum includes a category distinction between "light blue" and "dark blue" that is not made in English. Winawer et al. (2007) asked English-speakers and Russian-speakers to discriminate between two colors of blue. Russian-speakers did so more quickly when one color fell into the category for "light blue" and the other "dark blue" than they did when both colors were in the same linguistic category. Moreover, the effect was stronger when the colors were perceptually close. This performance difference also disappeared under a verbal interference condition, but not a spatial interference condition. Meanwhile, English-speakers asked to perform the same task did not demonstrate any of these task performance differences. The study suggests that language categories affect color discrimination performance in clear and measurable ways.

The domain of number is a good testing ground for the “language as tool-kit” approach to the linguistic relativity hypothesis, in part because of how we understand human beings to cognitively represent numerical information, in part because of what we know about how numeracy develops alongside language in humans, and in part because there are clear differences between the ways that adult speakers of different languages perform number-related tasks. Dehaene et al. (2003) localizes three numeracy functions in three areas of the left parietal lobe. Specifically, in Dehaene’s model, the horizontal segment of the intraparietal sulcus is associated with number processing, the bilateral superior posterior parietal system is responsible for the “mental number line,” and the left angular gyrus area is involved in verbal number manipulations. Furthermore, we are born with two systems for the cognitive representation of number—a parallel-individuation system that can track up to three or four discrete objects and an analog magnitude estimation system we use to approximate large quantities (Feigenson, Dehaene, & Spelke, 2004). While these cognitive systems are also found in other animals, humans appear to use exact number words as tools that enhance our capacity to do things with quantities. We distinguish ourselves from other species through “our reliance on linguistic representations of quantity—number words—to act as symbolic placeholders in complicated operations” (Frank, Everett, Fedorenko, & Gibson, 2008, p. 819). With a solid grasp of the inherent neurobiology of numeracy—how and where we think of and understand numbers—and the understanding that language plays some role in our comprehension of reality—including our comprehension of number—we may begin to focus our attempt to shed greater light on the relationships between number, language, thought and the world by examining previous research on these topics. Reviewing developmental research on numeracy will allow for better differentiation between those number skills humans are born with from those we develop, while examining cross-cultural studies of number language and exact quantity will illuminate which aspects of numeracy are common across cultures and languages and which are the product of the particular cultures and languages in which we develop. This opens the way to asking what happens to human abilities to recognize and represent exact quantity when language for exact quantity is interfered with or impaired.

Literature Review: Recent Studies in the Relationship Between Number Language and Concepts of Exact Quantity

This literature review examines recent studies which have directly and indirectly addressed aspects of the question of how language affects exact quantity recognition and representation in human beings. This review begins with a study that connects animal research in this area to human capacities for quantity recognition and representation. From there, the review turns to developmental literature, examining children who are learning the language of number and are also learning to apply that language to groups of objects identified with that language. These studies should help to clarify the capacities humans are born with, and those that come into being as a child grows and learns. The next group of studies in the review focus on an indigenous Amazonian culture that has no language for exact number and who are challenged by exact quantity representation. This body of work raises questions about whether the emergent number capacities found in the developmental literature are the product of maturation, and would therefore develop in any typical human, or whether these capacities for exact quantity recognition and representation are a result of learning that is specific to the language and culture the individual is born into. A final study in this review asked English-speaking college students to perform the same tasks as the above-mentioned indigenous culture when the college students' access to number language has been artificially disrupted. Here, the relevant question is whether interruptions to the students' access to language for exact quantity impacts their ability to recognize and represent exact quantity. The commonalities between populations in these studies help to highlight what is essentially human about how we all experience exact quantity, while the variances between populations can illuminate the role that development, culture, and language play in particular groups of people's experiences of understanding and using exact quantities.

Whalen et al. (1999) is often cited in the literature on number language and cognition as a landmark study. The experiments described in the paper attempted to evaluate the hypothesis that humans estimate large quantities using the same kind of cognitive mechanism as animals when they are unable to use language to count that quantity exactly. When animals are tasked to produce a particular number of bar presses in order to receive food, their responses are increasingly variable in proportion to the magnitude of the target number of bar presses. In other words, animals demonstrate a constant coefficient of variation (CoV) across magnitudes in these tasks. A CoV is equal to the standard deviation of a data set divided by its mean, effectively converting the spread of the data into a percentage and giving the reader a sense of the *relative* size of the data's variability when compared to other sets. As used in animal studies of magnitude estimation, a constant CoV across targets means that there is proportionally less variability in an animal's response when the target number of bar presses is five as when the target number of bar presses is nine. The range of responses an animal makes to a target of nine is larger than the range of responses made for a target of five (i.e., the standard deviation is larger for a target of nine than for five), but that range is always proportional to the size of the target—and it is always about 0.15.

Whalen et al. (1999) hypothesized that if they limited human participants' access to verbal counting of quantities by asking them to respond more quickly than they could count, participants would fall back on estimations of mental magnitude, and produce the same CoV seen in animals. To test this, they performed two experiments with seven undergraduate and graduate college students. In the first experiment, each participant was shown a number and then asked to press a button that many times as quickly as possible without counting. In the second experiment, the same participants were shown a flashing dot and asked to say how many times it flashed, again without counting. In both experiments, the participants were shown every odd number between seven and twenty-five and each participant performed forty trials of each experiment. In both experiments, the authors found a constant CoV across trials and participants, suggesting that without time to count, humans resort to estimating large quantities and that these estimations are noisy in proportion to the size of the number being estimated.

Although Whalen et al. (1999) drew its conclusions from a sample of seven American college students, raising questions about the size and representativeness of the sample, the authors accumulated 5,600 total data points by having each participant complete a large number of trials of each experiment. While the large data set may offset the potential sample size problem, it does not address whether the participants themselves are a good representation of the population. Although there is a tradition within psychophysics research of using many trials from a small sample of participants, such an approach might simply reinforce the idiosyncrasies of those individuals' responses, or highlight, in this case, the facility of American college students with the task, rather than suggesting a general human trait. Furthermore, such an approach invites participant acclamation—that is, it is possible the participants got better at the tasks with repeated exposure. The authors were careful in examining this and other possible explanations for the participants' performance and found that there was no evidence of participant acclamation, nor any evidence that the participants were secretly counting or using duration of time to estimate their responses

A better approach, one that would allow for more substantive conclusions about the role of language in the way humans think about and use number, involves studies employing participants who have no language for number, or studies which artificially induce conditions where participants are unable to access language for number. The members of these groups of study participants would share an incapacity to employ number language, albeit for different reasons, but this would allow a better sense of what humans are capable of with regard to quantity in the absence of language for number. The latter is difficult because it is difficult to say for sure that participants are not accessing number language despite efforts to the contrary, as in Whalen et al. (1999)—asking people not to count doesn't mean they won't, even if they are making a good faith effort—and as we will see in Frank et al. (2012)—it may be possible to count while simultaneously listening to and reciting a news broadcast. The former—studies with participants who have no language for number—is more easily possible by working with children who have not yet mastered number language.

Wynn (1992) marks the beginning of research into what pre-linguistic children know about **exact quantity**, suggesting that five-month-old children are able to distinguish between 1, 2, and 3. Thirty-two infants were shown a doll inside a display case, and then a screen was used to block the infant's view of the doll. Next, a researcher added another doll behind the screen in the case, but did so in plain view of the infant. Finally, the screen was removed to reveal the number of dolls behind the screen. Thus, under the control condition (called the "possible outcome"), the infant saw one doll before the screen goes up, sees another doll added while the screen is up, and sees two dolls when the screen is removed. However, under the experimental condition (or the "impossible condition"), the researcher also secretly removed the first doll when she added the second. Thus, when the screen was lifted, only one doll was visible, even though one appeared to have been added and none appeared to be taken away. The infants stared significantly longer at the impossible outcome, suggesting they expected there to be more dolls in the end state of the experiment when it appeared to them that more dolls had been added. Wynn (1992) also performed the same experiment in reverse—appearing to subtract one doll from the case—and got similar results. Being surprised (and therefore staring longer) at changes that violate our assumptions about addition and subtraction suggests that prelinguistic infants can distinguish between small quantities, lending support to the hypothesis of an inborn capacity for recognizing these differences. These results have been successfully replicated in the twenty-five years since Wynn (1992), and other research has found this capacity in infants as young as two days old (Izard, Sann, Spelke, & Streri, 2009)[1].

Meanwhile, Xu and Spelke (2000) provided evidence for an **approximate number sense** in prelinguistic infants, suggesting an innate human capacity to grossly distinguish between large quantities. In two experiments, sixteen six-month-olds were habituated to arrays of either eight or sixteen black dots of various sizes, configurations, and brightness. Then the infants were shown a novel array: sixteen dots if they had been habituated to eight dots, and eight dots if they had been habituated to sixteen. The infants stared longer at the new numbers, suggesting they could differentiate between eight and sixteen dots. However, the second experiment in Xu and Spelke (2000) showed the limitations of this ability to distinguish large quantities. The same infants who could differentiate between eight and sixteen did not differentiate between eight and twelve dots. We are born able to distinguish between sets of objects that are identical in every way but quantity, so long as the disparity between the two sets is of a ratio of at least 2:1.

Consequently, there appears to be good evidence that we are born with systems for distinguishing between small exact quantities and for differentiating between large, sufficiently different quantities. The question remains: how do human infants go from these innate capacities

[1] Izard, Sann, Spelke, & Streri (2009) also demonstrated that newborns have the capacity to differentiate between small exact quantities across sense modalities.

to being able to count—how do we go from being able to recognize small exact quantities and approximating large quantities to being able to differentiate between large exact quantities?

Condry and Spelke (2008) performed six quasi-experiments with sixteen to forty three-year-olds whose ability to recite the number list outpaced their comprehension of the meanings of particular numeracies. Such a child might be able to recite counting numbers (a.k.a. "natural numbers") in order up to ten but not correctly identify or create a set of seven objects. Taking advantage of the opportunity that this developmental moment presents, the authors attempted to provide evidence for one of two accounts of how children develop an understanding of number words and number concepts. In one account, children build from the concepts of a "numerically distinct individual" and a "set" to see that a set is a group of individuals—or, put another way, they understand singular and plural, and then they learn, through a bridge made of language, that a plural is some group of individuals. In the other account, children already understand these features of natural numbers, but they cannot name precise sets accurately because they have not gotten fine enough control over their capacity for analog magnitude estimation.

The results of Condry and Spelke (2008) supported the first of the two accounts of how children develop their understanding of number words and concepts. Their experiments provided evidence that an understanding of natural number does not precede the acquisition of language for number words. The rules of numeracy appear to develop alongside language, and the authors hypothesize that an understanding of natural number emerges as a consequence of trying to understand number words and learning to use counting to represent number. The authors conclude that without a language-independent representation of cardinality, children cannot learn the relevant difference between even small sets labeled "one," "two," and "three."

Although the development of number concepts and language is certainly of interest, it does not directly address the question of how our language for counting impacts our concepts of exact quantity. Furthermore, it is not always appropriate to draw a causal explanation of a skill from its development. While Condry and Spelke (2008) might show us how children come to understand certain concepts, there is no guarantee that their understanding does not evolve past that point in adulthood. This brings us to the second way to explore the relationship between language, number, and cognition—an adult population that lacks language for number concepts could show us a great deal about how number language affects number concept.

An Amazonian hunter-gatherer tribe called the Pirahã is of great interest to linguists and cognitive scientists alike, and for mostly the same reason: their language defies many of the concepts and rules that we had come to see as fundamental to thought and language. Relevant to the current study, what is interesting about the Pirahã language is that it has no words for number concepts. While other indigenous cultures sometimes have limited language for words or "impoverished" number concepts, the Pirahã are perhaps alone in having no words for number. The next three studies discussed all use the Pirahã as participants; all attempt to shed light on the relationship between number and language; and each subsequent study attempts to replicate the work that came before it.

In a pilot study with only seven participants and where data was primarily taken from four of those seven, Gordon (2004) engaged Pirahã tribe members in a series of matching tasks, where the participant was asked to create a visual array that matched the author's array. Surprisingly, the Pirahã struggled to accurately reproduce any set of objects consisting of more than three items, even when the model was still visible for the participant to copy. Furthermore, Gordon (2004) noted the Pirahã produced a consistent CoV of 0.15, matching the findings of

Whalen et al. (1999). The distribution of the Pirahã errors suggest the use of analog magnitude estimation—supporting the idea that in the absence of language for number and the ability to count exactly, people fall back onto slower, less accurate, inborn abilities to estimate quantities beyond three. Without language for discrete numbers, people can only guess at quantity.

Such surprising results deserve to be re-tested, especially with a potentially problematic sample size, methodology, and problems with translation between radically different languages and cultures. Frank et al. (2008) attempted to replicate the results from Gordon (2004) in a different Pirahã village and arrived at conflicting results in two experiments. The first experiment asked six members of the Pirahã tribe to describe varying numbers of objects to test whether the words Gordon (2004) claimed were number words actually were words for specific quantities. The authors presented participants with spools of thread, adding one at a time. For each quantity the participant was asked, "How many?" Unlike Gordon (2004), Frank et al. (2008) found that each of the three words that may have represented quantity was used to designate a range of quantities, suggesting they are relative, comparative terms, rather than representing numbers. From the results of the first experiment, the authors concluded the Pirahã have no method of expressing exact quantity. In their second experiment, Frank et al. (2008) attempted to recreate the experiments of Gordon (2004), and did so with fourteen participants. The authors got the same results for every task with the exception of the initial one-to-one matching task, where participants are asked to create a visual array that matched an experimenter's array with the model still visible for the participant to copy. Here, results were near ceiling for accuracy, suggesting that analog magnitude estimation was not used, as in Gordon (2004). These outcomes prompted Frank et al. (2008) to conclude that the startling results of Gordon (2004) were the result of either the participants not understanding the task or inconsistencies in the experiment.

Because the Pirahã can do the one-to-one matching task without words for number, Frank et al. (2008) argued that Pirahã performance on matching tasks only requires knowing that a set of items contains exactly some number of individuals, not necessarily the name of that number. Consequently, the authors argued for number language as a cognitive technology, rather than language for number being necessary to accurately recognize numbers beyond three. In other words, exact quantity is not created by language, but number language allows the user to transfer information about sets across time, space, and modality. For Frank et al. (2008), the lack of such a technology explains the difficulty the Pirahã demonstrated in the matching tasks that required either remembering an exact quantity or transposing it in space, but not the one-to-one matching task, where the target set was visible throughout the trial.

Everett and Madora (2012) sought to resolve the conflict between Gordon (2004) and Frank et al. (2008) with regard to the Pirahã performance on one-to-one matching tasks and an attempt to explain the anomalous results. The authors recreated the three tasks from Frank et al. (2008) with fourteen participants in a third Pirahã village, testing each participant on four random numbers between four to ten, garnering a total of fifty-six trials per task. Because the authors' primary interest was in the one-to-one matching task, the other tasks were used as controls (these consisted of a task where the array to be matched was hidden after presentation and a task where the array to be matched was presented perpendicularly to the participant). Everett and Madora (2012) found no significant differences when comparing any of the tasks to each other or when comparing tasks across the three studies with the exception of the Frank et al. (2008) one-to-one matching task, which was found to be significantly different from each control task and the Everett and Madora (2012) one-to-one matching task ($p < .001$ in each case). In a similar vein, the CoV for all tasks in Everett and Madora (2012) were 0.15 across targets, consistent with Gordon (2004) and Whalen et al. (1999) and suggested the Pirahã were employing analog estimation strategies for quantities greater than three—in other words, they were roughly estimating, not counting.

Of all the research of this kind conducted with the Pirahã, only the one-to-one matching task in Frank et al. (2008) produced a CoV markedly different from 0.15. Everett and Madora (2012) offers a speculative explanation: unlike the others, the village tested in Frank et al. (2008) had been exposed to math tutoring that included neologisms for number words. It is the neologisms for number words that stand out—all the villages had been exposed to the one-to-one matching task and other attempts at basic math training on the part of the Brazilian government, but only the village of Xagiopai (the site of Frank et al. (2008)) had been exposed to number word neologisms. The authors are clear that this is speculation on their part, and their caution seems warranted here, as it is anywhere research with the Pirahã is concerned. Practical concerns limit the sample size possible when testing this group—they are few in number and live in a remote location. Additionally, the Pirahã language and their life experiences are radically different from the language and life experiences of the researchers, so much so that any conclusions must be made tentatively and cautiously.

Coming full circle, Frank et al. (2012) brings together the experimental tasks performed with the Pirahã and the idea that verbal interference forces adults with number language to resort to analog magnitude estimation as in Whalen et al. (1999). The authors hypothesized that if language is not crucial to establishing exact number, then adults with language for number should successfully perform non-verbal number tasks under verbal interference (in this case, reciting a news broadcast to which they are simultaneously listening while performing the experimental tasks). On the other hand, should language be necessary for exact numeracy, these same English-speaking adults should be forced to fall back on analog magnitude estimation under verbal interference. Should the latter be the case, analysis of the data would reveal a constant CoV, as seen in other studies. To test this, Frank et al. (2012) had thirty-five students from the Massachusetts Institute of Technology attempt the matching tasks performed with the Pirahã while repeating a radio news broadcast aloud. Each participant performed nine trials of each of the five tasks; a total of forty-five trials per participant. The results were then compared to each other and to the results of the same experiments with the Pirahã from Frank et al (2008).

While the English-speakers were found to be more accurate than the Pirahã, both groups made "significant and systematic errors" (p. 79) on the "nuts-in-a-can" task, wherein the experimenter drops objects into a can one at a time and then asks the participant to present an array consisting of the same number of objects. In this task, where participants can neither see nor visually remember the number they are meant to match, the college students, like the Pirahã, produced a flat CoV of 0.15 across targets, suggesting the use of analog magnitude estimation. Based on these results—and apparently not aware of the findings of Everett and Madora (2012)—Frank et al. (2012) drew the conclusion that the concept of "exact match" does not require language, but language is crucial to storing and manipulating exact quantities greater than three. This conclusion is in line with the language as a technology or tool hypothesis, which argues that while human beings have pre-linguistic numerical capacities, language for exact quantities allows us to transcend these abilities, putting new possibilities in our grasp.

To answer the question that frames this review: the evidence to date strongly suggests that language for number has a significant influence on how quickly and accurately we comprehend and process quantities larger than three. At the same time, there is still much room for debate as to how fundamental number language is to the correct apprehension of exact quantity—the logical next frontier in this literature. The evidence from Everett and Madora (2012) supports the idea that language is a necessary precondition for the ability to exactly represent quantity and as such conflicts with the "language as technology" hypothesis supported by Frank et al. (2012). Each study has limitations—the representativeness of each sample is questionable and there are potential problems around translation on the one hand and the efficacy of verbal interference on the other—and so while they point the way to the next steps to be taken in this area of research, these studies also represent the major gaps in this literature. Studies which rely on verbal interference always leave open the possibility that the participants' ability to access number language has not been effectively or similarly disabled across studies, and studies which reach across languages and cultures are always in danger of being confounded by cultural and linguistic differences, especially in research of this kind. It is possible to fill in some of the questions left in the wake previous research by studying English-speaking people with aphasia, who have had access to number language throughout their lives and live in a culture steeped in symbolic representations of number, but who also have had their access to or capacity to employ number language impaired.

The Current Study

One largely unexplored route to understanding the relationships among language, number, and cognition that accounts for the abovementioned problems is through the population of people with aphasia. People with focal brain lesions—either as a result of infarcts or tumor resections—or other restricted lesions may have aphasia. Aphasia is an acquired impairment of a person's ability to comprehend and formulate language across multiple modalities, including speaking, reading, writing, and listening (Rosenbek, LaPointe, & Wertz, 1989). Consequently, people with aphasia may experience difficulty in the use of language for number and calculation (Dragoy et al, 2016). Meanwhile, McNeil and Pratt (2001) specify that aphasia is a processing or linguistic performance disorder, rather than an impairment of linguistic competence. In other words, a person with aphasia is thought to have difficulty accessing or employing the language for a cognitive concept, not difficulty with the concept itself. By this logic, if aphasia were to affect a person's ability to non-symbolically represent exact quantity, it would be in a similar fashion to the impact of verbal interference—by disrupting the person's ability to access language for number. However, it also remains possible that aphasia may impair an individual's ability to represent exact quantity in a manner more like the case of the Pirahã—where a person with aphasia's access to the concept of exact quantity itself may be impaired, rather than their having an impaired ability to deploy that language tool.

Certainly, there are differences between the population of people with aphasia, people of an anumeric culture, and English-speakers under verbal interference. What separates the Pirahã from the other populations under discussion here is that they exist in a world without exact quantity language and therefore may not have a concept of number to access. English speakers under verbal interference, meanwhile, are part of a numeric culture who have had their ability to use that concept temporarily disrupted, and people who have aphasia due to stroke are members of a numeric culture and have had, at least until the onset of their aphasia (in some cases), access to language for exact number for most of their lives. Both the English speakers and people with aphasia are surrounded by symbolic representations of number in their culture in a way that the Pirahã are not. Furthermore, it is difficult to impossible to control for the heterogeneity of an aphasia population in terms of lesion locations and correlations between those lesions and types of language impairments. This is unlike a population of English speakers under verbal interference, where participants are from the same cultural background and manipulated via the same method and respond predictably, and unlike the Pirahã, who are uniformly without access to words for exact quantity because such terms are not in their language.

While a few case studies have examined the impact of aphasia on calculation (e.g., Dragoy et al., 2016), only one study to date has examined the impact of language impairment on non-verbal and non-symbolic representation of quantity. Lemer, Dehaene, Spelke, and Cohen (2003) examined a person with acalculia due to a focal lesion of the left parietal lobe and another person with semantic dementia from predominantly left temporal hypometabolism to demonstrate dissociations between tasks associated with counting and tasks associated with innate quantity systems of number processing. As predicted by a deficit in innate quantity processing due to a lesion in the parietal lobe, the first person showed a severe slowness in approximation, and exhibited impairments in subitizing and numerical comparison tasks. Meanwhile, the second person had intact approximation abilities and showed preserved processing of non-symbolic small numbers—in other words, her "quantity processing" systems were functioning as expected—but she struggled with tasks that required intact verbal processing and counting. Given this previous finding and related findings with other populations, the presence of language impairment in the form of aphasia should negatively affect the individual's ability to produce non-verbal representations of quantity.

Consequently, in this study we recruited a group of sixteen participants with aphasia to perform the non-verbal numerical tasks of Everett and Madora (2012) and Frank et al. (2012) to address the question: *What is the effect of language impairment on an individual's ability to produce non-linguistic representations of quantity?*

In this research, participants with aphasia performed the same set of five, increasingly complex matching tasks used with both the Pirahã and English-speakers whose access to language was artificially compromised by verbal interference. I hypothesized that participants would make more frequent and larger errors (1) in proportion to target size; (2) on each subsequent, more difficult, task; and (3) produce a flat coefficient of variation (CoV) on each task and across target quantities, suggesting reliance on the analog magnitude system to estimate quantity. Such results would lend further evidential support to the hypothesis that access to language for exact number is necessary for the recognition and representation of exact quantities.

If language impairment in the form of aphasia limits a person's ability to represent non-symbolic exact quantities, that would add another human population to the list of those in which a relationship has been found to exist between exact number language and the ability to perform non-symbolic exact quantity tasks. When considered with similar or complementary evidence from previous studies with different populations—children raised in numerate cultures who have yet to develop number-language skills, adults in numerate cultures under verbal interference, and adults in an anumeric culture—it seems clear that access to exact number language influences how humans think about numbers. More broadly, these findings would support the linguistic relativity hypothesis that language influences nonlinguistic thought and behavior.

Methodology

Research Design

This dissertation is based on a sample of 16 participants with clinical diagnoses of aphasia. All sixteen participants completed aphasia assessments, the set of five non-verbal and non-symbolic exact quantity representation tasks from Everett and Madora (2012) and Frank et al. (2012), as well as additional number and counting tasks. A subset of eight participants completed tests of nonverbal semantic processing and short-term memory. Descriptive quantitative methods were used to analyze performance between and among these assessments and tasks, and then compared to data from the Pirahã (Everett & Madora, 2012; Frank et al., 2008; Gordon, 2004) and college students under verbal interference (Frank et al., 2012) performing identical tasks. Qualitative, case study approaches were used to capitalize on the diversity of the participant population by highlighting response patterns common to subsets of participants.

IRB Approval

This project first received IRB approval in February 2015 and is currently approved through January 2019.

Recruitment of Participants

Research team members included a clinical psychology graduate student (the author), a cognitive neuroscientist (Dr. Alex Kranjec), and a certified speech-language pathologist with expertise working with people with aphasia (Dr. Sarah Wallace). Participants were recruited through personal networks of research team members (e.g., clinical supervisors in the Duquesne

University Speech-Language Clinic) and through the distribution of fliers. Personal networks and employees at facilities that provide speech-language therapy were provided recruitment fliers to distribute to potential participants and their families. Family members were encouraged to explain details of the flier to potential participants who have difficulties reading.

A second method of recruitment included participants known to Dr. Wallace because they had given written permission to be included in the Wallace Participant Pool. These participants expressed interest in receiving information about upcoming research studies. The information in the Wallace Participant Pool was maintained using the same procedures as all other study information. All information was listed by participant code and the codes linking participants to the information were kept in a separate secure location. Information secured for the Wallace Participant Pool included phone numbers, diagnosis of aphasia, and in some cases email addresses or work phone numbers for caregivers or spouses. These participants were contacted via mail, email, or by phone.

Ethics, Consent and Confidentiality

All processes for obtaining consent and all documentation of consent in this study follow procedures and guidelines for informed consent in research with aphasia participants outlined in Kagan and Kimelman (1995) and are the standard research practice of the Duquesne University Speech-Language-Hearing Clinic.

Informed consent procedures. Consent was obtained either at the time of the preliminary session or prior to the testing session at a time, day, and location that was convenient for the potential research participant, the participant's legal guardian or adult family member, and the primary researcher. Potential participants and their legal guardians or adult family members were given the choice to complete consent procedures at the participant's home or at Duquesne University. The research study was explained to prospective participants through face-to-face interactions with a member of the research team. Decisions about participation in this study in no way affects the services that potential research participants received.

People with chronic aphasia are vulnerable research participants due to limitations in their capacity to understand written or oral information presented to them. However, many of these individuals have retained personal legal guardianship and must provide consent to participate in a research project. A person's status regarding the provision of consent was determined by questioning both the person with aphasia and family members of that person about legal guardianship. Specifically, the investigator asked the person with aphasia and their family members to name the legal guardian if the person with aphasia no longer maintains personal legal guardianship. The appropriate consent, permission, and assent forms were then presented for completion. When prospective participants had not retained personal legal guardianship, permission was obtained from the person's legal guardian and assent obtained from the person with aphasia.

In all cases, the research methods and procedures were described by the researcher to potential participants with aphasia and their family member(s) or legal guardian(s). Due to the nature of aphasia, additional time, written input, and drawing was provided as the researcher deemed it necessary to ensure comprehension of the information presented. Specifically, the researcher confirmed comprehension through the person with aphasia's use of facial expressions, gestures, writing, drawing or verbal expression. In addition, prospective participants and their family member(s)/legal guardian(s) were presented with written informed consent and assent forms. After listening to a verbal description of the project methods and procedures and reviewing the written consent and assent forms, participants and their family member(s)/legal guardians(s) were asked whether they understood the forms and whether they wished to discuss the decision to participate with their families. When they responded that they understood the forms and had no questions, participants and their family member(s)/legal guardian(s) were asked to give their assent, permission, or consent as appropriate. Participants were provided a copy of the consent form to retain for their records.

Confidentiality. The confidentiality of participants was maintained in the following ways: (a) participants were each assigned a participant code by the researcher and all data and documentation were identified with this code alone; (b) the forms, transcripts, and data associated with this project were stored in a locked file cabinet in the office or lab of the co-investigators for a five year period and then will be destroyed; (c) when study results are reported at scientific meetings or published in scientific reports, the identity of the participants will not be made public.

Target Population and Sample

Participants in this study included sixteen people with aphasia resulting from left hemisphere stroke (Table 1). Participants consisted of 13 men and 3 women, all of whom were right-handed prior to their stroke. Twelve participants used their left hand to complete the tasks of the study. All participants were adults between the ages of 43 and 75 years old at the time of testing and spoke English as their primary language. The mean and median participant age was 60.5 years old. All participants had completed between 12 and 20 years of education at the time of testing, with an average of 15.44 years of education and a median of 16 years of education. Participants were a minimum of nine months and maximum of 159 months post onset of aphasia, with a mean of 73.31 months post onset and a median of 63 months post onset. Irrespective of severity of disorder, all individuals diagnosed with aphasia—as determined by their score on the aphasia quotient portion of the Western Aphasia Battery-Revised (WAB-R) (Kertesz, 2006) and a speech-language pathologist—were included in the study. Participants had WAB-R Aphasia Quotient (AQ) scores that ranged from 15.30 to 97.40, with an average AQ score of 57.20 and a median of 52.65. Broken up by aphasia type, five participants are diagnosed as having Broca's type aphasia, six with Anomic type, two with Wernicke's type, two with Conduction type, and one with Global type aphasia. Participants were required to have adequate vision and hearing abilities to complete the study (see below for description of hearing and vision screening).

Table 1

Demographics of participant population.

Participant	Age	Gender	Education	Handedness	Months Post Stroke	WAB-R AQ	Apahasia Type
1	43	M	16	L	145	65.50	Broca's
2	58	M	16	L	59	97.40	Anomic
3	53	M	18	L	62	24.40	Broca's
4	74	M	12	L	64	31.20	Wernicke's
5	75	M	20	L	30	55.90	Conduction
6	51	M	16	R	119	49.40	Conduction
7	56	F	18	L	92	96.80	Anomic
8	68	M	16	L	59	46.00	Wernicke's
9	45	M	15	L	55	27.40	Broca's
10	68	M	12	L	159	83.80	Anomic
11	70	M	12	R	9	32.10	Broca's
12	59	M	14	R	14	96.50	Anomic
13	72	F	16	R	96	88.50	Anomic
14	62	F	16	L	39	15.30	Broca's
15	47	M	14	L	96	18.30	Global
16	67	M	16	L	75	86.70	Anomic
mean	60.50		15.44		73.31	57.20	
median	60.50		16.00		63.00	52.65	
std dev	10.57		2.25		42.91	30.64	
		13 M / 3 F		4 R / 12 L			

Notes. Age, Education, and Months Post Stroke are at time of testing. Handedness refers to which hand the participant used to perform the study's tasks. Aphasia type is determined by WAB-R scores.

Process Overview

Participants completed a minimum of two sessions: a "preliminary session" consisting of screening procedures and aphasia testing and a "testing session" in which participants completed the study's number-related tasks. Time for an optional third session was reserved if not all test materials were completed or if the researcher and participant agreed to a "debriefing interview." Each session lasted up to two hours. Participants were offered an optional ten minute break for every thirty minutes of participation in the sessions described below. Breaks were given more often as needed. All sessions occurred either at the Duquesne University Speech-Language-Hearing Clinic or the participant's home, and were conducted by the author.

Preliminary session. The preliminary session included testing and screening procedures. All participants completed a prescreening and medical history interview with the researcher prior to beginning the study. During the preliminary session, all participants were screened for normal, or corrected to normal, vision and hearing as a requirement of participation in the study. Participants completed the aphasia quotient portion of the WAB-R (Kertesz, 2006) and the Cognitive Linguistic Quick Test (CLQT) (Helm-Estabrooks, 2001). These assessments are part of a standard procedure followed at the Duquesne University Speech-Language-Hearing Clinic. Depending on the time of participant intake and consent into the present study, the data for the preliminary session was either accessed from records held at the clinic or administered at entry.

Testing session. All participants completed the set of five non-verbal and non-symbolic exact quantity representation tasks from Everett and Madora (2012) and Frank et al. (2012) in the following order: a one-to-one matching task (Task 1), an uneven matching task (Task 2), an orthogonal matching task (Task 3), a hidden matching task (Task 4), and a "nuts in-a-can" task (Task 5) (Figure 1, below). In every task, the experimenter presented a quantity of spools of thread (approximately 1” tall, ¾” in diameter) and asked the participant to construct a row of un-inflated balloons (approximately 4” long and 2” wide) that matched the number of spools of thread. Participants were tested once per task on each quantity from four to twelve in one of two random orders, for a total of forty-five trials per participant.

In addition, participants completed a number elicitation task, a confrontation naming task, and a counting task. Eight participants also completed the Semantic Category Probe Test (Freedman & Martin, 2001) and the Pyramids and Palm Trees Test (Howard & Patterson, 1992).

Debriefing interview. The optional debriefing consisted of discussing the results from the assessment as well as a full description of the study and answering any questions and/or discussion regarding the participant’s feelings about their experience taking the assessment. During this interview, the researcher asked participants to describe strategies they utilized during the testing session. If participants were unable to give a descriptive account of their process during testing, the researcher asked a series of Yes/No questions to elicit an account.

Procedures

Prescreening and medical history interview. Participants completed a prescreening and medical history interview with the researcher prior to the study.

Hearing screening. The hearing screening involved engagement in an informal conversation about activities the person has recently participated in or plans to do in the near future. The sole purpose of this screening was to ensure that participants had adequate hearing to engage in conversation.

Vision screening. An E chart was used for vision screening. The E chart is an ophthalmological chart that uses the same principle as the traditional Snellen distant vision chart to determine visual acuity. It is useful for people with aphasia because it does not require picture or letter naming but merely recognition of the orientation of an "E" when presented in various positions of 90° rotation (up, down, left, right). Participants were asked to point in the correct direction of the "E" as it is presented.

Number elicitation task. Participants were asked to name the number of spools of thread presented, increasing in order from one to twelve and then in decreasing in order from twelve to one. In each case, the researcher asked the participant, "How many spools of thread are there?" Participants were given up to thirty seconds to respond. The researcher recorded whether each target was named correctly, and if not, what response was given, if any. Divergence between performance on this task and on the matching tasks might illuminate whether the participant is having difficulty with recognizing, articulating, or representing the target quantity.

Confrontation naming task. Participants were asked to name the Arabic numerals one through twenty as presented individually on flashcards, in a random order. Numerals were 3" tall, and typed in black ink on 5x8" white cards. In each case, the researcher asked the participant, "What number is this?" Participants were given up to thirty seconds to respond. The researcher recorded whether each target was named correctly, and if not, what response was given, if any. This task aimed to assess the participant's ability to recognize and name Arabic numerals. Performance on this task compared to performance on the matching and counting tasks might help differentiate the participant's ability to recognize and name symbolic and non-symbolic numbers.

Count list recitation task. Participants were asked to recite the count list from one up to twenty and then from twenty down to one. The researcher said, "Please count from one up to twenty" and "Please count from twenty down to one." Participants were allowed five minutes to list the numbers. The researcher recorded whether each number was named correctly and in the correct order, and if not, what response was given, if any. Skipped numbers (e.g., "1…2…3…5…6") were counted as one incorrect response. Performance on this task indicates the participant's capacity to access and articulate counting numbers in order, a factor in the participant's matching task performance.

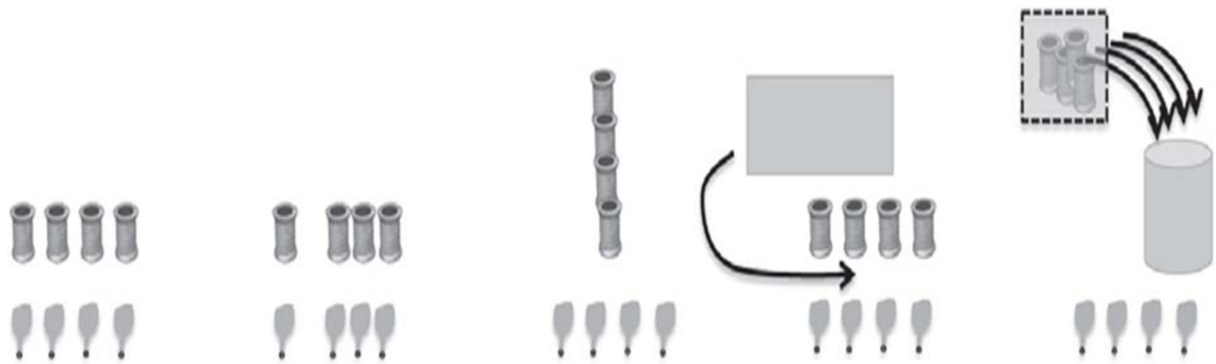

Figure 1: Schematic of each of the matching tasks. From left to right: (1) one-to-one match (Task 1), (2) uneven match (Task 2), (3) orthogonal match (Task 3), (4) hidden match (Task 4), (5) "nuts-in-a-can" (Task 5).

Matching tasks (Everett & Madora, 2012; Frank et al., 2012). Participants completed five object matching tasks: a one-to-one matching task (Task 1), an uneven matching task (Task 2), an orthogonal matching task (Task 3), a hidden matching task (Task 4), and a "nuts in-a-can" task (Task 5) (Figure 1). In every task, the experimenter presented a quantity of spools of thread (approximately 1" tall, ¾" in diameter) and asked the participant to construct a row of un-inflated balloons (approximately 4" long and 2" wide) matching the number of spools of thread. Participants were allowed as much time as needed to create a response array. For each trial, the researcher recorded whether the target array was correctly represented, and if not, what number of un-inflated balloons was presented. Participants were tested once per task on each quantity from four to twelve in one of two random orders; a total of forty-five trials per participant.

In Task 1, the experimenter placed the spools one at a time in an evenly spaced line from left to right. In Task 2, the spools were presented in the same manner as in the one-to-one task, but broken randomly into smaller groups of one to four. Task 3 was identical to the one-to-one task except that the row of spools was presented in a line perpendicular to the participant. Task 4 was identical to Task 1 except that the row of spools was subsequently hidden from the participant after being presented. In Task 5, the "nuts-in-a-can" task, the experimenter placed spools one by one into an opaque cup.

Instruments

Language and cognition assessments. All participants completed the aphasia quotient portion of the WAB-R (Kertesz, 2006) and the CLQT (Helm-Estabrooks, 2001). These formal assessments include tasks such as answering questions, describing pictures, manipulating and naming common objects, following directions, repeating words, and matching pictures to printed words and sentences.

Pyramids and Palm Trees Test (Howard & Patterson, 1992). Eight participants completed this test of non-spoken semantic processing. In it, participants match a pictured item to the closest associate among a set of two pictured choices (e.g., fish matched to: cat, table). This task assessed the participant's capacity to process non-verbal semantic information. Distinguishing between semantic and verbal impairments may help explain performance on the matching tasks.

Semantic Category Probe Test (Freedman & Martin, 2001). Eight participants completed this test of semantic short-term memory. Participants listened to a list of three or more words and determined whether the final word was from the same category as any of the preceding words by saying or pointing to “Yes” or “No.” This task aimed to assess the participant’s capacity to retain semantic information in their short-term memory. Short-term memory impairment might interfere with performance on the matching tasks.

Collection of Data and Method of Data Analysis

Data from the assessments conducted in the preliminary and testing sessions were compiled. Frequencies and descriptive statistical analysis were completed across all data.

Performance on the five matching tasks was measured by participant accuracy on each trial. For each trial, the researcher recorded whether the target array was correctly represented, and if it was not, what number of un-inflated balloons was presented. For the number elicitation task, the confrontation naming task, and the count list recitation task, performance was measured by participant accuracy on each trial. In the number elicitation and confrontation naming tasks, the researcher recorded whether each target was accurately named, and if not, what response was given, if any. In the count list recitation task, the researcher recorded whether each number was named correctly and in the correct order, and if not, what response was given, if any. In the count list recitation task, a skipped number (e.g., “1…2…4…5”) counted as one incorrect response.

A database of trial results across participant, aphasia sub-type, task, and target response was compiled. The raw accuracy results across those four domains, both in aggregate and across domains, were compared. Additionally, the CoV across task, participant, and target response were calculated and the results were compared across domains, within and between aphasia sub-types. Finally, results were compared to the results from similar studies with different populations—the Pirahã and college students under verbal interference.

Results

Overview

This chapter will describe (1) participant performance of the tasks of this study, including a comparison of the results of this study to previous studies using the same tasks with different populations, (2) correlations between participant performance within and across the procedures and instruments of the study, and (3) connections and dissociations found in participant performance among categories of tasks. The tasks of the study can be divided into three broad categories—matching tasks, aphasia assessments, and counting tasks—and the chapter is structured to follow those categories.

First, I will describe the performance of participants on the five matching tasks at the center of this study. Then I will compare the performance of this study's participants to participants in other studies who have performed the same tasks. From there, I will shift to reporting the results of the study's participants on the two aphasia assessments—the WAB-R and CLQT. Then I can introduce important correlations between matching task performance and aphasia assessment performance, including subscales of the WAB-R and domains and individual subtests of the CLQT. Next, I will turn to participant performance on the counting tasks—the number elicitation, confrontation naming, and count list recitation from 1-20 and 20-1. Then, again, I will attend to important correlations—this time between the counting tasks and matching tasks and then the counting tasks and aphasia assessments. In the final section of the chapter, I will consider the important correlations and dissociations between matching task performance and performance on subtests of the CLQT that require spoken or non-spoken responses. Finally, I will explore descriptive connections between performance on the matching tasks and counting tasks that makes use of a distinction between “counting” numbers and “naming” numbers found in participant performance.

Table 2
Participant errors across tasks sorted by total errors, WAB-R AQ scores, CLQT Severity scores

		Task					Total Errors	% Correct	WAB-R AQ	CLQT Severity
		1	2	3	4	5				
Participant	2	0	0	0	1	0	1	97.8	97.4	4.0
	16	0	0	0	1	1	2	95.6	86.7	4.0
	7	0	0	0	1	1	2	95.6	96.8	4.0
	1	0	0	0	1	1	2	95.6	65.5	3.0
	10	0	0	0	1	1	2	95.6	83.8	2.8
	12	1	0	0	1	1	3	93.3	96.5	3.6
	13	0	1	3	1	2	7	84.4	88.5	3.2
	11	0	0	0	5	2	7	84.4	32.1	2.8
	15	0	0	0	4	5	9	80.0	18.3	2.8
	3	2	0	1	3	4	10	77.8	24.4	2.4
	9	1	0	2	2	5	10	77.8	27.4	2.4
	6	4	0	2	2	2	10	77.8	49.4	1.6
	14	2	0	2	4	4	12	73.3	15.3	2.0
	4	5	1	0	3	4	13	71.1	31.2	2.2
	8	5	0	2	6	3	16	64.4	46.0	1.6
	5	4	1	3	7	6	21	53.3	55.9	2.2

Notes. The maximum number of errors on each task is nine. Darker colors indicate more errors and greater severity. Participants sorted by total errors on matching tasks.

Matching Tasks

Participant accuracy. Percent correct scores for all tasks ranged from 53% to 98%, with an average of 82.4% correct and a median of 82.2% correct (Table 2). There was notable variation across participants and matching tasks. Participants made as many as 21 errors (Participant 5) across the combined 45 trials (9 trials per each of the 5 matching tasks), but every participant made an error on at least one trial (Participant 2 was the only participant who made only one error). The mean number of errors for all tasks was 7.9 errors, with a median of 8 errors. Participants responded correctly on 83% of Task 1 trials (with individual participants making as few as 0 and as many as five errors in 9 trials), 98% of Task 2 trials (range of 0 to 1 errors), 90% of Task 3 trials (range of 0 to 3 errors), 70% of Task 4 trials (range of 1 to 7 errors), and 71% of Task 5 trials (range of 0 to 6 errors) (Figure 2).

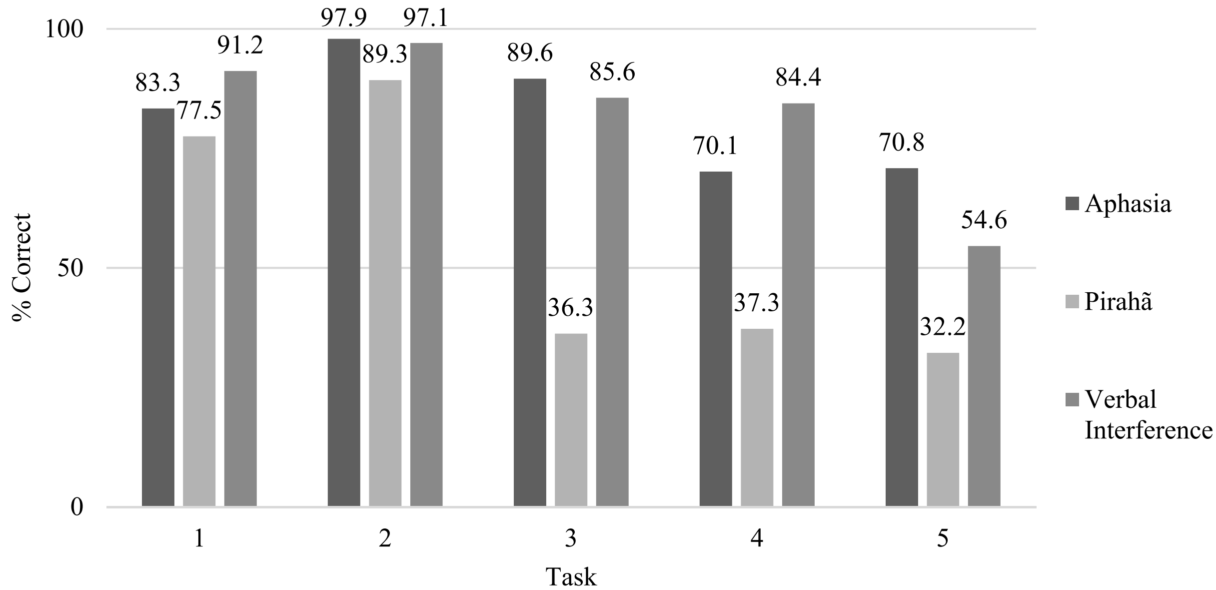

Figure 2: Matching task accuracy per task for participants with aphasia, Pirahã, and adults under verbal interference.

Participants' accuracy decreased as the target quantity increased across all tasks (r^2 = 0.87, p = 0.001) (Figure 3) and for each individual task (Figure 4). Across all tasks, participants responded correctly to 97.5% of trials involving the targets 4 and 5, and became steadily more inaccurate as targets increased in magnitude, until they were correct in only 63.8% and 73.8% of trials involving the targets 11 and 12, respectively. In Task 1, participants made no errors when targets were 4, 5, or 6, but dropped to 68.75% accuracy for targets 10 and 12 and 62.5% accuracy for trials where the target was 11.

By contrast, in the nearly-identical Task 2, participants only made errors when the target was 10 (93.75% accurate) or 12 (87.50% accurate). In Task 3, participants performed at ceiling for targets 4, 5, 6, and 8, but dropped to 75% accuracy for targets 10 and 11, and 68.75%

accuracy when 12 was the target. On Task 4, participants were accurate on 93.75% of trials for targets 4 and 5, but were unable to accurately respond more than 75% of the time for larger targets, correctly responding to only 56.25% of trials for target 9, 62.5% of trials for target 10, 37.5% of trials for target 11, and 68.75% of trials for target 12. Finally, on Task 5, participants responded correctly 93.75% of the time when targets were 4 and 5, dropped off in performance for targets 6 and 7 (62.5% and 68.75% correct, respectively), performed somewhat better with targets 9, 10, and 12 (81.25%, 75%, and 75% correct, respectively), and struggled most with targets 8 and 11, where they were correct on only 43.75% of trials.

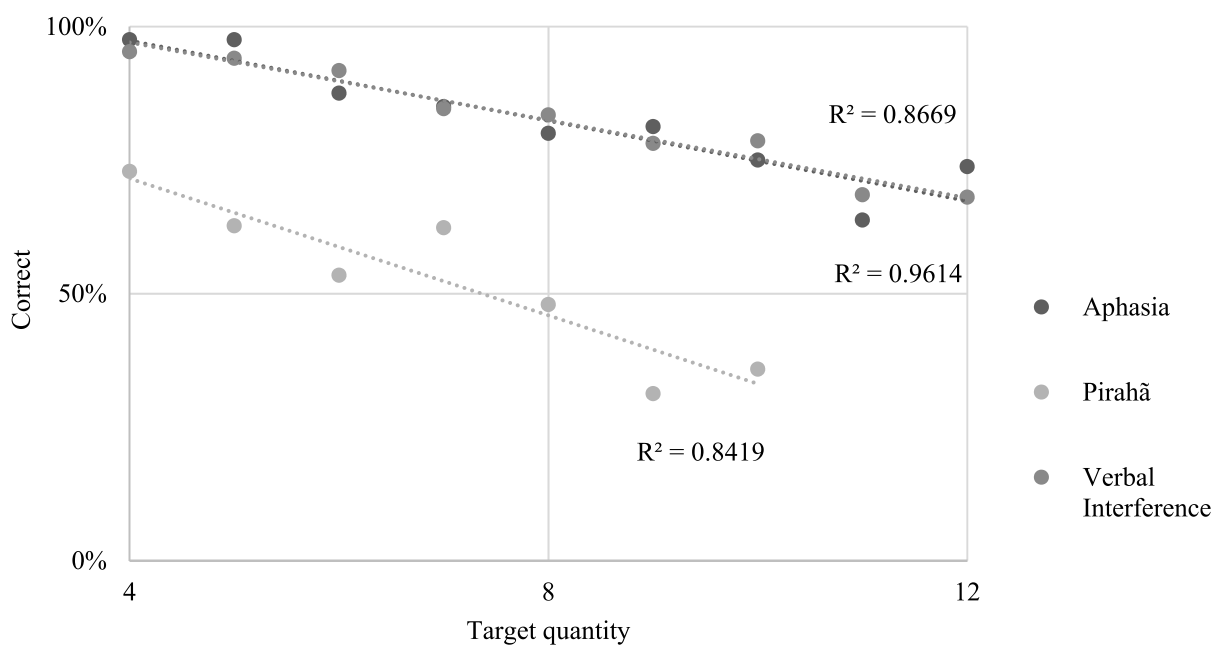

Figure 3: Matching task accuracy across targets for participants with aphasia, Pirahã, and adults under verbal interference.

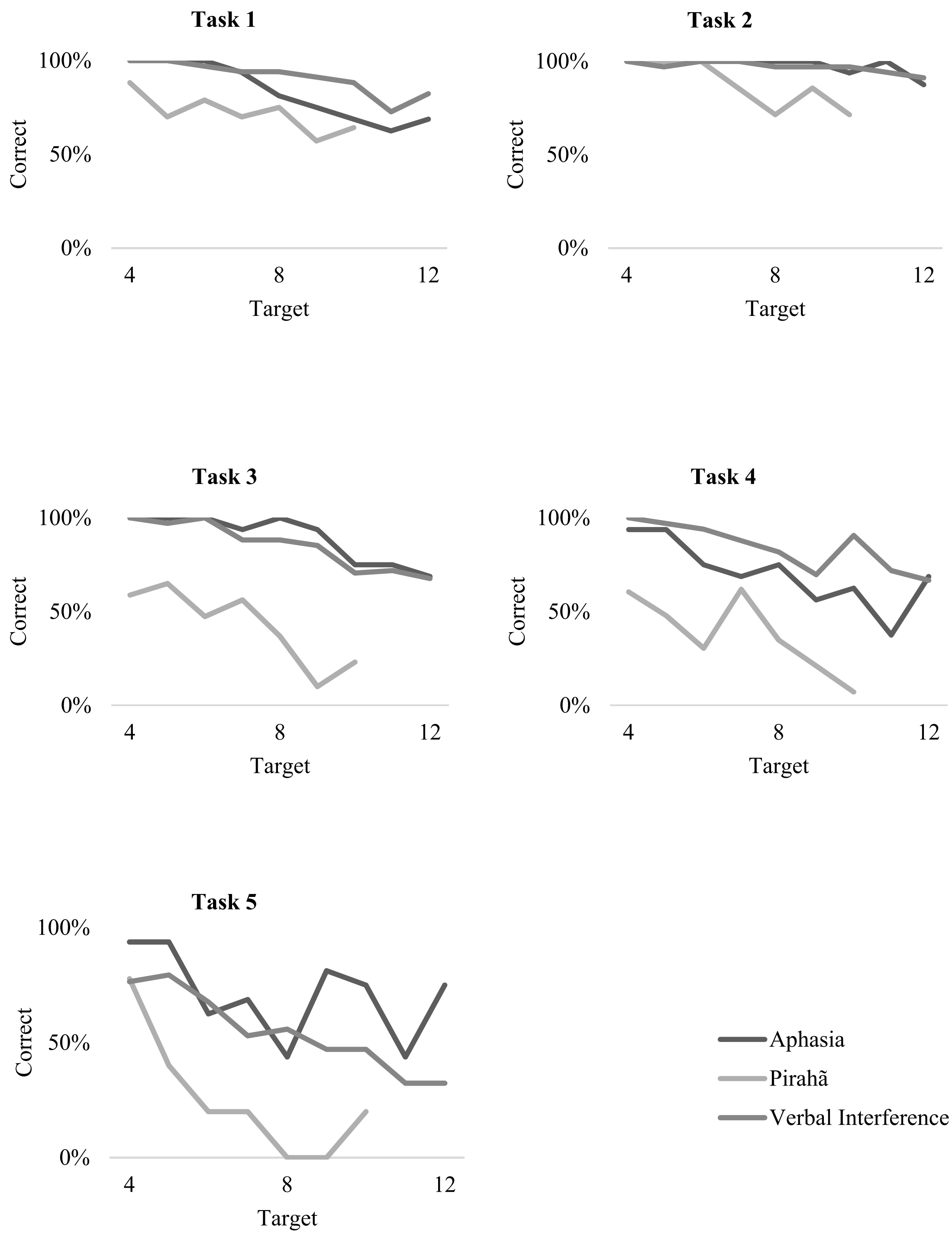

Figure 4: Matching task accuracy per task across targets for participants with aphasia, Pirahã, and adults under verbal interference.

Error magnitude. Across all tasks, average error magnitude—calculated as the mean of the size of the errors made by participants—increased as target quantity increased (r^2 = 0.88, p = 0.001) (Figure 5). In other words, as the size of the target increased, so did the size of the errors made by participants. Average error magnitude was as low as 0.03 and 0.05 for targets 4 and 5 (i.e., participants made relatively smaller errors), increasingly steadily to 0.61 and 0.49 for targets 11 and 12 (i.e., participants made relatively larger errors). The trend toward larger errors as targets increased held for individual tasks as well. In Task 1, participants performed at ceiling for targets less than 7, but the magnitude of errors increased stepwise from there: the average size of a participant's error was only 0.063 for target 7, but reached 0.563 for target 12. In Task 2, participants made few errors, and only gave incorrect responses on targets 10 and 12, where the average magnitude of those errors was 0.125 in both cases. In Task 3, participants performed perfectly (no errors) on trials involving targets up to 7, and then, as in Task 1, made increasingly larger errors as the target increased, with the exception of target 8, where no errors were made.

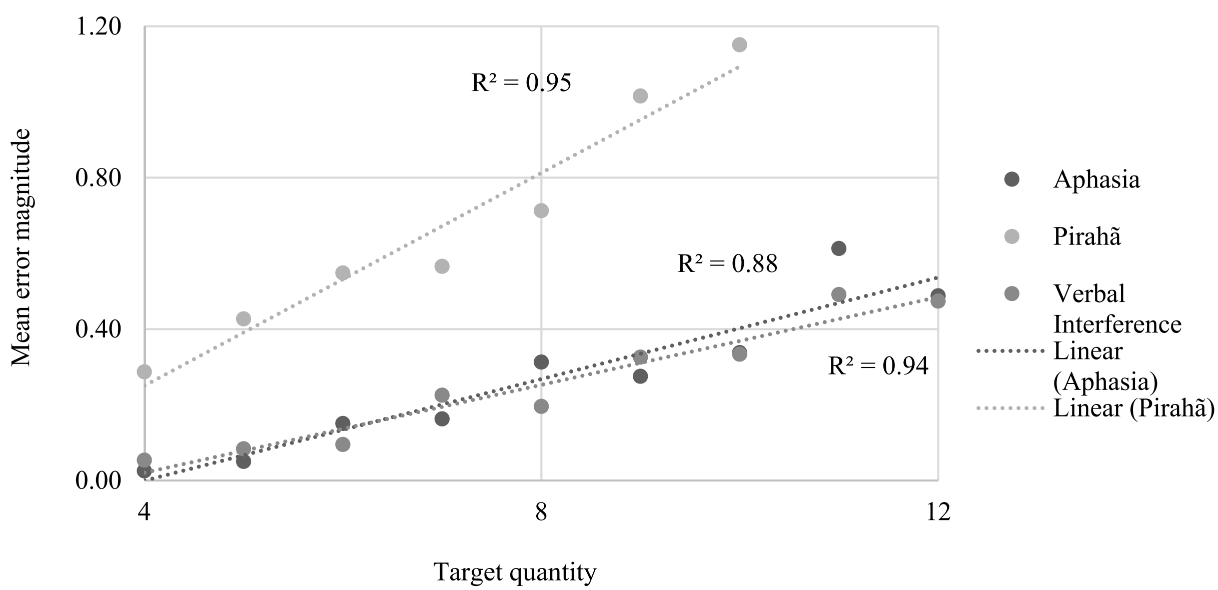

Figure 5: Matching task mean error magnitude across targets for participants with aphasia, Pirahã, and adults under verbal interference.

As in Task 1, participants made errors of an average magnitude of 0.063 for target 7 on Task 3, and reached an average magnitude error of 0.563 for target 12. Participants made errors for each target in Task 4, increasing from errors of an average magnitude of 0.063 for target 4 to an average error magnitude of 0.563 for target 12. However, average magnitude error did not increase in step with the increase of target size: in two cases—targets 9 and 11—the average magnitude of error was higher than not only the previous target, but the subsequent target as well, making those two targets stand out from the rest. Average magnitude error increased from 0.375 for target 8 to 0.625 for target 9, then receded to 0.500 for target 10, before increasing again to 1.063 for target 11 and dropping to 0.563 for target 12. For Task 5, average error magnitude was identical (0.063) for targets 4 and 5, hovered around 0.4 for targets 6 (0.438), 7 (0.375), 9 (0.313), and 10 (0.375), and spiked to 0.938 for target 8, 1.125 for target 11, and 0.625 for target 12. The conspicuous jump in average error magnitude for target 8 on Task 5 is the result of an unusually large error on this trial by participant 12, who responded with 12 to a target of 8. The error was uncharacteristic for the target number on that task, across tasks, and for participant 12. By contrast, target 11 on Task 5 included three errors of a magnitude larger than 1 or 2, consistent with how participants performed on other tasks in response to target 11.

Coefficient of variation. Dating back to Whalen et al. (1999), coefficient of variation (CoV) has been used to measure whether study participants are estimating when asked to represent large quantities. Whalen et al. (1999) showed that, like other animals, humans have to estimate the size of a set, they produce increasingly variable responses in proportion to the magnitude of the set they are trying to quantify. This also means that they produce a constant CoV across magnitudes. A CoV is the standard deviation of a data set divided by its mean. A constant CoV across targets means that there is proportionally less response variability when the target is smaller as when the target is larger. The range of responses to a larger target is larger than the range of responses for a smaller target—that is, the standard deviation is larger for a larger target. However, that range is always proportional to the size of the target—about 0.15.

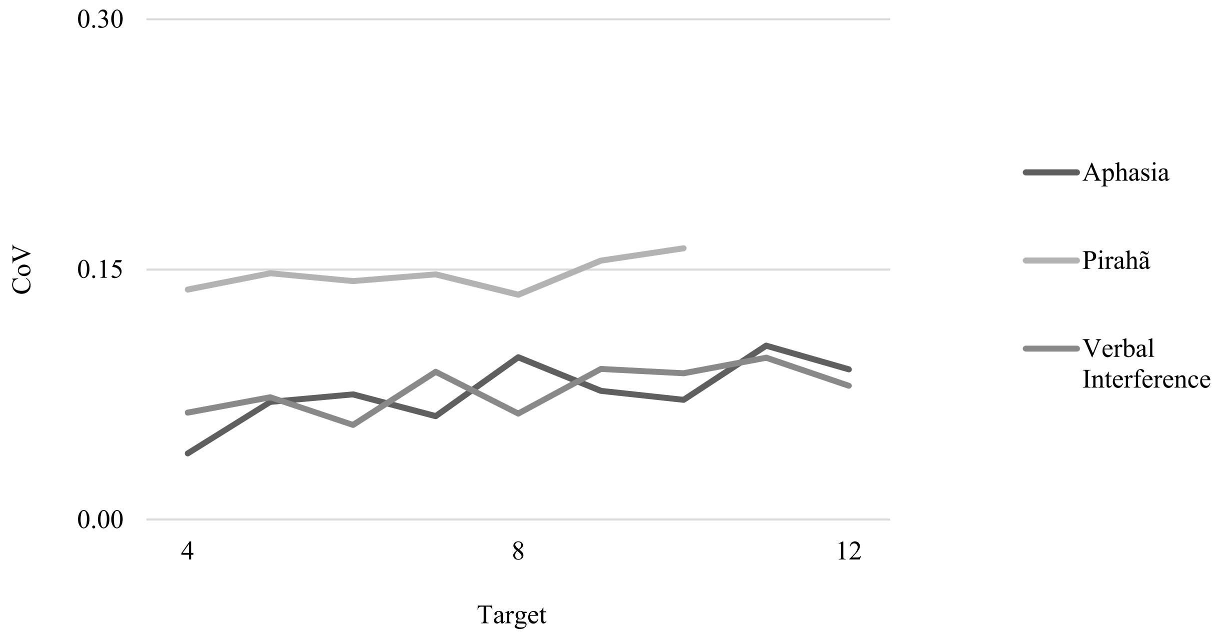

Figure 6: Matching task CoV across targets for participants with aphasia, Pirahã, and adults under verbal interference.

In this case, participants produced an average CoV of 0.8 across all tasks and targets—smaller than expected if participants were estimating quantities, and smaller than what had been seen in studies with the Pirahã, but in line with English speakers under verbal interference from Frank et al. (2012). Participants produced a flat CoV across target quantities and tasks (Figure 6).Participants produced a range of CoV between 0.6 and 0.10 across all tasks and all targets with the exception of target 4, which produced the lowest CoV at 0.4. Targets 8 and 11 produced the highest CoV across tasks at 0.10. On Task 1, participants produced a CoV of 0.05, largely influenced by participants performing at ceiling for targets 4, 5, and 6. The CoV for target 7 increased to 0.04, then 0.08 for targets 8, 9, and 10, 0.7 for target 11, and 0.09 for target 12.

Because there were hardly any errors on Task 2, the CoV was a predictably low 0.01 across targets, only rising to 0.05 and 0.03 for targets 10 and 12, respectively. As in Task 1, at ceiling performances for targets 4, 5, and 6 produced a CoV of 0.00 for those targets on Task 3. The average CoV across targets for this task was 0.03, with target 12 producing the highest CoV, 0.09. On Task 4, the average CoV increased to 0.10 across all targets, with a low CoV 0.06 for target 4, but otherwise hovering between 0.8 (targets 7 and 10) and 0.12 (target 11). Task 5 produced an average CoV of 0.11, distributed between lows of 0.5 and 0.6 for targets 4 and 5, and highs of 0.18 and 0.16 for targets 8 and 11. As was the case across all tasks and targets, individual tasks produced flat CoV across targets—the CoV slope was 0.013 for Task 1, 0.004 for Task 2, 0.012 for Task 3, 0.000 for Task 4, and 0.008 for Task 5 (Figure 7). The similarity between the CoVs produced by the aphasia participants and English speakers under verbal interference could be accounted for by the fact that those two groups of participants, unlike the Pirahã, exist in a culture that conceptualizes exact quantity and gives language to it. The difference between the CoVs of those participant groups and the CoVs of the Pirahã could likewise be explained by participants' "differential cultural experience with mathematics and other uses of exact numerosity," which "led to their relatively more precise representation of analog magnitude" (Frank et al., 2012, p. 82).

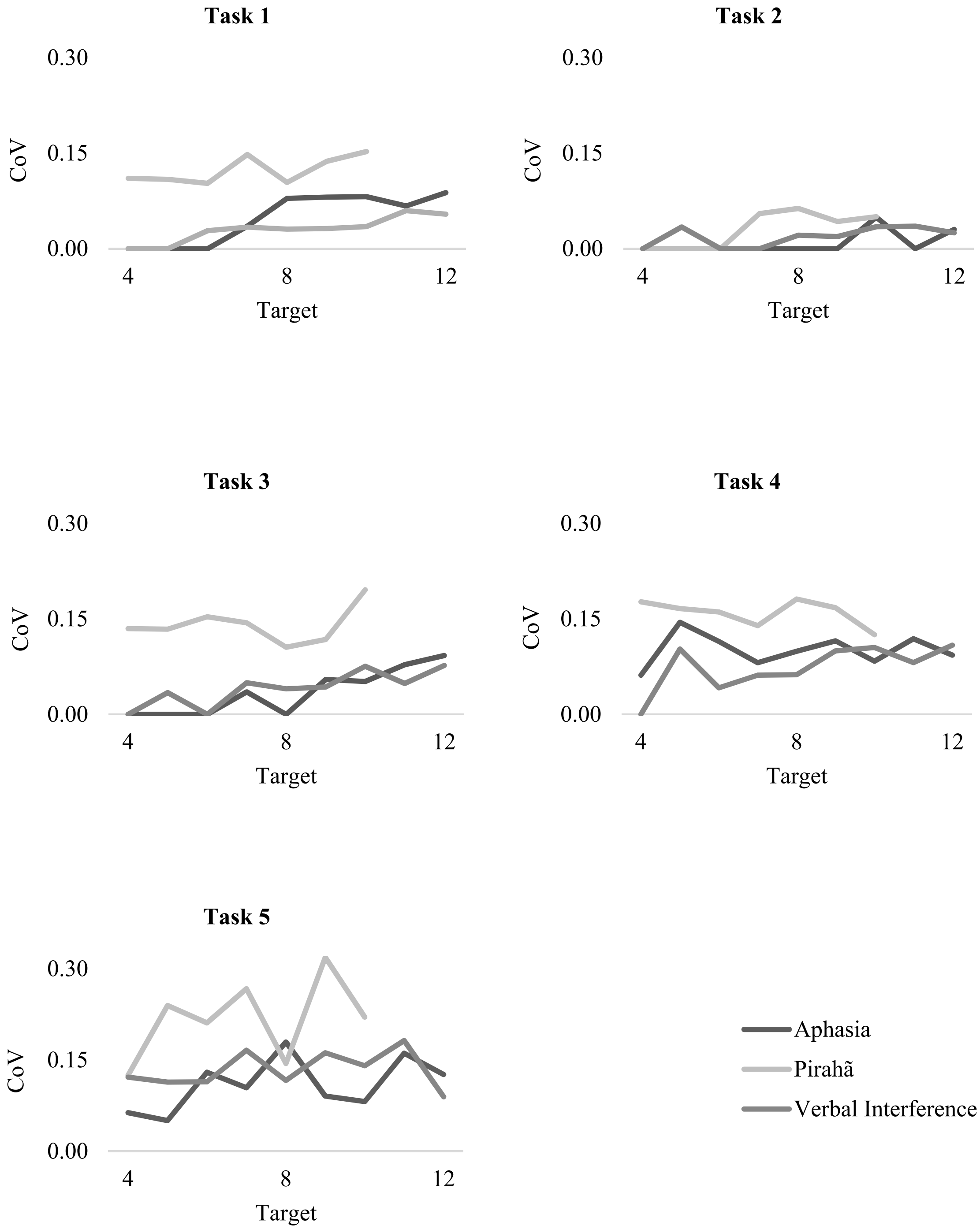

Figure 7: Matching task CoV per task across targets for participants with aphasia, Pirahã, and adults under verbal interference.

Performance compared to previous studies' populations. Across analyses, aphasia participants' performance was remarkably similar to the performance of English speakers under verbal interference from Frank et al. (2012) (Figures 2 – 7). Across tasks and targets, aphasia participants and English speakers under verbal interference produced nearly identical declines in accuracy as target quantity increased, nearly identical increases in error magnitude as target quantity increased, and nearly identical CoV. Participants with aphasia and English speakers under verbal interference showed only small differences in accuracy on Tasks 1, 2, and 3—English speakers were 8% more accurate on Task 1, while aphasia participants were 1% more accurate on Task 2 and 4% more accurate on Task 3. On Task 4, English speakers under verbal interference were 14% more accurate than participants with aphasia, while on Task 5, aphasia participants were 16% more accurate than English speakers under verbal interference. Responses to targets on individual tasks mirrored the groups' performance on those tasks overall: the aphasia group and English speakers both demonstrated decreasing accuracy as targets increased on Task 1, but the aphasia group was consistently less accurate for every target. The same was true on Task 4, but reversed on Tasks 3 and 5, where the aphasia group was consistently more accurate than the English speakers under verbal interference, despite both groups growing progressively less accurate as the targets increased. Similarly, the aphasia participants and the English speakers under verbal interference showed nearly identical CoV across all targets and tasks, and each group's mean CoV for each task was within 0.03 of the other group's (Figures 6 and 7). The slopes of these CoV across targets per task were also similarly flat, with no greater than a statistically insignificant 0.009 difference between them.

Compared to the Pirahã (Figures 2 – 7, aggregated from Everett & Madora, 2012; Frank et al., 2008; and Gordon, 2004), participants with aphasia and English speakers under verbal interference were generally more accurate and made smaller errors, but all three groups showed similar patterns of responding. Across tasks and targets, aphasia participants and the Pirahã declined in accuracy as target quantity increased, made larger errors as target quantity increased, and produced flat CoV. Participants with aphasia and Pirahã were similarly accurate across targets on Task 1 (Pirahã = 78% correct, Aphasia = 83% correct), but their performances diverged from there. Although the Pirahã—like the English speakers under verbal interference and the aphasia participants—performed best on Task 2, they were only 89% accurate, whereas the English speakers and aphasia participants were 97% and 98% accurate, respectively. After the first two tasks, however, Pirahã performance dropped off significantly: they were no better than 37% accurate in Tasks 3, 4, and 5.

Another way of expressing this would be to say that Pirahã performance dropped off after Task 2, while aphasia participants' performance dropped off after Task 3, and the English speakers under verbal interference performed notably worse only on Task 5. Responses to targets on individual tasks mirrored the groups' performance on those tasks overall: all three groups demonstrated decreasing accuracy as targets increased on Task 1, but the English speakers under verbal interference were most accurate at every target, the aphasia group was consistently less accurate for every target, and the Pirahã were least accurate at every target. The same pattern was true on Tasks 2, 3, 4, and 5, with the exceptions of Tasks 3 and 5, where the aphasia group was more accurate than the English speakers, as noted above. While the aphasia participants and the English speakers under verbal interference showed nearly identical CoV across all targets and tasks (approximately 0.10), the Pirahã produced at CoV of 0.15. With the exception of Task 2, where all three groups had a similar CoV (English speakers = 0.02, Pirahã = 0.03, Aphasia = 0.01), the Pirahã consistently produced a CoV that was 0.06 (Task 4) to 0.10 (Task 3) higher than the next closest group. The slopes of the Pirahã CoV across targets per task were as flat as the other groups.

Western Aphasia Battery-Revised (WAB-R) and Cognitive Linguistic Quick Test (CLQT)

Participant performance. Aphasia participants completed the WAB-R and CLQT, two formal assessments or language and cognition. The WAB-R computes an Aphasia Quotient (AQ) with a maximum score of 100, a Spontaneous Speech subscale with a maximum score of 20, and Auditory Verbal Comprehension, Repetition, and Naming and Word Finding subscales, each with a maximum score of 10. In all cases, a lower score indicates greater impairment. Participants ranged in AQ from 15.30 to 97.40 (mean = 57.20, median = 52.65), had Spontaneous Speech scores that ranged from 0.00 to 19.00 (mean = 11.50, median = 11.50), had Auditory Verbal Comprehension Scores that ranged from 3.45 to 10.00 (mean = 7.38, median = 7.75), had Repetition Scores that ranged from 0.40 to 10 (mean = 5.15, median = 4.80), and Naming and Word Finding Scores that ranged from 0.00 to 9.90 (mean = 5.14, median = 4.40).

The CLQT gives both raw scores and severity ratings (on a scale of 1 to 4, where lower scores indicate greater impairment) across Attention, Memory, Executive Function, Language, and Visuospatial domains. Those domains are used to calculate a Composite Severity Rating (on a scale of 1 to 4, where lower scores indicate greater impairment). Individual subtests load on to multiple domains, except for a clock drawing task, which does not load onto any domain. Participants had Composite Severity Ratings that ranged from 1.60 to 4.00 (mean = 2.79, median = 2.80), Attention Domain scores ranging from 68 to 202 (mean = 164.06, median = 175.00), Memory Domain scores ranging from 40 to 168 (mean = 102.69, median = 94.50), Executive Function Domain scores ranging from 12 to 32 (mean = 23.56, median = 23.50), Language

Domain scores ranging from 1 to 33 (mean = 16.16, median = 14.50), and Visuospatial Domain scores ranging from 52 to 102 (mean = 83.13, median = 85.00). Participants had average severity ratings of 3.00 or better for the Attention (mean = 3.25, median = 3.00), Executive Function (mean = 3.00, median = 3.00), and Visuospatial (mean = 3.81, median = 4.00) Domains, and average severity ratings below 2.00 for the Memory (mean = 1.94, median = 1.00) and Language (mean = 1.94, median = 1.00) Domains. Participants' Clock Drawing subtest scores ranged from 0 to the maximum score of 13, with a mean of 8.69 and median of 8.50.

Correlations between matching task and WAB-R performance. There were *no* significant correlations between performance on Tasks 1, 2, and 3 with any WAB-R scores. Conversely, there were significant correlations (at or below the 0.05 level) between performance on Tasks 4 and 5 with AQ score and *every* WAB-R subscale except for Auditory Verbal Comprehension (Table 3). Performance on Task 5 significantly correlated at the 0.01 level with every WAB-R subtest and the AQ score, while performance on Task 4 correlated at the 0.01 level with only the Naming and Word Finding scores (r = 0.66). Unsurprisingly, there were also significant correlations at the 0.01 level between WAB-R AQ and subtest scores and tasks where the target is not visible when responding—Tasks 4 and 5 considered together.

Table 3
Correlations between matching task performance and WAB-R scores.

Task(s)	Aphasia Quotient	Spontaneous Speech	Auditory Verbal Comprehension	Repetition	Naming and Word Finding
1	.408	.298	.426	.378	.487
2	-.022	-.039	.041	-.007	-.102
3	.229	.181	.097	.165	.330
4	.614*	.622*	.373	.498*	.661**
5	.774**	.741**	.686**	.675**	.721**
all	.633**	.578*	.507*	.541*	.674**
target visible (1, 2, 3)	.363	.267	.330	.318	.446
target not visible (4, 5)	.748**	.736**	.566*	.632**	.748**

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Correlations between matching task and CLQT performance. Overall, correlations were significant at the 0.01 level between CLQT Composite Severity Ratings and performance on Task 1 ($r = 0.77$), Task 4 ($r = 0.64$), and Task 5 ($r = 0.64$) (Table 4). CLQT Visuospatial domain correlated most strongly with performance on Task 1, while the Language and Memory domains correlated most strongly with performance on Task 5. Correlations were significant at the 0.01 level between Task 1 and CLQT Attention ($r = 0.64$), Executive Function ($r = 0.71$), and Visuospatial Skills ($r = 0.75$) domains. Task 3 also correlated significantly at the 0.01 level with Visuospatial Skills ($r = 0.67$). Tasks where the target is not visible during response (Tasks 4 and 5) correlated significantly at the 0.01 level with the Memory ($r = 0.87$) and Language ($r = 0.84$) domains. Tasks 4 and 5, considered separately, demonstrated similar and significant at the 0.01 level correlations with the Memory and Language domains of the CLQT.

Table 4
Correlations between matching task performance and CLQT Domain and Composite Severity Rating scores.

Task(s)	Composite Severity Rating	Attention	Memory	Executive Function	Language	Visuospatial Skills
1	.767**	.643**	.572*	.710**	0.424	.748**
2	0.157	0.036	0.308	0.020	0.094	0.187
3	.562*	.550*	0.432	.543*	0.338	.669**
4	.635**	0.402	.722**	0.408	.708**	0.331
5	.640**	0.286	.884**	0.353	.852**	0.228
all	.798**	.555*	.824**	.598*	.726**	.583*
target visible (1, 2, 3)	.763**	.658**	.601*	.698**	0.434	.798**
target not visible (4, 5)	.691**	0.376	.867**	0.414	.842**	0.306

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Five CLQT subtests did not load onto either the Language or Memory domains—Symbol Cancelation, Symbol Trails, Mazes, Design Generation, and Clock Drawing. Amongst these, there were few significant correlations with matching task performance (Table 5). There were significant correlations at the 0.05 level between performance on the Mazes subtest and performance on Task 1 (*r* = 0.59) and Task 3 (*r* = 0.55), which in turn drove a significant correlation at the 0.01 level between performance on the Mazes subtest and task performance when the target was visible when responding (*r* = 0.66). There was a significant correlation at the 0.01 level between performance on the Design Generation subtest and task performance on Task 1 (*r* = 0.67). There were significant correlations at the 0.01 level between Clock Drawing subtest scores and performance on Task 1 (*r* = 0.76) and significant correlations at the 0.5 level between Clock Drawing subtest scores and performance on Task 5 (*r* = 0.60). The Clock Drawing subtest doesn't load on to any of the CLQT domains.

Three CLQT subtests load onto both the Language or Memory domains—Personal Facts, Story Retelling, and Generative Naming (Table 5). These subtests all correlated significantly with performance on Tasks 4 and 5—and all but performance on Task 4 and the Story Retelling subtest correlated at the 0.01 significance level. Consequently, all of these subtests correlated quite strongly with tasks where the target is not visible on response (Tasks 4 and 5 considered together). The only other significant correlation found here was between Task 1 and the Story Retelling subtests (*r* = 0.61).

The Design Memory subtest only loads on the Memory Domain of the CLQT. There were significant correlations at the 0.01 level between Design Memory subtest scores and performance on Task 2 (*r* = 0.65), Task 3 (*r* = 0.67), and significant correlations at the 0.05 level between Design Memory subtest scores and performance on Task 1 (*r* = 0.59) and Task 5 (*r* = 0.52). The strong correlations with Tasks 1, 2, and 3 led to a significant correlation at the 0.01 level between the Design Memory subtest and matching tasks where the target was visible when responding (*r* = 0.76).

Table 5

Correlations between matching task performance and CLQT subtest scores.

	Language Domain	Language and Memory Domains			Memory Domain	Neither Language nor Memory Domain				No Domain
Task(s)	Confrontation Naming	Personal Facts	Story Retelling	Generative Naming	Design Memory	Symbol Cancelation	Symbol Trails	Mazes	Design Generation	Clock Drawing
1	.233	.389	.605*	.347	.587*	.446	.025	.585*	.668**	.762**
2	-.101	.227	.116	.119	.647**	-.042	-.242	.296	-.147	.282
3	.197	.231	.393	.446	.670**	.385	.181	.554*	.096	.132
4	.585*	.732**	.585*	.602*	.333	.341	-.057	.185	.144	.495
5	.724**	.827**	.767**	.711**	.518*	.201	-.068	.046	.037	.596*
all	.537*	.702**	.730**	.643**	.653**	.403	-.013	.405	.292	.656**
target visible (1, 2, 3)	.222	.386	.580*	.431	.755**	.449	.056	.657**	.465	.602*
target not visible (4, 5)	.706**	.842**	.729**	.709**	.457	.297	-.068	.129	.100	.589*

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

The Confrontation Naming subtest only loads on the Language Domain of the CLQT. A significant correlation at the 0.01 level was found between Confrontation Naming subtest scores and Task 5 performance ($r = 0.72$) and a significant correlation at the 0.05 level was found between Confrontation Naming subtest scores and Task 4 ($r = 0.59$). The strong correlations with Tasks 4 and 5 led to a significant correlation at the 0.01 level between Confrontation Naming and matching tasks where targets were not visible when responding ($r = 0.71$).

Counting Tasks

Number elicitation task. In the number elicitation task, participants were asked to name the number of spools of thread in a set as they increased from 1 to 12, and then decreased from 12 to 1. Participants averaged 8.44 correct responses as the target set of items increased in quantity (range: 0 – 12 correct responses). However, the median number of correct responses was 11 and 7 participants performed at ceiling. Participants averaged 7.81 correct responses as the target array decreased in quantity (range: 0 – 12 correct responses), with 9 as the median number of correct responses, and only 5 participants performed at ceiling. Combining the increasing and decreasing number elicitation totals, participants ranged from 0 to 24 correct responses, averaging 16.25 correct responses and a median of 19.5 correct responses. Five participants correctly identified the number of items in all 24 sets and 2 participants were unable to correctly identify the number of items in any set.

Confrontation naming task. Participants were asked to identify Arabic numerals 1 to 20 on flashcards presented in a random order. Participants averaged 13.25 correct responses and a median of 16 correct responses. Six participants scored at ceiling and 2 failed to give a correct response.

Table 6
Correlations between matching task performance and counting task performance.

Task(s)	Number Elicitation Up	Number Elicitation Down	Number Elicitation Total	Confrontation Naming	Count List Recitation Up	Count List Recitation Down	Count List Recitation Total	All Counting Tasks
1	.211	.165	.191	.074	.062	.268	.195	.172
2	.048	-.099	-.024	-.250	-.219	.046	-.072	-.105
3	.220	.134	.181	.119	.237	.199	.232	.199
4	.571*	.397	.494	.345	.311	.468	.433	.447
5	.605*	.462	.543*	.524*	.556*	.637**	.651**	.616*
all	.503*	.356	.438	.314	.334	.492	.459	.436
target visible (1, 2, 3)	.237	.151	.198	.063	.108	.266	.215	.181
target not visible (4, 5)	.637**	.464	.561*	.467	.464	.595*	.582*	.572*

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Count list recitation task. Participants were asked to count from 1 to 20 and from 20 to 1. Reciting the count list from 1 up to 20, participants averaged correctly naming 14.94 numbers, with a median of 19 correct responses. Eight participants made no errors and 2 gave no response. When reciting the count list from 20 down to 1, participants correctly named 8.88 numbers of the count list on average, with a median of 4.5 correct responses. Five participants made no errors, and 2 gave no response. Combining the increasing and decreasing count list recitation task totals, participants averaged 23.81 correct responses and had a median of 22 correct responses out of 40. Five participants made no errors on the task and 2 gave no correct responses.

Across all counting tasks, participants averaged 53.31 correct responses and a median of 58.5 correct responses out of 84 possible correct responses. Five participants scored at ceiling, and two did not give any correct responses.

Correlations between matching task and counting task performance. There were several significant correlations between performance on the counting tasks and performance on the matching tasks (Table 6). There were significant correlations at the 0.05 level between performance on the increasing elicitation task and performance on Task 4 (r = 0.57) and Task 5

(r = 0.61). This drove a significant correlation at the 0.01 level between performance on the increasing elicitation task and on the matching tasks when the target was not visible when responding (r = 0.64). Otherwise, significant relationships were restricted to Task 5 performance, with significant correlations found at the 0.05 level between performance on Task 5 and both elicitation tasks (up and down totaled) (r = 0.54), Task 5 and confrontation naming (r = 0.52), Task 5 and the ascending count list recitation task (r = 0.56), and all the counting tasks considered together (r = 0.62). There was a significant correlation at the 0.01 level between performance on the descending count list recitation task and performance on Task 5 (r = 0.64). Because of the correlations with the ascending and descending count list recitation tasks, there was a significant correlation at the 0.01 level between performance on the combined count list recitation tasks and performance on Task 5 (r = 0.65).

Table 7
Correlations between counting task performance and WAB-R scores.

	Number Elicitation			Confrontation Naming	Count List Recitation			All Counting Tasks
WAB-R	Up	Down	Total		Up	Down	Total	
AQ	.706**	.678**	.702**	.819**	.657**	.756**	.771**	.796**
Spontaneous Speech	.781**	.722**	.763**	.850**	.693**	.762**	.791**	.831**
Auditory Verbal Comprehension	.415	.437	.432	.558*	.453	.487	.510*	.521*
Repetition	.727**	.724**	.736**	.836**	.607*	.741**	.739**	.793**
Naming and Word Finding	.663**	.636**	.659**	.780**	.628**	.687**	.715**	.745**

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Correlations between counting task and aphasia assessment performance. Performance on the counting tasks correlated strongly with WAB-R subtest scores and AQ (Table 7). There were significant correlations at the 0.01 level between performance on each counting task and AQ, Spontaneous Speech subtest scores, and Naming and Word Finding subtest scores. There were also significant correlations at the 0.01 level between Repetition subtest scores and performance on every counting task except for the ascending counting list task (*r* = 0.61), where there was a significant correlation at the 0.05 level. Meanwhile, there were only three significant correlations at the 0.05 level between performance on counting tasks and Auditory Verbal Comprehension subtest scores: the confrontation naming task (*r* = 0.56), overall performance on the count list recitation task (*r* = 0.51), and overall performance on all counting tasks (*r* = 0.52).

Table 8

Correlations between counting task performance and CLQT Domain and Composite Severity Rating scores.

	Number Elicitation			Confrontation Naming	Count List Recitation			All Counting Tasks
CLQT	Up	Down	Total		Up	Down	Total	
Composite Severity Rating	.498*	.524*	.518*	.526*	.389	.496	.487	.526*
Attention	.374	.375	.380	.315	.349	.141	.248	.313
Memory	.603*	.512*	.567*	.575*	.438	.728**	.653**	.636**
Executive Function	.214	.282	.251	.307	.195	.279	.263	.281
Language	.671**	.616*	.653**	.726**	.616*	.755**	.751**	.748**
Visuospatial Skills	.213	.236	.228	.173	.134	.088	.116	.168

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Correlations between counting task and CLQT performance. Performance on counting tasks correlated significantly with the Memory and Language CLQT domain scores, and with CLQT Composite Severity Ratings, but not with Attention, Executive Function, or Visuospatial Skills domain scores (Table 8). Performance on nearly every counting task—number elicitation, confrontation naming, and count list recitation—correlated significantly with CLQT Language domain scores at the 0.01 level, and the few exceptions were significant at the 0.05 level. Similarly, performance on nearly every counting task correlated significantly with CLQT Memory domain scores at the 0.05 level, while there were significant correlations at the 0.01 level between CLQT Memory domain scores and descending count list recitation task performance ($r = 0.73$), count list recitation task performance ($r = 0.65$), and overall counting task performance ($r = 0.64$). There were also significant correlations at the 0.05 level between Composite Severity Ratings and performance on the number elicitation tasks—ascending ($r = 0.49$), descending ($r = 0.52$), and overall ($r = 0.52$)—the confrontation naming task ($r = 0.53$), and overall counting task performance ($r = 0.53$).

Table 9
Correlations between counting task performance and CLQT subtest scores.

CLQT		Number Elicitation			Confrontation Naming	Count List Recitation			All Counting Tasks
Domain	Subtest	Up	Down	Total		Up	Down	Total	
Language	Confrontation Naming	.671**	.652**	.671**	.797**	.717**	.696**	.761**	.776**
Language and Memory	Personal Facts	.634**	.493	.573*	.548*	.437	.654**	.607*	.608*
	Story Retelling	.528*	.550*	.546*	.598*	.485	.698**	.657**	.638**
	Generative Naming	.517*	.480	.506*	.615*	.522*	.639**	.637**	.620*
Memory	Design Memory	.141	.061	.103	.091	-.051	.368	.206	.156
Neither Language Nor Memory	Symbol Cancelation	.380	.344	.368	.281	.383	.071	.219	.287
	Symbol Trails	-.143	-.067	-.108	-.005	.038	-.293	-.165	-.116
	Mazes	.138	.185	.163	.072	-.018	.057	.027	.079
	Design Generation	-.085	.004	-.042	-.035	-.142	.120	.010	-.016
N/A	Clock Drawing	.397	.420	.414	.391	.254	.471	.409	.423

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Breaking those CLQT domains into their constituent subtests highlights the ways that specific subtests drove the correlations we saw above. Most CLQT subtests did not correlate significantly with performance on the counting tasks. Those that did are subtests that either load onto both the Language and Memory domains or that load only onto the Language domain (Table 9). The Confrontation Naming subtest is the lone subtest that loads exclusively onto the CLQT Language domain, and performance on every counting task correlated significantly with Confrontation Naming subtest scores at the 0.01 level. By contrast, the Design Memory subtest only loads on the CLQT Memory Domain—and here there were no significant correlations with performance on the counting tasks.

The three CLQT subtests that load onto both the Language or Memory domains—Personal Facts, Story Retelling, and Generative Naming—correlated significantly with nearly all of the counting tasks. The exceptions were the descending number elicitation task and the ascending count list recitation task. Meanwhile, correlations were particularly strong between these subtests and the descending count list recitation task.

Table 10

Correlations between matching task performance and CLQT subtests grouped by whether they require a spoken response.

Task(s)	Spoken Response	Non-spoken Response	
1	.424	.866**	^^
2	.094	.138	
3	.338	.467	
4	.708**	.426	
5	.852**	.366	^^
all	.726**	.653**	
target visible (1, 2, 3)	.434	.788**	
target not visible (4, 5)	.842**	.431	^

Notes. *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed). ^Difference between correlations is significant at the 0.05 level (2-tailed). ^^Difference between correlations is significant at the 0.01 level (2-tailed).

Spoken and Non-Spoken CLQT Subtests

Because of the above pattern of correlations, individual subtests of the CLQT were divided into those that require a verbal response and those that do not, and then compared to counting task performance and matching task performance. Unsurprisingly, six of eight correlations between performance on the counting tasks—all of which require a spoken response—and CLQT subtests that require spoken responses were significant at the 0.01 level and all eight were significant at the 0.05 level. There were no significant correlations between counting task performance and performance on CLQT subtests that require a non-spoken response. Importantly, significant correlations at the 0.01 level were found between CLQT subtests requiring a spoken response and performance on Task 4 ($r = 0.71$), Task 5 ($r = 0.85$), and tasks were the target is not visible during response ($r = 0.84$) (Table 10). At the same time, there were significant correlations at the 0.01 level between CLQT subtests not requiring a-spoken response and performance on Task 1 ($r = 0.87$) and tasks were the target is visible during response ($r = 0.79$).

Using guidelines suggested by Lee and Preacher (2013), there were significant differences found at the 0.01 level for correlations between performance on Task 1 and CLQT subtests requiring spoken responses ($r = 0.42$) and performance on Task 1 and CLQT subtests requiring non-spoken responses ($r = 0.87$). Conversely, there were significant differences found at the 0.01 level for correlations between performance on Task 5 and CLQT subtests requiring spoken responses ($r = 0.85$) and performance on Task 5 and CLQT subtests requiring non-spoken responses ($r = 0.37$).

Matching, Counting, and Naming

In this section, I will take a closer look at the relationships between performance on the matching tasks and counting tasks. First, I look at whether individual targets proved more difficult for participants across tasks. Then I turn examining individual participant performance, comparing matching task and counting task performance and adding details about their aphasia assessment performance as appropriate. From there, I will attempt to isolate the impact of participant counting ability on matching task performance, and then differentiate the impact of "matching" (performance on the matching tasks), "naming" (performance on the Confrontation Naming task), and "counting" (performance on the ascending Number Elicitation and Count List

Recitation tasks). In these sections I will examine not only the overall performance of the study sample, but also the performances of individual participants.

Matching and counting task performance across targets. There were no relationships found between participant responses to individual target numbers on the matching and counting tasks. That is, there was no evidence that any participant's aphasia manifested with particular targets on both matching and counting tasks, interfering with the participant's ability to correctly match, name, or count certain numbers (e.g., a specific "seven deficit"). Just as participants generally made more frequent and larger errors on the matching tasks when targets were larger, there were more frequent errors for larger targets across the counting tasks. That said, participants made between 22 and 26 total errors for counting targets 4 through 9, and then made 28, 32, 38 total errors for targets 10, 11, and 12, respectively. This may hint at either the increasing difficulty in tracking larger target quantities or struggling more with reciting the count list past 10.

Individual participant performance on the matching and counting tasks. There were two participants (8 and 13) who appeared to be much better at the counting tasks than at the matching tasks, as defined by these participants having a difference of five or more in the number of errors they made on the matching tasks compared to the counting tasks. Participant 8 made sixteen matching task errors but no counting task errors, while participant 13 made seven matching task errors and no counting errors. There were seven participants whose performance on matching and counting tasks were within five errors of each other. Five of these participants (1, 7, 10, 12, and 16) did well on both matching and counting tasks, making fewer than ten errors on each set of tasks. Two participants (4 and 5) made more than ten errors on the matching tasks and ten or more counting tasks errors. Seven participants appeared to be much better at the matching tasks than they were at the counting tasks, again defined by these participants having a difference of five or more in the number of errors they made on the counting tasks compared to the matching tasks. Four of those seven participants (3, 6, 9, 14) also made more than ten errors on both the matching tasks and then counting tasks. Two participants (11 and 15) made more than ten counting task errors but fewer than ten matching task errors. Participant 2 was the lone participant who made fewer than ten errors on both matching and counting tasks but made more than five more counting task errors than matching task errors. In brief, the participant population produced a diverse array of results. Seven participants were much better at the matching tasks than the counting tasks, seven were equally good or bad at both types of tasks, and two were better at the counting tasks than the matching tasks. At the same time, there was diversity within these groups. There were five participants who were equally good at both the matching and counting tasks and two participants who were equally bad. Meanwhile the participants who were stronger at one set of tasks or the other could be divided into groups that struggled across both sets of tasks and those who did not.

All participants who made ten or more errors on the counting tasks made more counting errors than matching errors. All participants who made ten or more errors on the matching tasks made more counting errors than matching errors, except participant 8, who made no counting errors and sixteen matching errors. Looking more closely at their matching task performance, participant 8 made five errors on Task 1 and two errors on Task 3, for a total of seven errors when the target was visible, and made six errors on Task 4 and three errors on Task 5, for a total of nine errors when target was not visible. The possibility that participant 8 did not understand the tasks is undercut by their perfect performance on Task 2 and the fact that they correctly responded on the first three trials of Task 1. Participant 8's scores on the WAB-R and CLQT were also noteworthy—while all their scores were below the mean of our participants, they had the lowest scores of any participant on the Naming and Word Finding section of the WAB-R (1.4), shared the lowest CLQT Composite Severity Rating score (1.6), and had the second lowest score on the CLQT Visuospatial Domain (59). Meanwhile, participant 13, who made seven matching errors but no counting errors, presented somewhat differently, making no errors on Task 1, an error on Task 2, and three errors on Task 3—a total of four errors when the target was visible on response—and one error on Task 4, and two errors on Task 5—a total of three errors when the target was not visible on response. Participant 13's WAB-R and CLQT were all above the sample's mean, except for the CLQT Visuospatial Domain (77). Notably, participant 13 was the only participant with a CLQT Visuospatial Domain score below the mean to complete Task 1 without making an error. Here we can see a distinction between the performance of two participants who made far more matching task than counting task errors. While both participant 13 and participant 8 had low Visuospatial Domain CLQT scores and greater difficulty with the matching tasks than the counting tasks, participant 8 had lower scores on a larger number of aphasia assessment domains and made more than twice as many matching task errors as participant 13.

Three participants (2, 11, and 15) made fewer than ten errors on the matching tasks and made more than five more counting errors than matching errors. Participant 2 made one matching mistake, on Task 4 (target 9, to which they presented a response array of 8), and nine counting errors, seven of which they made while attempting to count backwards. Participant 2 had among the highest scores on every WAB-R scale and CLQT domain. By contrast, participant 11 and 15 made seven and nine matching task errors, respectively. All their errors were made when the target was not visible during response: participant 11 made five errors on Task 4 and two errors on Task 5, while participant 15 made four errors on Task 4 and five errors on Task 5. They were both among the worst performers on the counting tasks, with participant 11 making 31 counting errors (out of 45) and participant 15 making 45 counting errors. Both participants presented generally similar profiles on the WAB-R and CLQT. They both had WAB-R AQ and subscale scores that were below the sample mean, but CLQT Composite Severity Scores and CLQT Attention, Executive Function, and Visuospatial Domain scores that were at or above the mean. Importantly, both participants had CLQT Memory and Language Domain scores below the sample mean—participant 11 scoring 76 on the Memory Domain and 8.5 on the Language Domain, and participant 15 scoring 80 on the Memory Domain and 4 on the Language Domain. They were the only two participants to make multiple errors exclusively on Tasks 4 and 5. They were also the only two participants to only score below the mean exclusively on the CLQT Memory and Language Domains.

In brief, these three participants—who had a relatively more difficult time with the counting tasks than the matching tasks—erred exclusively on matching Tasks 4 and 5 and had above average aphasia assessment scores except on the Language and Memory Domain of the CLQT and on the WAB-R. Difficulty counting, performance on aphasia assessments that draw on language and memory abilities or require verbal responses, and singular difficulty with matching Tasks 4 and 5 appear to go hand-in-hand.

Participant counting ability and matching task performance. The skill of counting backwards is not intuitively involved in successful performance of the matching tasks and the Confrontation Naming counting task draws on substantially different skills from the other counting tasks. To develop a clearer picture of the impact of a participant's ability to count—that is, to correctly name the quantity of items in a set by assigning each item in that set an identity in turn, whereby the last identity assigned is equal to the quantity of items in the set—on the participant's ability to non-symbolically represent exact quantity, it is necessary to isolate those tasks that most closely resemble the counting activity presumably involved in the matching tasks.

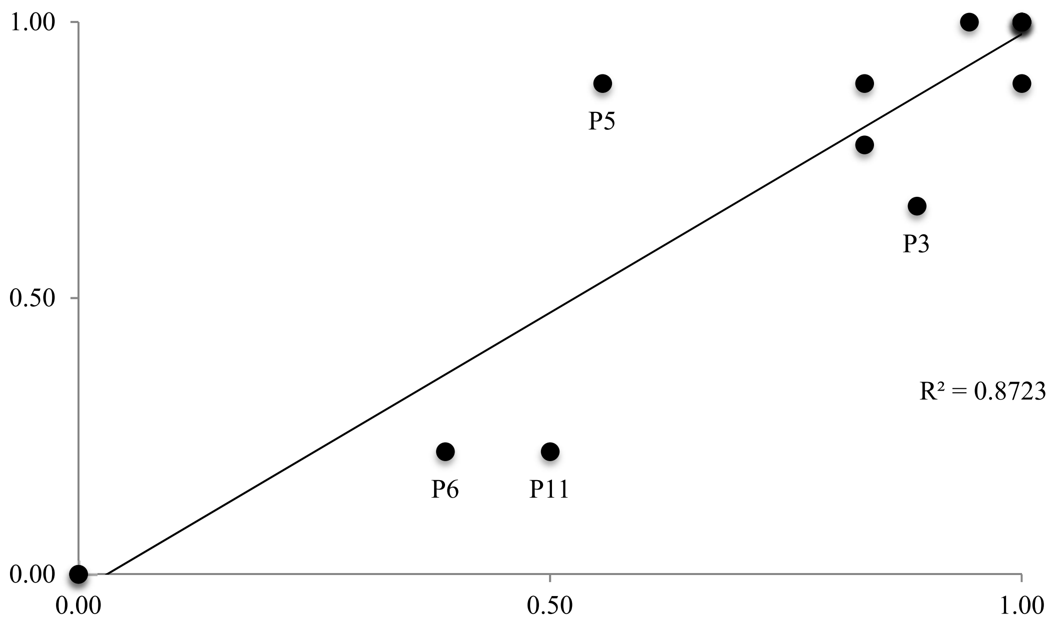

Figure 8: Participant counting task performance (x) and confrontation naming task performance (y), expressed as percentages of correct responses for targets 4-12.

When the descending Number Elicitation, descending Count List Recitation, and Confrontation Naming tasks are removed from the set of counting tasks, an interesting picture emerges (Figure 8). A large group of participants could reliably recite the count list, match the count list to objects in an array, and name Arabic numerals presented on flashcards (participants 1, 2, 4, 7, 8, 9, 10, 12, 13, and 16), while a pair of participants could not reliably perform any of those tasks (participants 14, 15). Then there were four participants who did not fall neatly into either of these categories. Participants 3, 6, and 11 were somewhat worse at confrontation naming than what was predicted by their counting abilities, but while participant 3 profiled similarly to participant 11 with regard to matching tasks—greater percentage of errors on Tasks 4 and 5, where the target is not visible on response—and CLQT domain scores—below average

Language and Memory domain scores, above average Visuospatial Domain scores—participant 6 made more matching task errors when the target was visible on response (Tasks 1, 2, and 3) and their CLQT domain scores were the inverse of participants 3 and 11 (higher than average Language and Memory domain scores, lower than average Visuospatial domain score). What all three participants have in common are their below average AQ, Repetition, and Naming and Word Finding subscale scores on the WAB-R. On the other hand, participant 5 was better at confrontation naming than was predicted by their ability to count. Participant 5 made more errors on the matching tasks than any other participant, both in total and per task. Their CLQT scores were all below average, as were most of their WAB-R scores. Interestingly, they had scores above the mean for the Repetition and Naming and Word Finding WAB-R subscales, an inversion of the performance of participants 3, 6, and 11. In other words, despite their difficulties with the matching tasks, counting tasks, and most of the aphasia assessments, participant 5 outperformed expectations when asked to name objects and symbolic images, such as in the WAB-R Naming and Word Finding subtest and on the Arabic numeral Confrontation Naming task. Object and symbol recognition tasks appear to call on a different set of skills than the matching tasks and counting tasks.

Differentiating between matching, naming, and counting. Similar, but not identical, results were found when taking the difference between the percent of correct responses on the matching tasks from (1) the percent of correct responses on the ascending Number Elicitation and Count List Recitation tasks and (2) the percent of correct responses on the Confrontation Naming task (Figure 9). Taking this calculation allows us to distinguish between participants to performed relatively well or poorly when "matching" (performance on the matching tasks), "naming" (performance on the Confrontation Naming task), and "counting" (performance on the ascending Number Elicitation and Count List Recitation tasks). Overall, ten participants named, counted, and matched about equally well—defined here as their percent of correct responses in the above-named tasks all being within 15% of each other. This suggests that, for most participants, their performance on one set of tasks could be anticipated by their performance on the other grouping of tasks.

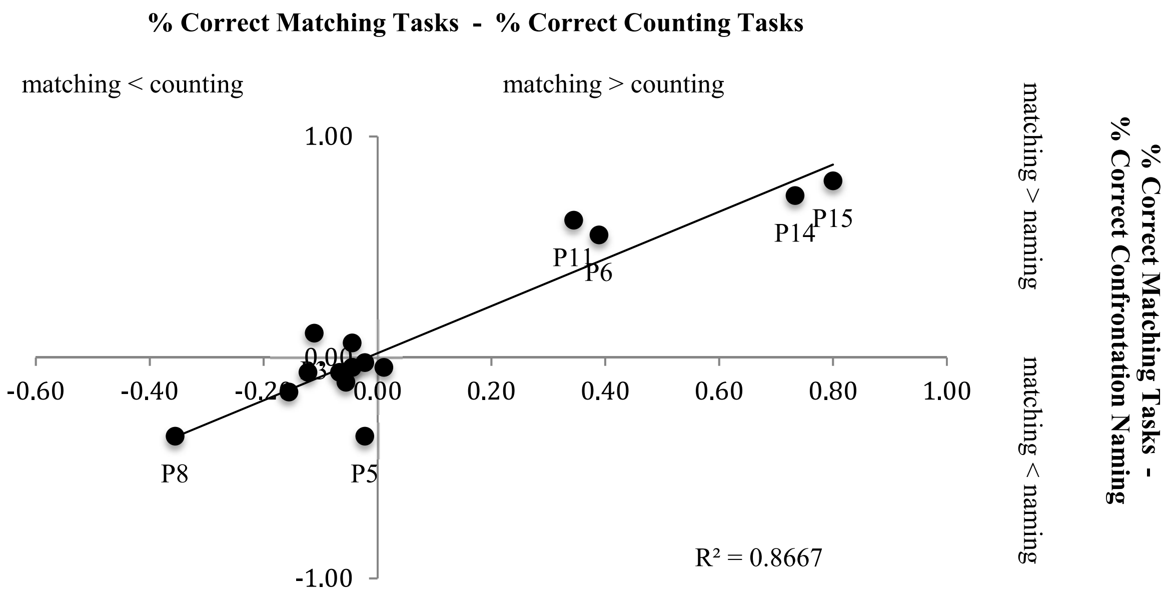

Figure 9: Participant difference scores between matching task performance and counting task performance (x) and matching task performance and confrontation naming task performance (y), expressed as percentages of correct responses for targets 4-12.

Meanwhile, Participants 14 and 15 matched far better than they counted or named but, as we saw above, they were unable to answer correctly on any trial of the counting tasks. That said, participants 14 and 15 present important contrasts. Participant 14 made twelve errors across all tasks, including two errors on Task 1 and two errors on Task 3, for a total of four errors when the target was visible during response. Participant 15 made nine errors, but only on Tasks 4 and 5, where the target was not visible during response. Both participants had similarly poor scores across the WAB-R, but participant 15 only produced CLQT domain scores below the mean for the Language and Memory domains.

Participants 6 and 11, also substantially better at the matching tasks than on the counting or naming tasks, reflect similar trends. Participant 11, like participant 15, only made errors on Tasks 4 and 5, and had below the mean scores on the CLQT Language domain and Memory domain, while participant 6, like participant 14, made errors across the matching tasks and had a CLQT Visuospatial domain score below the mean. Interestingly, participant 6, unlike participant 14, had CLQT Language domain and Memory domain scores above the sample mean, which might help explain how their matching errors were distributed: participant 6 made four total errors on Tasks 4 and 5 and six total errors on Tasks 1, 2, and 3, while participant 14 had nearly reversed results: four errors on Tasks 1, 2, and 3, and eight errors on Tasks 4 and 5.

Both participants 5 and 8 were better counters than they were matchers, although participant 8 was perfect at counting and naming, while participant 5 was better at naming than they were at counting. Participant 8's counting performance is difficult to account for against the backdrop of their matching task errors and performance on the WAB-R and CLQT. Participant 5, meanwhile, was unique in performing poorly on the CLQT, but at or above the mean on the WAB-R subscales.

Generally, it appears that there are more meaningful relationships between performance on the matching tasks and CLQT domain scores than there is among counting, naming, and matching. Participants 1, 2, 7, 10, 12, and 16 made one or no errors on each of the matching tasks, had aphasia battery assessments that were above the mean, and had little difficulty with the counting or naming tasks. Participants 11 and 15 made multiple errors on Tasks 4 and 5 only, where the target was not visible during response, struggled with counting and naming, and had CLQT domain scores that were below the mean only on the Language domain and Memory domain. Meanwhile, participants 3 and 9 could count and name, but made multiple errors and the majority of their errors on matching Tasks 4 and 5 and had low scores on the CLQT Language domain and Memory domain. Participant 13, with above average scores on the CLQT Language and Memory domains and perfect counting and matching scores, did well on Tasks 4 and 5, but made three errors on Task 3 and had a low CLQT Visuospatial domain score. Nearly the same can be said of participant 6, except that they did not perform well on counting and naming tasks. The remaining participants (4, 5, 8, and 14) made the greatest number of errors across all matching tasks, had consistently low aphasia assessment scores—especially on the CLQT Language, Memory, and Visuospatial domains—but included a mix of good and bad performances on the counting and naming tasks. Visuospatial skills are clearly related to performance on Tasks 1 and 3, while language and memory abilities are related to performance on Tasks 4 and 5.

Discussion

The task of this section is several-fold. First, I will review the results of and relationships found between the non-symbolic matching tasks, aphasia assessments, and counting tasks that stand out as noteworthy and provide plausible explanations for them. Then, I will describe the present results in the context of findings within previous studies using the same matching tasks with two other populations: the Amazonian Pirahã culture and English-speaking college students under verbal interference. Finally, I will suggest ways that the findings may add to the literature concerning human numeracy, especially with respect to how language and other domains of thought interact.

Review of Results

Of note, every participant made at least one error across all tasks, half of the participants made more than 8 errors, and multiple participants made more than 15 errors. Meanwhile, control groups of English-speakers performed at ceiling with these tasks (Frank et al., 2012). In general, participants with aphasia consistently (1) made more errors for larger target quantities, (2) made errors of greater magnitude for larger target quantities, and (3) had more difficulty with tasks where targets were not visible during response. There was consistency among those participants with the greatest overall task impairments, that is to say that participants who made ten or more incorrect responses also made errors across Tasks 1, 3, 4, and 5—their poor performance was not limited to a single task, but spread across them. These participants also tended to have lower CLQT composite severity ratings. Participant 15 is an exception to these tendencies—they made 9 errors, but *only* on Tasks 4 and 5, and had a median CLQT composite severity rating for our participants. Conversely, participants who made fewer errors had higher

CLQT composite severity ratings and generally only erred on Tasks 4 and/or 5. This suggests that participants with more severe language impairments are more challenged by the matching tasks than participants with milder language impairments. This might also mean that Tasks 4 and 5 are generally more difficult for participants, and/or that Tasks 4 and 5 are more sensitive to mild language impairments than the first three tasks.

Looking at the individual tasks, we see that eight participants responded incorrectly to at least one trial of Task 1, where the target remained visible and did not require conservation in space or time, nor, presumably, counting: producing a correct responding only required participants to match one object to an object presented to them in the array. Participants who performed poorly on this task tended to also perform poorly on Tasks 4 and 5, and have lower CLQT composite severity ratings driven primarily by poorer performance on the Executive Function and Visuospatial Skills domains and the Clock Drawing subtest (which does not load onto any CLQT domain). It appears that poorer performance on Task 1 would predict poor performance on later tasks, and only participant 6 did not follow this pattern. They made four errors on Task 1, but only two errors on each of Tasks 3, 4, and 5. However, they did fit the pattern with below average CLQT domain scores on Attention, Executive Function, and Visuospatial Skills and above average Language and Memory domain scores. Overall, these results suggest that participants who are challenged by Task 1 have more severe impairments in the visuospatial and language domains.

The results of Task 1 stand in stark contrast to near-ceiling results on Task 2. In Task 2, targets were presented in groups of 2 and 3. This is the only difference between Tasks 1 and 2, suggesting that many participants subitized the visible targets on Task 2 in order to answer accurately, but were unable to do so consistently on Task 1. One possible explanation for this result is that subitizing may be functionally independent from systems that reflect moderate visuospatial and executive impairments. However, near-ceiling performance on Task 2 also suggests that severe perceptual and/or attentional impairments (e.g., field cuts, neglect) do not explain poor performance on Tasks 1, 3, 4, and 5; this represents an important control condition in a stroke population with expected neurological and behavioral heterogeneity. Even the most severely impaired participants (by any measure) completed the task accurately.

Surprisingly, performance on Task 3 was superior to performance on Task 1, despite the required spatial translation between the perpendicular target array and horizontal response. As in Task 1, there was a tendency for participants who erred on Task 3 to also respond incorrectly on other the other tasks (with the exception of Task 2) and had lower Visuospatial domain scores on the CLQT. And, as we saw in Task 1, these results suggest that participants who have difficulty with Task 3 have more severe language impairments, but that their impairments may be more related to visuospatial, rather than verbal, abilities. The possibility that finding Tasks 1 and 3 challenging for visuospatial reasons, rather than because of verbal language impairments, might also help to account for the discrepancy between Pirahã performance on Task 3, where participants responded correctly on only 36% of trials (more in line with their performance on Tasks 4 and 5), and the performances of aphasia participants and English speakers under verbal interference, whose performance on Task 3 was more in line with their performance on Tasks 1 and 2. It may be the case—and I will explore this possibility in the next section—that there are multiple factors driving divergent performance on these tasks.

While participant accuracy was 83%, 98%, and 90% on Tasks 1, 2, and 3, respectively, participant accuracy on Task 4 dropped to 70%. Task 4 is also the first task on which participants responded without the target visible while they formulated their response. Here, a new set of correlations emerges. Performance on Task 4 correlates with scores on the Memory and Language domains of the CLQT and particularly to two of the subtests that load onto both of those domains (Personal Facts and Generative Naming). Participants performed similarly on Task 5, and there were similar correlations with CLQT scores: the strongest correlations were with the Memory and Language domains and the subtests that load onto both. However, as compared to Task 4, on Task 5, there were stronger correlations with WAB-R scores and with the Confrontation Naming CLQT subtest, which only loads onto the Language domain. We also find correlations here between Task 5 performance and performance on the counting tasks. This makes intuitive sense: without a visible, accurate counting and memory become more important.

Participants responded correctly on 90% of Task 1, 2, and 3 trials, where the target remained visible for comparison, matching, and recounting. Performance on Tasks 4 and 5 was poorer, as expected: both involve responding without the target array still visible. When we group tasks by whether the target is visible or not during response, a distinct pattern of correlations emerges. When targets are visible, performance on those matching tasks correlates well with the Attention, Executive Function, and Visuospatial Skills domains, and the Design Memory, Mazes, and Clock Drawing subtests on the CLQT. Meanwhile, Tasks 4 and 5, where the target isn't visible on response, correlate with the Language and Memory domains, the three subtests that load onto both those domains, and the Confrontation Naming subtest on the CLQT. These tasks also correlate with WAB-R AQ and subtest scores, and better with performance on the counting tasks.

The counting tasks, meanwhile, correlate exclusively with the CLQT Language and Memory domains, and all of the WAB-R scores with the exception of the Auditory Verbal Comprehension subtest. Looking closer, there's a distinctly verbal component to these correlations. All of the counting tasks require a verbal response, and the WAB-R subtests and CLQT domains that correlate with them are also the only ones that require a spoken response. That distinction stands out starkly when the CLQT subtests are grouped according to whether they require a spoken or non-spoken response (see Table 10).

Keeping in mind that none of the matching tasks require a spoken response, it is worth emphasizing that difficulties in producing a spoken response on an aphasia assessment predicts difficulties with non-symbolic exact quantity matching tasks only when the target is not visible during response. While it might be less surprising to see that performance on CLQT subtests that do not require a spoken response correlate with performance on matching tasks where the target is visible during response, it is noteworthy that performance on non-spoken CLQT subtests do not reliably predict performance on any task except Task 1. These results are remarkable because they suggest that the ability to speak is related to the ability to represent non-symbolic exact quantity (an activity that does not appear to involve speech), but only when the quantity being represented is invisible. The intuitive connection between speech and exact quantity recognition and representation is counting—that is, the articulation of the count list in one-to-one correspondence to a set of objects where the last number matched to an object stands for the number of objects in the group and can be held in memory. Meanwhile, tasks where the target are visible during response, such as Task 1, might rely less on memory and counting, and performance on these types of tasks would therefore be less sensitive to impairments in speech.

A picture that emerges from these data could be summarized as follows: for people with aphasia, non-symbolic exact quantity representation is more difficult than for people without aphasia, except when the target quantities are presented in subitizable groups and remain visible during response. They have a more difficult time representing larger quantities than smaller quantities, and struggle when the target is not visible when the participant represents it. For those who have difficulty when the target is visible during response, it is likely that they also will have difficulty with tasks when the target is not visible during response, but it is not the case that those who have difficulty with tasks where the target is not visible during response will also err on tasks where it is. Participants who had difficulty with non-verbal aphasia assessment subtests are more likely to err on non-verbal quantity representation tasks where the target is visible during response, while participants who had difficulty with aphasia assessment subtests that require verbal responses are more likely to err on non-verbal quantity representation tasks where the target is not visible during response.

Taking this one step further, four groups of participants with aphasia can be described post hoc with respect to their impairments: (1) those who have only mild impairments in their access and use of language, mild visuospatial impairments, and who have little difficulty with recognizing and representing quantities (henceforth the “mild verbal / mild visuospatial” impairment group); (2) those who have mild language impairments, but more severe visuospatial impairments, resulting in relatively more errors on tasks where the target is visible on response than when it isn’t (the “mild verbal / severe visuospatial” impairment group); (3) those who have language-specific impairments in verbal expression, counting, but relatively better visuospatial abilities, who have greater difficulties in representing exact quantities when they have to count and mentally retain the name of the quantity (the “severe verbal / mild visuospatial” impairment group) ; and (4) those who have both significant verbal and non-verbal impairments related to their brain injury and have difficulty representing exact quantities except when presented in subitizable groups that remain visible during representation (the “severe verbal / severe visuospatial” impairment group) (Figure 10).

	mild verbal / mild visuospatial	mild verbal / severe visuospatial	severe verbal / mild visuospatial	severe verbal / severe visuospatial
Participants	1, 2, 7, 10, 12, 16	6, 13	3, 9, 11, 15	4, 5, 8, 14
WAB-R AQ	87.78	68.95	25.55	37.10
CLQT spoken (Language domain)	27.50	22.50	5.13	5.25
CLQT non-spoken (Visuospatial domain)	92.33	64.50	91.75	70.00
Matching tasks 2 (target visible and subitizable)	100%	94.4%	100%	94.4%
Matching tasks 1 & 3 (target visible on response)	99.0%	75.0%	91.7%	68.1%
Matching tasks 4, 5 (target not visible on response)	89.8%	80.6%	58.3%	48.6%
Number elicitation	98.6%	66.7%	50.0%	50.0%
Count list recitation	95.8%	72.5%	51.2%	67.5%
Confrontation naming task	94.2%	60.0%	30.0%	63.8%

Figure 10: Visualization of study participants broken into clusters. The "mild verbal / mild visuospatial" group is characterized by relatively high aphasia assessment scores with good matching and counting task performance. The "mild verbal / severe visuospatial" group is characterized by higher than average aphasia assessment scores on spoken-language-related domains but below average visuospatial scores, and making relatively more of their matching task errors on Tasks 1 and 3. The "severe verbal / mild visuospatial" group is characterized by lower than average aphasia assessment scores on spoken-language-related domains but higher than average visuospatial scores, difficulty with counting, and making most of their matching task errors on Tasks 4 and 5. The "severe verbal / severe visuospatial" group is characterized by generally lower aphasia assessment scores and difficulty across counting and matching task performance. WAB-R and CLQT data are reported as raw scores. Matching and counting task data are reported as percentage of correct responses on the task or tasks. Darker shading indicates poorer performance.

Members of the mild verbal / mild visuospatial impairment group would show up in this sample as participants with higher than average aphasia assessment scores, be good "counters" and "namers," and perform well on the matching tasks, making infrequent errors on these tasks and only with higher targets on Tasks 4 and 5. Participants 1, 2, 7, 10, 12, and 16 fit this description. Members of the mild verbal / severe visuospatial impairment group would have *higher than average* aphasia assessment scores on the WAB-R (where all the subtests require verbal responses except Auditory Verbal Comprehension) and on the CLQT Language and Memory Domains, but *below* average CLQT Visuospatial Domain scores, and make relatively more of their matching task errors on Tasks 1 and 3. Only participants 6 and 13 fit this category. Meanwhile, members of the severe verbal / mild visuospatial impairment group would have *lower than average* aphasia assessment scores on the WAB-R and on the CLQT Language and Memory Domains, but *above average* Visuospatial Domain scores. They would struggle with counting and naming, and make the majority or all of their matching task errors on Tasks 4 and 5. Participants 3, 9, 11, and 15 fall neatly into this category. Members of the severe verbal / severe visuospatial impairment group would have low scores on the aphasia assessments, especially all three of the CLQT Language, Memory, and Visuospatial Domains, struggle with counting and naming, and make errors across matching tasks. Participants 4, 5, 8, and 14 fit here.

The relationship between matching task performance, aphasia assessment measures, and counting and number naming task performance can be described as follows. Because Task 1 performance correlated most significantly with CLQT subtests that called for a non-verbal response and, in particular, the CLQT Visuospatial Domain, it seems reasonable to conclude that poor performance on Task 1 is best predicted by poor visuospatial abilities, more so than the ability to count, or any language and memory impairments. That said, since all participants who were challenged by Task 1 and had low CLQT Visuospatial scores also had low CLQT Language and Memory Domain scores, low WAB-R scores, and were challenged by Tasks 4 and 5, it isn't possible to say that counting and language and memory abilities don't play any role in the capacity to do one-to-one matching tasks like Task 1. It is possible, however, to see that performance on Tasks 4 and 5, where the target is not visible on response, is tightly connected to counting, language, and memory capacities. Performance on Tasks 4 and 5 correlate strongly with CLQT subtests that call for a verbal response, the CLQT Language and Memory Domains, and all of the WAB-R scales that require verbal responses. In turn, those CLQT Domains, WAB-R scales, and performance on Task 5 correlate strongly with counting task performance. Examining individual cases bears this description out.

Perfect performance on Tasks 4 and 5 requires participants to be able to count—that is, to be able to recite the names for quantities in the correct order, assign each of those names to one and only one object in an array, and keep that association in mind long enough to be able to name the set as containing some certain number of objects—but counting is not sufficient for perfect performance on Tasks 4 and 5. The tasks require a second counting performance when reproducing the target array. However, without the capacity to count, participants are able to provide correct or nearly correct answers by estimating, but are unable to perform the task perfectly, as we might ordinarily expect.

Accurate performance on Tasks 1 and 3 would also certainly benefit from the capacity to count, but it doesn't appear to be necessary to perform the same set of operations outlined above in order to perform them well—participants need only to check that they have correctly corresponded the target array to the response array, they do not need to name individuals, the set, or retain that name. An impairment in visuospatial matching ability, however, would either make it very difficult to perform Tasks 1 and 3, or force participants to rely on their abilities to count in order to compensate for their visuospatial difficulties. If those counting abilities are also impaired—as they often were in our sample—then performance on Tasks 1 and 3 would be further compromised. Again, even with all these challenges, participants would be expected to produce correct or nearly correct responses simply through estimation.

The one task where neither counting nor visuospatial abilities do not appear to play a role is in Task 2, where participants are able to give correct responses because the array is broken into subitizable groups. They do not have to count or estimate in order to provide a correct response. This is borne out by the difference in participant performance on Task 2 compared to other tasks.

Participants' performance lend support to established ways of thinking about number, thought, and language. According to the model put forth by Feigenson, Dehaene, and Spelke (2004), we are born with two systems for the cognitive representation of number—a parallel-individuation system that can track up to three or four discrete objects (subitizing) and an analog magnitude estimation system we use to approximate large quantities. This can be seen in the participants' performance on Task 2, and in those cases where impairments in language, memory, and visuospatial abilities appear to require participants to estimate their responses. While these cognitive systems are also found in other animals, humans appear to use exact number words as tools that enhance our capacity to do things with quantities by bridging these systems. In other words, counting makes it possible for participants to perform Tasks 1 and 3 more easily, although without counting language, it is possible to perform those tasks well by relying on visuospatial matching abilities instead of counting—or vice versa: counting can compensate for visuospatial impairments. At the same time, without the capacity to employ counting language effectively, Tasks 4 and 5 are impossible to perform better than would be expected through estimation.

A Bridge to Other Studies and Broader Claims

From here, a bridge can be built to the results of other studies that have used these same tasks with other populations. The results of this study mirror those of previous studies with the Pirahã (Everett & Madora, 2012; Frank et al., 2008; and Gordon, 2004) and adult English-speakers under verbal interference (Frank et al., 2012). While our aphasia participants and the English-speakers under verbal interference produced nearly identical results in accuracy, error magnitude, and CoV within and across tasks, the Pirahã results reflected similar patterns to those groups but revealed made more frequent and larger errors and higher CoVs within and across tasks (see Figures 6 and 7). And while all three groups did well on Task 2, we can see that the aphasia participants' accuracy was closer to the Pirahã performance than the English-speakers' on Task 1, but more like the English-speakers' under verbal interference on Tasks 3, 4, and 5.

However, we can make a few useful comparisons to verbal interference study participants and the Pirahã study participants if we split the aphasia participants into the four groups suggested above—a mild verbal / mild visuospatial impairment group (those who are able to perform the matching tasks well, are good counters, and have milder aphasia impairments), a mild verbal / severe visuospatial impairment group (those who are relatively more challenged by Tasks 1 and 3, and whose impairments show up primarily in the visuospatial domain and not in the domains of language and memory), a severe verbal / mild visuospatial impairment group (those who struggle with matching Tasks 4 and 5, are poor counters, and have aphasia impairments that are primarily limited to the domains of language and memory), and a severe verbal / severe visuospatial impairment group (those who are challenged by all the matching tasks—except the subitizing involved in Task 2—are poor counters, and have aphasia impairments that include not only language and memory domains, but also visuospatial domains). For instance, the accuracy of English-speakers under verbal interference dropped below 85% correct and CoV rose past 0.05 for Tasks 4 and 5, like the severe verbal / mild visuospatial impairment group, who had difficulty only with counting and language and memory domain measures. Meanwhile, the Pirahã were only accurate on more than 85% of trials or had a CoV below 0.05 for Task 2, like the more severely impaired aphasia participants in the severe verbal / severe visuospatial impairment group. In other words, the severe verbal / mild visuospatial impairment group performed more like the English-speakers under verbal interference, while the Pirahã and the severe verbal / severe visuospatial impairment group more closely resemble each other (Figure 11).

While it isn't hard to imagine that the aphasia participants and English-speakers under verbal interference would perform similarly, the connection between aphasia participants and an indigenous Amazonian culture without concepts of exact quantity might be harder to accept. Studies such as Yoon et al. (2014) have suggested that the Pirahã organize the visuospatial field

	Verbal Interference	Aphasia Participant Groups			Pirahã
		severe verbal / mild visuospatial	mild verbal / severe visuospatial	severe verbal / severe visuospatial	
Challenged by Task 1			✓	✓	✓
Challenged by Task 2					
Challenged by Task 5	✓	✓		✓	✓

Figure 11: Visualization comparing matching task performance of Pirahã study participants, English-speakers under verbal interference, and current study participants.

differently than adults from Western cultures. If true, and their lack of counting language makes counting impossible, it seems reasonable the Pirahã and aphasia participants with impairments in language, memory, and visuospatial domains would perform similarly on these tasks.

The studies involving the Pirahã clearly suggest a reliance on analog magnitude estimation in attempting to represent target quantities. Of all the research of this kind conducted with the Pirahã, only the one-to-one matching task in Frank et al. (2008) produced a CoV markedly different from 0.15. Everett and Madora (2012) offered a speculative explanation: unlike the others, the village tested in Frank et al. (2008) had been exposed to math tutoring that included neologisms for number words. It is the neologisms for number words that are exceptional—all the villages had been exposed to the one-to-one matching task and other attempts at basic math training by the Brazilian government, but only the site of Frank et al. (2008) had been exposed to number word neologisms. The authors are clear that this is speculation on their part, but it dovetails with a possible explanation as to the task performance differences between the Pirahã on the one hand and the verbal interference and aphasia participants on the other. In attempting to account for the lower CoVs and greater accuracy of the verbal interference participants, Frank et al. (2012) suggests that participants' "differential cultural experience with mathematics and other uses of exact numerosity led to their relatively more precise representation of analog magnitude" (p. 82). The same could be suggested of the aphasia participants in this study—like the English speakers under verbal interference, their cultural experience with the concept of exact quantity (and the belief in its importance), would lead to more precise analog magnitude estimation.

Certainly, there are differences between the current population of people with aphasia, people of an anumeric culture, and English-speakers under verbal interference. What separates the Pirahã from other populations under discussion here is that they exist in a world without exact-quantity language and may not have a concept of number to access. English speakers under verbal interference, meanwhile, are members of a numeric culture who have had their ability to use language temporarily disrupted, and people who have aphasia are members of the same culture with a more permanent disruption. Also, an aphasia population consists of individuals with distinct lesions, resulting in a range of verbal and nonverbal impairments and significant heterogeneity is to be expected, compared to a population of English speakers undergoing experimental manipulation via verbal interference. While diversity within the current aphasia population is viewed as a potentially rich source for identifying particular aspects of language (e.g., comprehension, speech) that may uniquely affect particular aspects of number use (e.g., mental representation of exact quantity, counting), it also suggests caution before drawing definitive conclusions based on group performance.

With that caveat made, the evidence from this study suggests that visuospatial ability correlates with the capacity to reliably recognize and reproduce exact quantity above what would be expected by estimation when the target quantity is presented in a way that is not subitizable and *remains visible* for the participant to use for reference. Here, counting may compensate for visuospatial impairments, but doesn't appear necessary for success. Aphasia participants who are challenged by this task have difficulties with visuospatial, language, and memory domains. The Pirahã struggle with this task not only because they lack counting language, but perhaps also because they organize the perceptual field differently than those from industrialized cultures do. English-speakers under verbal interference, meanwhile, can rely on visuospatial abilities to compensate for their impaired counting in order to complete these tasks successfully. All three groups, across multiple studies, struggle when asked to perform tasks where the target is not visible while responding, which strongly suggests that counting language and the ability to accurately count are necessarily skills in order to recognize and reproduce exact quantities with greater accuracy than expected by analog magnitude estimation.

This study helps to add evidence to hypotheses about the significant interaction between counting language and the capacity to recognize and represent exact quantity. In addition to examining a previously unexamined population this study also helps to add texture and nuance to these hypotheses by looking closely at the role of underlying language, memory, visuospatial and other cognitive processes involved in counting, exact quantity recognition, and representation.

Language impairments that limit an individual's ability to cognitively represent exact quantities impact an individual's ability to recognize and represent those quantities in non-symbolic, non-linguistic ways. We have returned to the linguistic relativity hypothesis that language influences non-linguistic perception, thought, and action. To follow the "language as tool-kit" metaphor for how that language influences what we can think about, how we can think about it, and how we act in and upon the world: words for exact numbers are tools that allow us to perceive and manipulate the world in certain ways. With them, we can perform complicated feats of engineering, economics, and mathematics. Without those tools, however, we have a difficult time doing matching tasks that seem so "basic" to those of us who've grown up in a world full of exact quantities that we have a hard time believing that they require tools at all.

Limitations and Directions for Future Research

Between the current study, the studies that have used the same tasks with different populations, and developmental research that has pointed to similar conclusions, there is mounting evidence that counting language plays an important role in the recognition and use of exact quantities larger than four. In situations where exact number language does not exist within a culture, has not yet been developed, or is temporarily or more permanently interfered with in individuals, human beings struggle not only with tasks that involve counting, but also tasks that involve exact matching, which would not appear to require counting or number language. This is evidence of the linguistic relativity hypothesis: access to (or lack of access to) certain language tools enables and restrains what humans can think about and do in the world.

As with any study, there are questions that have remained unanswered and which future studies might effectively take up. For instance, the Pirahã studies relied on approximately the same number of participants for their findings, but the studies with English-speakers under verbal interference used twice as many participants—it remains to be determined whether with 35 participants, this study's results (or the Pirahã studies' results) would come to resemble those from English-speakers under verbal interference. In the case of the Pirahã, the sample size is limited by the population size, but in the case of this study, many more participants could be studied if there were time and access. To take another example of where data points might provide more information: data from the Semantic Category Probe task and Pyramids and Palm Trees assessment might have yielded greater insights, given the correlations between those assessments and performance on matching tasks, counting tasks, and aphasia assessments. Unfortunately, only half of the sixteen participants completed these tasks. Adding those tasks to the standard battery of assessments may provide another avenue to describing more closely the impact different language impairments have on exact quantity representation tasks.

In a related vein, it would prove worthwhile to conduct additional tasks with the same population. Recent research in developmental psychology have developed tasks that ask children to represent exact number concepts but do not involve counting (Izard, Streri, and Spelke, 2014; Jara-Ettinger, Piantadosi, Spelke, Levi, and Gibson, 2016). Asking people with aphasia to perform these kinds of tasks might help us to better understand a basic question about whether counting is necessary for Tasks 1, 2, and 3, further delineating the necessary and sufficient conditions by which it is possible for a human being to recognize and use exact quantities greater than 4. To complement any potential results from such developmental tasks used with an aphasia population, it may be potentially illuminating to do the inverse: asking children to perform the matching and counting tasks from the current study. Children in the process of learning the count list and/or how to identify exact quantities of objects could help refine or revise notions of how necessary counting language is to exact quantity recognition and representation.

It is also worth noting that the aphasia participants have been treated as a homogenous group, differentiated by measures of severity and impairments in certain domains, rather than by the type of aphasia that they have been diagnosed with. The reasoning here is that in the clinical diagnosis of aphasia in these cases was based upon participant performance on these assessments, rather than by lesion location. In other words, with this group of participants, to sort them by general severity and their relative domain impairments is to do the same thing as to break them up into groups labeled “Broca’s,” “Conduction,” “Wernicke’s,” and “Global.” That said, while diagnosis based on observable behaviors is both expedient and more clinically useful, it may obscure some of what we could come to know about our participant’s impairments and their effects with access to brain scans and map participants' lesions relative to one another. Future studies with the capacity to acquire and utilize participants' brain scans would provide a different, neuroanatomical approach to understanding the way that language impairment impacts exact quantity representation.

The present study also suggests that experiments involving people with aphasia may serve to further refine an understanding of how language and thought interact. In the last 15 years, a great deal of interesting empirical research has emerged that has helped to illuminate and refine aspects of the linguistic relativity hypothesis. To my knowledge, this study is one of few to date that employs participants with aphasia to investigate linguistic relativity. Other domains of interest to linguistic relativity—such as spatial and temporal perception—could be explored in an aphasia population and could supplement research that has already been done in these areas. Paluy, Gilbert, Baldo, Dronkers, and Ivry (2011) ventured into this territory by examining color discrimination abilities in people with aphasia using tasks like those in Winawer et al. (2007). Their studied aimed to test a "lateralized Whorfian effect" seen in reaction time differences when color discrimination targets are presented to the right or side of the visual field. Much as was the case with verbal interference in Winawer et al. (2007), participants in this study with aphasia due to left hemisphere stroke did not demonstrate the performance difference found in people who had suffered right-hemisphere strokes and in neurotypical controls. Would aphasia function similarly to verbal interference in studies exploring other cognitive domains? More research could teach us more about both aphasia and the linguistic relativity hypothesis. It is possible, too, that new areas of interest to the linguistic relativity hypothesis might be opened by exploring the relationship between language, thought, and action in people who have aphasia. The horizon of study for the linguistic relativity hypothesis might be broadened by examining those areas of daily life where people who have aphasia struggle and considering the possibility that language impairment is a factor in these activities that do not appear to involve language.

Finally, this study has been focused on questions of relatively abstract questions of the relationships between language, cognition, and behavior in human experience, broadly considered. This approach has ignored the day-to-day challenges of many of the study's participants, people who struggle to count and recognize exact quantities in a culture where counting and identifying exact quantities is a necessary activity of daily life. Their difficulties are no doubt compounded by the frustration of being unable to, or having difficulty with, tasks that were likely so routine as to be taken for granted prior to the onset of aphasia. It would be not only of academic interest, but moreover clinically valuable, to explore the rehabilitative possibility of re-learning counting language as an avenue to improved exact quantity recognition and representation. Future studies with the same or similar participants might be able to not only begin to address questions of the impact of re-learning counting language on non-symbolic exact quantity task performance, thereby helping to refine the linguistic relativity hypothesis, but also may light the way to effective strategies for helping people with aphasia be able to regain some of the functioning they have lost and improve their overall quality of life.

REFERENCES

Condry, K.F., & Spelke, E.S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*(1), 22-38.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20(3), 487-506.

Dragoy, O., Akinina, Y., & Dronkers, N. (2016). Toward a functional neuroanatomy of semantic aphasia: A history and ten new cases, *Cortex*, 1-19.

Everett, C. (2013). *Linguistic relativity: Evidence across languages and cognitive domains*. Berlin: De Gruyter Mouton.

Everett, C., & Madora, K. (2012). Quantity recognition among speakers of an anumeric language. *Cognitive Science*, *36*(1), 130-141.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307-314.

Frank, M.C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, *64*(1-2), 74-92.

Frank, M.C., Everett, D.L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819-824.

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

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press.

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, *306*(5695), 496-499.

Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test*. New York, NY: Pearson.

Howard, D., Patterson, K.E. (1992). *Pyramids and palm trees: A test of semantic access from pictures and words*. Bury St. Edmunds, UK: Thames Valley Test Company.

Izard, V., Streri, A., and Spelke, E. (2014). Toward exact number: Young children use one-to-one correspondence to measure set identity but not numerical equality. *Cognitive Psychology 72*, 27–53.

Izard, V., Sann, C., Spelke, E., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences USA 106*, 10382-10385.

Jara-Ettinger, J., Piantadosi, S., Spelke, E., Levy, R., & Gibson, E. (2016). Mastery of the logic of natural numbers is not the result of mastery of counting: evidence from late counters. *Developmental Science*, 1–11.

Kagan, A., & Kimelman, M.D.Z. (1995). Informed Consent in Aphasia Research: Myth or Reality. *Clinical Aphasiology 23*, 65-75.

Kertesz, A. (2006). *Western Aphasia Battery–Revised*. New York, NY: Pearson.

Lee, I.A., & Preacher, K.J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia*, *41*(14), 1942-1958.

Lucy, J. (1997). Linguistic relativity. *Annual Review of Anthropology 26*, 291-312.

McNeil, M.R. & Pratt, S.R. (2001). Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. *Aphasiology, 15*(11), 901-911.

Paluy, Gilbert, Baldo, Dronkers, and Ivry (2011). Aphasia patients exhibit a reversal of hemispheric asymmetries in categorical color discrimination. Brain and Language, 116 (3), 151-156.

Rosenbek, J.C., LaPoint, L.L., & Wertz, R.T. (1989). *Aphasia: A clinical approach*. Boston, MA: Little, Brown & Co.

Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104* (19), 7780-7785.

Whalen, J., Gallistel, C.R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130-137.

Whorf, B.L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf* (J.B. Carroll, Ed.). Cambridge, MA: MIT Press.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature 358*, 749-750.

Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition 74*, B1-B11.

Yoon, J., Witthoft, N., Winawer, J., Frank, M.C., Everett, D.L., & Gibson, E. (2014). Cultural differences in perceptual reorganization in US and Pirahã adults. *PLoS ONE 9*(11): e110225.