

Characterizing the spatial, functional, and representational properties of the ventral occipitotemporal “Number Form Areas”

By

Darren J. Yeo

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

May 31, 2021

Nashville, Tennessee

Approved:

Gavin R. Price, Ph.D.

Laurie E. Cutting, Ph.D.

Isabel Gauthier, Ph.D.

James R. Booth, Ph.D.

ACKNOWLEDGEMENTS

I am eternally indebted to my advisor, Dr. Gavin Price, for taking a chance on me 7 years ago in supporting my transition from classroom teaching to neuroscience research. Gavin has provided me with countless opportunities, advice, and encouragement for my growth as a researcher, a mentor to others, and as a person. I am also extremely appreciative of his constant reminders to pursue what I am truly passionate about and to think of the bigger picture, his unwavering efforts to have me make time for a life outside of work, and for his patience in putting up with the numerous downs I have had in completing this journey.

I am also very grateful to my committee members, Dr. Laurie Cutting, Dr. Isabel Gauthier, and Dr. James Booth, for their time and generosity in offering me guidance and feedback to make my ideas and work better, for sharing their expertise and knowledge, for accommodating the multiple changes I have had to make to my dissertation plans. I am truly humbled to be able to learn from these giants in their respective fields in and outside the committee meetings.

The daily grind would not have been as enjoyable and motivating without my fellow members of the Numerical Brain Lab (including Olive!), past and present: Eric Wilkey, who has been a dependable senior member of the lab whom I can always go to for advice about anything, and for inviting me to gatherings to help me get out of lab; Benjamin Conrad, who inspires me a lot with his capacity for learning and self-taught expertise in neuroimaging methods, programming, and computing, and for making our lives so much easier with the automated pipelines he has written; Courtney Pollack, who is always available to offer detailed feedback and thoughtful questions to help me convey my ideas better; Andrew Lynn, for his refreshing ideas outside of numerical cognition and for willing to play the much needed devil's advocate to put us on the right track; Lisa Venanzi, Adam Kaminski, Amanda Goodwin, Sharo Casto, and Sarah Wiesen for being such dependable, trustworthy, and empathetic co-workers.

I would like to thank Dr. Nina Martin, Dr. Vicki Harris, and Ally Armstead who welcomed me to the Vanderbilt community through the Child Studies Program, and have supported my research career in one way or another over the past years. I am also extremely grateful to Dr. Lisa Fazio and Dr. Bethany Rittle-Johnson who welcomed me into their labs and invited me to collaborate with them on various projects on math learning. My heartfelt thanks to Lisa in particular who mentored me on my very first multi-experiment paper and has continued to include me as part of the Building Knowledge Lab even after I switched my primary labs.

I am also like to express my gratitude to Nanyang Technological University and the Ministry of Education (Singapore) for funding my doctoral studies. My heartfelt thanks to Dr. Annabel Chen for her mentorship and support before and throughout my PhD journey.

Last but not least, I would like to thank my parents for giving me the freedom to pursue whatever I desire throughout my life, and trusting that things will turn out just fine; Pearlyn and Mervin for holding the fort at home during these couple of years; and Oli for his unconditional love and everlasting memories.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
Chapter	
1. Introduction	1
How Does the Brain Process Numerals? The “Visual Number Form” Hypothesis	2
Beyond the “Visual Number Form”: Triple-code Model of Number Processing	4
Do The “Visual Number Form” and “Visual Word Form” Systems Share Neural Resources?	7
A VNFA Can Be Reproducibly Localized, but fMRI May Not Be Optimal	10
Potential Hemispheric Asymmetries in the Recruitment of the Bilateral NFAs	16
Overview of Current Studies	18
2. The Search for the “Number Form Area”: A Functional Neuroimaging Meta-Analysis	22
Introduction	22
Methods	25
Results	41
Discussion	46
Conclusions	57
Authorship Contributions	57
3. Probing the Representational Content in the “Number Form Area”: A Representational Similarity Analysis	58
Introduction	58
Methods	63
Results	80
Discussion	87
Conclusions	97
Authorship Contributions	97

4. Probing the Hemispheric Asymmetry of Function and Representations in the Bilateral “Inferior Temporal Numeral Areas”	98
Introduction	98
Methods	105
Results	120
Discussion	131
Conclusions	140
Authorship Contributions	141
5. General Discussion	142
Summary of Findings	143
Future Directions	151
Final Remarks.....	156
Appendix	
A. Supplementary Materials for Chapter 3	158
Supplementary Methods.....	158
Supplementary Results	171
B. Supplementary Materials for Chapter 4	191
Supplementary Methods.....	191
Supplementary Results	194
REFERENCES	203

LIST OF TABLES

Table	Page
1-1. Evidence of lateralization of numeral processing in the ventral occipitotemporal (vOT) in adult fMRI, EEG and MEG studies	18
2-1. fMRI studies that contrasted Arabic numerals with other meaningful written symbols	29
2-2. Participant demographics, tasks, contrasts, foci of interest, and statistical threshold for the fMRI studies included in meta-analyses	35
2-3. ALE results for functional specialization of symbolic number processing in studies with all relevant contrasts (Meta-analysis I)	41
2-4. ALE results for functional specialization of symbolic number processing in studies with suitably specific contrasts only (Meta-analysis II)	44
3-1. Meta-analyses for the degree of similarity between each model RDM and neural RDMs in left and right ITG	82
A-1. Notable experimental and neuroimaging acquisition comparisons among datasets	158
A-2. Number of errors of omission and commission across all runs in each dataset	160
A-3. Visual properties of stimulus sets: Mean (SD)	161
A-4. Pairwise comparisons of luminance (above diagonal) and perimetric complexity (below diagonal) of stimulus categories for Digits Set, Letters Set 1, and Letters Set 2 (bold)	162
A-5. Mean degree of dataset-specific similarity between model and neural RDMs in left and right ITG	171
A-6. Relation between k -medoids cluster membership in the right ITG in Dataset 1 and the actual category membership	174
A-7. Relation between k -medoids cluster membership in the right ITG in Dataset 2 and the actual category membership	175
A-8. Relation between k -medoids cluster membership in the right ITG in Dataset 3 and the actual category membership	178
A-9. Meta-analyses for the degree of similarity between each model RDM and neural RDMs in left and right parietal lobules (PL)	180
A-10. Meta-analyses for the degree of similarity between each model RDM and neural RDMs in right premotor (PMC) and inferior frontal (IFG) regions	187
B-1. Stimulus list for a single run of digit detection and letter detection	191
B-2. Spearman's rank bivariate correlations of model RDMs (upper triangle: Digits only; lower triangle: Digits and letters combined)	193
B-3. Mean, standard deviation, and range for in-scanner (included runs only) and behavioral measures	194
B-4. Fisher's z -transformed representational similarity (RS_z) between co-occurring characters from target and non-target categories	198
B-5. Pairwise difference in the similarity between the left IT Digits-RDMs and model RDMs ($N = 32$)	200
B-6. Pairwise difference in the similarity between the right IT Digits-RDMs and model RDMs ($N = 32$)	200

LIST OF FIGURES

Figure	Page
1-1. Cohen and Dehaene's (1991) cognitive model of Arabic numeral-to-word transcoding	4
1-2. Triple-code model of number processing	7
1-3. “Number Form Area”, “Visual Word Form Area”, and “Letter Form Area”	11
2-1. Visualization of ALE results in meta-analysis I	42
2-2. Visualization of ALE results in meta-analysis II	45
3-1. Stimulus sets and a schematic of the fixation-color change-detection task	64
3-2. Regions of interest (ROIs), and neural and candidate representational dissimilarity models (RDMs)	71
3-3. Schematic of the categorical and control models	75
3-4. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-preferring regions in (a) left and (b) right inferior temporal gyrus (ITG)	81
3-5. Group-averaged representational dissimilarity matrices and representational geometry of exemplars in right ITG	86
4-1. Regions of interest (ROIs) and a schematic of the key components of the representational similarity analyses (RSA)	107
4-2. Hemispheric asymmetries of regional mean response amplitudes and their relation to calculation skills for digit detection	123
4-3. Category discriminability in the left and right IT ($N = 31$)	124
4-4. Relation between calculation skills and category discriminability	125
4-5. Exemplar discriminability in left and right IT	126
4-6. Relation between calculation skills and digit discriminability	127
4-7. Similarity between the model RDMs and Digits-RDMs of IT ROIs ($N = 32$)	130
4-8. Hemispheric asymmetry in similarity between the model RDMs and neural Digits-RDMs ($N = 32$)	131
5-1. Cognitive models of reading aloud numerals and words	153
A-2. Size (a) and temporal signal-to-noise ratio (b) in each region of interest across three datasets	165
A-3. Representational geometry of the 36 exemplars in two-dimensional space based on pixel overlap and shape distance for Letters Set 1	168
A-4. Representational geometry of the 36 exemplars in two-dimensional space based on pixel overlap and shape distance for Letters Set 2	169
A-5. Comparison of candidate representational dissimilarity models (RDMs) between Letters Set 1 (Datasets 1 and 2) and Letters Set 2 (Dataset 3)	170
A-6. k -medoids clustering of the neural representations of 36 exemplars in the numeral-preferring right ITG region in Dataset 1	174
A-7. k -medoids clustering of the neural representations of 36 exemplars in the numeral-preferring right ITG region in Dataset 2	176
A-8. k -medoids clustering of the neural representations of 36 exemplars in the numeral-preferring right ITG region in Dataset 3	178
A-9. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-preferring regions in (a) left and (b) right parietal lobule (PL)	179
A-10. Group-averaged representational geometry of the 36 exemplars in two-dimensional space	

in the numeral-preferring (a) left and (b) right parietal lobules (PL) in Datasets 1 and 2	185
A-11. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-preferring regions in the right (a) premotor cortex (PMC) and (b) inferior frontal gyrus (IFG)	186
A-12. Group-averaged representational geometry of the 36 exemplars in two-dimensional space in the numeral-preferring right inferior frontal gyrus (IFG) in Datasets 1 and 3	190
B-1. Representational geometry of the digit and letter exemplars in two-dimensional space based on phonological similarity	192
B-2. Representational geometry of the digit exemplars in two-dimensional space based on ratio	192
B-3. Representational geometry of the digit exemplars in two-dimensional space based on joint frequency	193
B-4. Representational geometry of the digit and letter exemplars in two-dimensional space based on shape	193
B-5. Hemispheric asymmetries of regional mean response amplitudes and their relation to calculation skills for letter detection	196
B-6. Relation between calculation skills and letter discriminability	199
B-7. Similarity between the alphanumeric model RDMs and Full-RDMs in the left and right IT ROIs ($N = 31$)	202
B-8. Hemispheric asymmetry in similarity between the alphanumeric model RDMs and neural Full-RDMs ($N = 31$)	202

CHAPTER 1

INTRODUCTION

Modern societies around the world use written characters as tools to visually represent spoken communication. Besides words written in script(s) used in one's native language, much of our waking lives are also governed by Arabic (or Hindu-Arabic) digits 0 to 9, such as in reading and writing time, date, phone numbers, addresses, as well as making measurements and performing arithmetic computations related to discrete quantities, length, mass, capacity, time, and money. Although numbers can be represented by characters of the script of any language (e.g., the concept of *five* can be written as “five” or “V” using the Roman script, “五” using the Chinese script, “חמשה” using the Hebrew script), the Arabic numeral system has been adopted almost universally for representing numbers, due in part to its efficiency in allowing numbers of any size to be easily represented and manipulated (J. Zhang & Norman, 1995).

The ability to read Arabic numerals and understand what they represent has critical consequences throughout lifespan. Such basic numeral understanding bridges preschooler's informal mathematical knowledge (e.g., counting, comparing collections of objects, solving verbal/story problems) to formal mathematical knowledge (e.g., algorithmic calculations) (Purpura et al., 2013), predicts later academic achievement in both math and reading (Duncan et al., 2007), and are associated with individual differences in mathematical competencies across lifespan (Schneider et al., 2017, 2018). In turn, low mathematical competencies are associated with lower likelihood of high school graduation and college attendance (Bynner & Parsons, 1997; Parsons & Bynner, 2005), lower employment opportunities and advancement, and income

(Bynner & Parsons, 1997; Parsons & Bynner, 1997, 2005; Rivera-Batiz, 1992), poorer decision-making related to personal finances and health care (Gerardi et al., 2013; Hibbard et al., 2007; E. Peters, 2012), poorer physical and psychological well-being (Parsons & Bynner, 2005), as well as huge annual economic costs to societies (Gross et al., 2009; R. Martin & Hodgson, 2014).

Despite the importance of Arabic numeral reading, we know much less about how we read numerals than we do about words. Over the past three decades, we have learned how learning to read words in a particular script transforms the perceptual pathways in our brains (for an in-depth review, see Dehaene et al., 2015). Of particular interest here is the experience-driven development of a left-lateralized neural circuit that connects early visual cortices to spoken language regions via the so-called ventral temporal “Visual Word Form Area” – the region where the brain first automatically recognizes that “READ” and “read” are the same word even though they look different (L. Cohen & Dehaene, 2004; Dehaene & Cohen, 2011; McCandliss et al., 2003). The “Visual Word Form Area” is thought to represent the orthographic knowledge of the script(s) that we are familiar with, and is a powerful biomarker of reading fluency or difficulty (Dehaene-Lambertz et al., 2018; Dehaene et al., 2010; Olulade et al., 2015; van der Mark et al., 2009, 2011). Does learning to read numerals use part of the same perceptual pathway that has been sculpted for words, or does it carve out its own pathway? In other words, do our brains develop pathways that automatically distinguish “15” and “IS”? The goals of this thesis are to search for a numeral-supporting perceptual pathway, and probe the functional and representational properties of its key perceptual nodes.

How Does the Brain Process Numerals? The “Visual Number Form” Hypothesis

Many of the conjectures about how the brain supports Arabic numeral reading stem from neuropsychological findings from patients with brain lesions and split brains. In particular,

Cohen and Dehaene (1991) introduced a cognitive module called “Visual Number Form” in the case study of patient YM who had a left temporal lobectomy. YM made many errors when asked to read aloud in French a list of 1- to 8-digit numerals. The reading errors YM made were mostly substitution errors that maintained the string length and syntax, and tended to be perseveration of the digit in the same relative position in the preceding numeral (e.g., “78, 233, 6534, 52, 6453” read as “78, 733, 7534, 73, 7453”). The errors also affected the leftmost digit more frequently than other digit positions even though YM showed no clinical signs of a general left visual hemifield neglect. As the perseverations still respected the place value of the affected digit (e.g., digit 7 could be perseverated with “sept” (seven) or “soixante-dix” (seventy)), an impairment at the level of extracting the verbal labels was ruled out. Errors were predicted by visual similarity predominantly, but also partially by numerical distance (e.g., digit 3 is more likely to be substituted by 4 than by 9), and affected digits 0 and 1 substantially less than other digits. Taken together, the observations suggest that the impairment was at the level of digit representations and their visuo-spatial properties.

Interestingly, the deficit was specific to reading aloud of numerals, and not when comparing their magnitudes or performing calculations on them, thereby ruling out a deficit at the level of visual input or attentional mechanisms. To account for the patterns of findings, Cohen and Dehaene (1991) hypothesized a visual buffer or workbench called the “Visual Number Form” that is accessed twice during numeral reading (**Figure 1-1**): First, information about string length and special digit identities serving as syntactic markers (e.g., 0 and 1) are extracted to create a syntactic frame, and then a word frame; Second, the remaining digit identities are accessed to fill the word frame and to retrieve their phonological representations. The functional locus of YM’s deficit was in the second access of the “Visual Number Form”

because his construction of the syntactic and word frames was intact.

In summary, Cohen and Dehaene (1991) hypothesized that the “workbench storing the input data for number reading is visual rather than semantic in nature... [it] encodes several visuo-spatial characteristics of the stimulus, including the shape and the relative positions of the digits, presumably in a stimulus-centred co-ordinate system... We propose to name this store visual number form, by analogy to the visual word form, which is at the level at which a string of characters is identified as an orthographic entity” (p. 54). Although it is unclear whether YM had similar patterns of errors with alphabetic reading, it appears that the “Visual Number Form” is assumed to be conceptually distinct from the “Visual Word Form”.

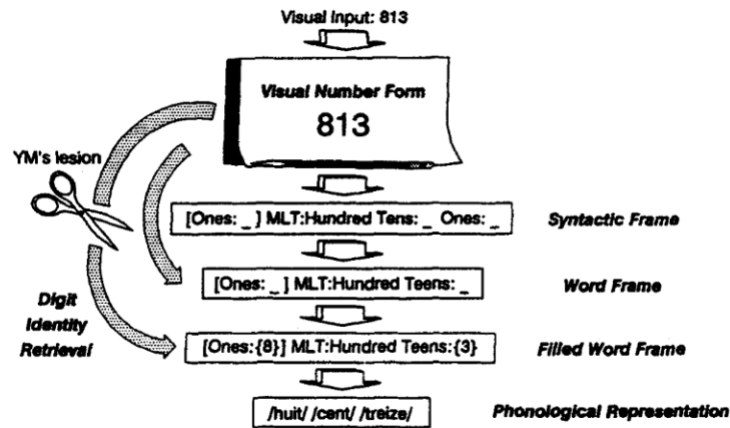


Figure 1-1. Cohen and Dehaene's (1991) cognitive model of Arabic numeral-to-word transcoding

Reprinted from Cohen & Dehaene (1991) with permission from Taylor and Francis.

Beyond the “Visual Number Form”: Triple-code Model of Number Processing

Incorporating the “Visual Number Form” hypothesis with other neuropsychological

findings from a variety of numerical tasks, Stanislas Dehaene and Laurent Cohen put forth a cognitive-anatomical “triple-code model” of number processing (Dehaene, 1992; Dehaene & Cohen, 1995). In the triple-code model, as illustrated in **Figure 1-2**, Dehaene and Cohen hypothesized three types of mental representations of number, and their corresponding neural substrates (Dehaene, 1992; Dehaene & Cohen, 1995).

Depending on the external format in which numbers are presented to us – collection of objects, Arabic numerals, or spoken and written number words – and what we do with them, different combinations of any of three distinct mental (and neural) representations could be recruited: First is the *visual number form code* in which numbers are “represented as strings of digits on an internal visuo-spatial scratchpad” (Dehaene & Cohen, 1995, p. 85) (e.g., 53 as [5][3] instead of [3][5]). This code is thought to be subserved by the bilateral ventral occipitotemporal cortices (Dehaene & Cohen, 1995) that comprise a mosaic of sub-regions thought to be specialized for the recognition of different object categories including letter strings (Allison et al., 1994; Grill-Spector & Malach, 2004; Grill-Spector & Weiner, 2014; Kanwisher, 2010). Second is a *verbal code* (or verbal word frame) in which numbers are represented lexically, phonologically, and syntactically as organized sequences of words (e.g., 485 → Hundreds[4] Tens[8] Ones[5] → “four hundred eighty-five”). The verbal code also supports verbally encoded arithmetic facts (e.g., “four times five is twenty”), and is thought to be subserved by left-lateralized regions implicated in general-purpose language processing, including the left inferior frontal, middle and superior temporal gyri, that extend to the inferior parietal lobule, comprising the angular and supramarginal gyri (Dehaene et al., 2003; Dehaene & Cohen, 1995). Evidence of a dissociation between reading number words and non-number words in patients with aphasia, however, suggest that the verbal code may be domain-specific (Cohen et al., 1997; Dotan &

Friedmann, 2015; Marangolo et al., 2004, 2005; for reviews, see Messina et al., 2009; Piras & Marangolo, 2009). Third is an *analogue magnitude code* in which numbers are represented by distributions of activation on a “mental number line” (Dehaene & Changeux, 1993). This code allows us to estimate, compare, and manipulate numerical magnitudes, and is thought to be subserved primarily by the bilateral intraparietal sulci, with support from the posterior superior parietal cortices for attentional orientation on the mental number line (Dehaene et al., 2003; Dehaene & Cohen, 1995).

The extent to which the visual, verbal, and magnitude codes are recruited depends on the stimulus input and task. The visual number form code is thought to be recruited as an “input stage common to *any* task involving number manipulations, including reading, magnitude comparison, calculation, etc.” (L. Cohen & Dehaene, 1991, p. 54, emphasis mine). Comparing whether “42” or “39” is greater is hypothesized to also recruit the magnitude code. The verbal code is not thought to be critical because patients with alexia who could not read aloud numerals could nonetheless compare their magnitudes with perfect accuracy (L. Cohen & Dehaene, 1995, 1996). In contrast, reading aloud multi-digit numerals not for the purpose of quantity manipulation is predicted to also recruit the verbal code, but not the magnitude code.

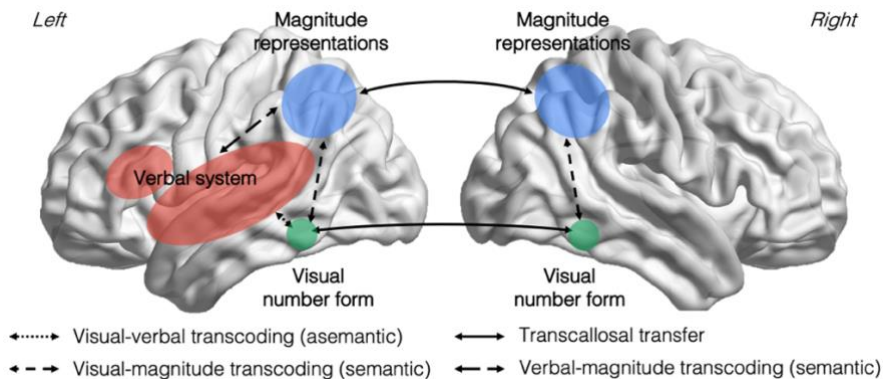


Figure 1-2. Triple-code model of number processing
Three types of mental representations of numbers and their corresponding hypothetical neural substrates. Adapted from Dehaene & Cohen (1995).

A considerable proportion of numerical cognition research has focused on the magnitude and verbal codes, and have found consistent empirical support for neural circuits underlying these two representational codes (for reviews and meta-analyses, see Ansari, 2008; Arsalidou & Taylor, 2011; Cohen Kadosh et al., 2008; Dehaene et al., 2003; Houdé et al., 2010; Kaufmann et al., 2011; Moeller et al., 2015; Sokolowski et al., 2017; Zamarian et al., 2009). However, there has been a lack of robust and convergent evidence of any high-level visual brain regions that appeared to be functionally specialized for processing Arabic numerals. The neural circuits underlying the visual number form code has therefore received less attention.

Do The “Visual Number Form” and “Visual Word Form” Systems Share Neural Resources?

Functional magnetic resonance imaging (fMRI) has made possible the discovery of a mosaic of brain regions in the ventral occipitotemporal cortex (vOT) that appear to be the key perceptual nodes of functionally specialized circuits for high-level visual recognition of different object categories, such as faces, scenes, body parts, and written words (for a review, see Kanwisher, 2010). Typically, a category-preferring vOT region is functionally defined based on a statistically greater blood-oxygenation-level-dependent (BOLD) response for the preferred category than non-preferred categories. For instance, a region preferring words would show a greater BOLD response to words than to line drawings of objects, false fonts, consonant strings, and Arabic numerals (e.g., Baker et al., 2007). At the spatial resolution enabled by fMRI, it is

more common that a BOLD preference for a particular category is not all-or-nothing, but a matter of degree (L. Cohen & Dehaene, 2004; Kanwisher, 2010). Importantly, compared to other brain recording techniques such as positron emission tomography and intracranial electroencephalogram, fMRI has been a highly effective and preferred method for the functional localization of cognitive processes due to its combination of being non-invasive and non-restrictive in the types of populations that can be studied, as well as allowing whole-brain coverage with high spatial resolution.

Over the past two decades, the use of fMRI has led to the discovery and an extensive study of a highly localized neural correlate of the visual word form called the "Visual Word Form Area" (VWFA) (for reviews, see Cohen et al., 2000, 2002; Cohen & Dehaene, 2004; Dehaene & Cohen, 2011; McCandliss et al., 2003; C. J. Price & Devlin, 2003, 2011). The VWFA, found in the mid-fusiform gyrus and along the left occipitotemporal sulcus (see **Figure 1-3**), is thought to represent letter strings as "an ordered set of abstract letter identities" (L. Cohen et al., 2002, p. 1054), and it fulfils three criteria for a functionally specialized region (L. Cohen & Dehaene, 2004; Dehaene & Cohen, 2011). First, the VWFA is *functionally specialized* for word reading in that it is tuned to processes required for reading in a specific script (e.g., invariance to size, letter case, and font). Second, the VWFA responds more to words and pronounceable pseudowords than to consonant strings, digit strings, and other object categories, suggesting some degree of *regional selectivity*. Third, the VWFA has a *reproducible location* across individuals, and across cultures with vastly different scripts (e.g., alphabetic versus ideographic) (L. Cohen & Dehaene, 2004; see Dehaene et al., 2015, for a review; Dehaene & Cohen, 2011). In alphabetic readers, the VWFA is also found to be the terminal stage of a hierarchical perceptual processing system that shows sensitivity to increasingly larger fragments

of alphabetic strings along the posterior-to-anterior gradient of the vOT – from strings of infrequent letters, to strings of frequent but rare bigrams, to strings of frequent bigrams and rare quadrigrams, and to real words (Vinckier et al., 2007).

Digit strings, however, do not seem to be represented in the VWFA to the same extent as words do. Not only is the VWFA less engaged by digit strings than by words, pseudowords, and consonant strings (Baker et al., 2007; James et al., 2005), but it also shows equivalent engagement to digit strings and words in an unfamiliar script (e.g., Hebrew words to non-Hebrew readers), which suggests that the VWFA does not care about numerals more than it does about non-meaningful characters (Baker et al., 2007). Moreover, Arabic digits do not form frequent combinations analogous to bigrams, trigrams, or quadrigrams that can leverage the same processing hierarchy as words do. Hence, although single digits and letters may be processed by a common representational system during the earlier stages of visual processing such as the identification of character features and allographs (McCloskey & Schubert, 2014), they likely diverge in their processing pathways at the level of character strings (Dotan & Friedmann, 2019).

Given that the VWFA can easily be localized using fMRI, and that digits are at least as ubiquitous as characters used for word reading (e.g., letters), it is reasonable to assume that an analogous “Visual Number Form Area” (VNFA) can also be detected using the same method. However, many fMRI studies failed to find a numeral-preferring region in the vOT, despite many of them having sufficient sensitivity to localize word-preferring or letter-preferring regions within the same sample or individual (e.g., Baker et al., 2007; James et al., 2005; Polk et al., 2002). Even when evidence of a numeral-preferring region along the ventral visual stream is reported in some studies, they do not appear to have a reproducible location in each hemisphere, spanning from the fusiform gyrus (e.g., Holloway et al., 2013; Pinel et al., 2001; Vogel et al.,

2015), to the inferior temporal gyrus (e.g., Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016), to the lateral occipital cortex (e.g., Park et al., 2012; L. Peters et al., 2015). This is particularly puzzling given that the location of the VWFA is consistent even between visually distinct alphabetic and logographic scripts (Bolger et al., 2005; Szwed et al., 2014; A. C. N. Wong et al., 2009), but Arabic digits being the most common script for numerals appear not to recruit a region that is as highly localized as the VWFA. Thus, if reproducible localization is one of the key criteria for defining a functionally specialized region (L. Cohen & Dehaene, 2004; Dehaene & Cohen, 2011), the question of whether a VNFA exists in the vOT remains open.

A VNFA Can Be Reproducibly Localized, but fMRI May Not Be Optimal

Initial intracranial electrophysiological recordings in epilepsy surgery patients found numeral-selective N200 (a negative-going event-related potential typically associated with visual processing around 200 ms post-stimulus onset) in several sites across the fusiform and inferior temporal gyri, but letter string N200 was also recorded from the same sites (Allison et al., 1994). These findings led the authors to conclude that processing of letter strings and numerals may be less spatially and functionally distinct (Allison et al., 1994). More recent intracranial stimulation and electrophysiological work in epilepsy surgery patients suggests that the most likely location of a numeral-preferring region may be within the posterior inferior temporal gyrus (pITG). Roux and colleagues (2008) stimulated many intracranial sites across the cerebral cortex, and found sites in the left pITG, inferior frontal gyrus, and supramarginal gyrus, in which stimulation selectively impaired reading aloud of multi-digit numerals, but not number words and sentences. Their findings thus suggest that the pITG is causally involved in the processing of Arabic numerals, and is a likely location of a VNFA. The fact that stimulation of any one of the three regions impaired numeral reading also suggests that although each of them is necessary for

numeral reading, they are clearly not sufficient alone.

Using intracranial electroencephalogram (iEEG) in epilepsy surgery patients, Shum and colleagues (2013) also found a circumscribed set of sites in the right pITG (but also in a much limited set of electrodes in the left pITG of some patients) that were more responsive to single Arabic digits than to characters with similar curvilinear features (letters, scrambled digits and letters, and numerals from a foreign script), as well as to single- and double-digit numerals than to semantically similar number words (e.g., “twenty”) and phonologically similar non-number words (e.g., “plenty”). Their findings thus suggest that the region may be tuned specifically to the overall visual forms or configurations of Arabic digits, rather than driven by the mere presence of constituent visual features, or by phonological or semantic processing. Importantly, this was the first study to provide a precise location of the candidate VNFA in a standardized stereotactic space: MNI 51, -54, -24 (see **Figure 1-3**).

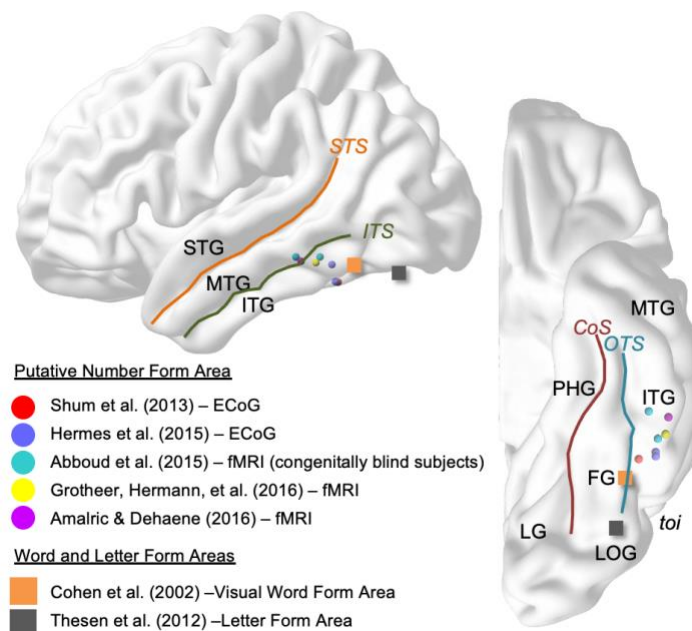


Figure 1-3. “Number Form Area”, “Visual Word Form Area”, and “Letter Form Area”

Candidate location of each area on the lateral temporal and ventral occipitotemporal cortices in the left hemisphere. STG: Superior temporal gyrus. MTG: Middle temporal gyrus. ITG: Inferior temporal gyrus. STS: Superior temporal sulcus. ITS: Inferior temporal sulcus. OTS: Occipitotemporal sulcus. CoS: Collateral sulcus. *toi*: Temporo-occipital incisure. FG: Fusiform gyrus. PHG: Parahippocampal gyrus. LG: Lingual gyrus. LOG: Lateral occipital gyrus. ECoG: Electrocorticography. fMRI: Functional magnetic resonance imaging.

Similar numeral-preferring sites were subsequently reported in other iEEG studies from the same research group (Daitch et al., 2016; Hermes et al., 2017; Pinheiro-Chagas et al., 2018), but because they relied on overlapping cohorts of subjects, its spatial reproducibility with different samples, particularly neurotypical ones, was not clear.

Based on its location being more lateral to the word- and letter-preferring fusiform regions, Shum and colleagues (2013) hypothesized that previous fMRI studies were not successful in localizing a VNFA because of its proximity to the air-bone interface of the auditory canal and venous sinus, and is thus susceptible to complete magnetization dephasing and BOLD signal attenuation. This insight led to a set of fMRI studies that adopted various advanced acquisition and post-acquisition methods to mitigate signal loss, and were all successful in localizing a VNFA in the left and/or right pITG (Abboud et al., 2015; Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016).

Interestingly, the first fMRI study to functionally localize a “visual” NFA proper in the pITG following the revelation of possible signal dropout around the region was in congenitally blind individuals (Abboud et al., 2015). Based on emerging evidence that category-preferring regions such as the VWFA can develop in the vOT even in individuals without visual experience (Reich et al., 2011; Striem-Amit et al., 2012), Abboud and colleagues (2015) examined the hypothesis that the VNFA is not involved in the computation of visual form per se, but to

“decipher [the shapes of] symbols for the purpose of connecting them to a quantity representation” independent of the sensory modality (p. 2). To test this hypothesis, they trained congenitally blind participants to associate auditory soundscapes with the shapes I, V, and X (the time, frequency, and intensity dimensions of sounds are isomorphically mapped to the visual features), which could then be decoded as letters or Roman numerals (e.g., V as five). They found a region in the right pITG (MNI 53, -44, -12)¹ that responded more when participants identified each soundscape as a specific numeral than as a letter. They also found a region in the left occipitotemporal sulcus close to the canonical VWFA site that responded more when the same soundscapes were identified as letters than as numerals. In other words, the regions were sensitive to abstract categories (numeral or letter) based on the *same* sensory form, indicating some degree of top-down control on which region was recruited for the task. They further showed using resting-state functional connectivity analyses in blind subjects and healthy controls that the meta-modal NFA had greater intrinsic connectivity with parietal regions implicated in quantity processing compared to the letter-preferring region, which had greater intrinsic connectivity with regions implicated in language processing. Such intrinsic functional connectivity biases were further replicated in children, and were found to be present even prior to formal schooling (Nemmi et al., 2018). Hence, these findings suggests that the NFA location may be constrained by structural connectivity biases between the vOT and quantity processing regions to link numerals to quantity representations (Hannagan et al., 2015).

To mitigate signal dropout, Abboud and colleagues (2015) used a novel statistical algorithm to detect voxels containing attenuated BOLD signals (Peer et al., 2016) and exclude

¹ A left pITG cluster was observed in a symmetric location (MNI -56, -50, -8) at an uncorrected threshold, but did not survive correction for multiple comparisons.

them prior to analyses. However, the authors did not compare their results with and without signal-intensity thresholding. Hence, their study could not confirm that signal attenuation would have led to a failure to localize an NFA in the first place; it was simply assumed to be the case.

Despite the sparse evidence available from only two studies with neuro-atypical populations as described above (i.e., epilepsy and congenitally blind), Hannagan and colleagues (2015) concluded that “the VWFA and the NFA are always localized and highly reproducible in the occipitotemporal cortex across subjects, fonts, and even sensory modalities” (p. 374). However, it cannot be ruled out that there could be neural reorganization in neuro-atypical populations that may not generalize to neuro-typical populations. Compared to the VWFA, there was also a lack of rigorous testing of the stimulus properties that the NFA is invariant to or sensitive to.

Nonetheless, subsequent fMRI studies provided evidence that an NFA in the pITG can also be localized in neurotypical individuals. In a study with neurotypical adults, Grotheer and colleagues (2016) went to great lengths to mitigate signal loss. For scan acquisition, they used a 64-channel head coil to enhance signal-to-noise ratio (SNR), parallel imaging acceleration factor of 3 to reduce susceptibility-based image distortion, high spatial resolution of 1-mm isotropic voxels to reduce intravoxel magnetization dephasing, and localized shimming to improve magnetic field homogeneity. They also used liberal spatial smoothing (8 times the voxel size) to increase the SNR. Using a 1-back repetition detection task with Arabic numerals and control categories that included letters and scrambled alphanumeric characters, they succeeded in localizing NFAs in the bilateral pITG (MNI -60, -57, -17 and 61, -45, -17), and also in the fusiform gyri (MNI -46, -57, -15 and 42, -49, -17). For unknown reasons, the authors focused on the pITG regions for further analyses, and ignored the findings of the fusiform regions as

additional candidate NFAs. Using a region-of-interest (ROI) analysis focusing specifically on the pITG regions, they also found that the preference for numerals in the pITG held for single digits, which suggests that single digits are sufficient to engage the NFAs and are not exclusive to multi-digit numerals. The authors then used the pITG peak voxels as ROIs in a separate experiment to compare the quality of the BOLD signals in the regions across different scan sequences that varied in voxel resolution (1 mm vs. 3 mm) and shimming method (standard vs. localized). The combination of a high-resolution scan and liberal spatial smoothing was most effective in mitigating signal attenuation in the pITG. However, because a priori regions were used, the authors did not directly demonstrate that the low-resolution scans failed to localize the NFAs in the first place.

Amalric and Dehaene (2016) also used a repetition detection task with Arabic numerals and control categories that included words and math formulas, and were successful in localizing NFAs in the bilateral pITG (MNI -56, -51, -19 and 62, -39, -17) in neurotypical adults. The authors attributed their successful localization of the NFAs to their “fast high-resolution fMRI sequence”, which included a multiband factor of 4, parallel imaging acceleration factor of 2, and 1.5 mm isotropic voxel resolution (Amalric & Dehaene, 2016, p. 4915). Nonetheless, without comparing their sequence with a slower and standard 3-mm resolution sequence, there was no direct evidence supporting their claim.

In sum, none of the studies above provided convincing evidence that directly supports the hypothesis that the prior lack of fMRI reports of an NFA was primarily due to signal dropout. Moreover, a retrospective inspection of fMRI studies revealed that some studies found numeral-preferring regions in the ITG (Cui et al., 2013; Cummine et al., 2015), and some in the fusiform gyrus (Grotheer, Herrmann, et al., 2016; Pinel et al., 2001). Hence, while the recent set of studies

offered a qualitative sense of a reproducible location of an NFA (see **Figure 1-3**), whether it lies within the pITG or fusiform gyrus, spans both gyri, or has different locations in each hemisphere remains to be settled. If signal dropout did not affect all studies equally, there could be more evidence of the existence of an NFA in the extant fMRI literature, but such evidence might not have been highlighted because functional localization in the vOT was not a primary aim of those studies. Therefore, an immediate outstanding question is “Where is the most probable reproducible location of the putative NFA?”. Knowing more confidently where in the vOT to look for numeral preference can then allow targeted investigations of its role in numeral processing.

Potential Hemispheric Asymmetries in the Recruitment of the Bilateral NFAs

Another factor that could affect a robust and consistent localization of an NFA in either hemisphere is the possibility that the left and right NFAs may not be engaged to the same extent across tasks. Although evidence from split-brain patients suggest that the visual systems in both hemispheres are capable of processing single Arabic digits (Cohen & Dehaene, 1996; Colvin et al., 2005; see Dehaene & Cohen, 1995, for a review; Sergent, 1990; Seymour et al., 1994; Teng & Sperry, 1973), studies in patients with left vOT lesions (L. Cohen & Dehaene, 1991, 1995, 2000; Miozzo & Caramazza, 1998) or split brains (L. Cohen & Dehaene, 1996; Gazzaniga & Smylie, 1984; Seymour et al., 1994) suggest that the bilateral NFAs are neither a single functional unit, nor functional duplicates working in parallel, but are functionally dissimilar and independent depending on the task contexts. For instance, left vOT lesions affected reading aloud of numerals and retrieval of verbally encoded arithmetic facts, but spared numeral magnitude comparison and some types of mental calculation (L. Cohen & Dehaene, 1991, 1995, 2000). Such neuropsychological evidence led Cohen and Dehaene (1995) to propose that we

possess “two number identification systems, possibly residing in different hemispheres, and which may be separately called upon depending on the task” (p. 123). In particular, the left hemispheric number identification system is hypothesized to be necessary for tasks that rely on verbal processes (e.g., multi-digit numeral reading and retrieval of verbally encoded arithmetic facts) due to its co-lateralization with the left-lateralized language system (L. Cohen & Dehaene, 1995; Dehaene & Cohen, 1997; Pinel & Dehaene, 2010) (see **Figure 1-2**).

Although there is now a significant body of evidence supporting the existence of bilateral NFAs in adults, a non-systematic and non-exhaustive review in **Table 1-1** shows that lateralization of numeral processing is not uncommon. One possibility is that the left NFA could be more variable in its location than the right NFA, possibly resulting from competition for neuronal space by spatially varied left-lateralized letter- and word-preferring regions (Glezer & Riesenhuber, 2013). Another possibility is that varied task demands involving the left-hemispheric language system may recruit the left NFA to different extents across studies.

Among the fMRI studies that observed a bilateral engagement of the numeral identification systems in **Table 1-1**, some found no hemispheric differences in the response profiles of the bilateral NFAs to different object categories including letters and novel characters (Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016), while others do report evidence of hemispheric asymmetry in certain functional properties of the bilateral NFAs (Amalric & Dehaene, 2016; Pollack & Price, 2019). Therefore, it remains unclear when the left and/or right NFA is recruited, and whether they represent numerals differently. Knowing this can not only shed light on why localization of the NFAs has been elusive with fMRI, but also how we should approach the study of their specialization(s) (i.e., whether we should treat the left and right regions interchangeably or distinctly).

Table 1-1. Evidence of lateralization of numeral processing in the ventral occipitotemporal (vOT) in adult fMRI, EEG and MEG studies

Left lateralization	Right lateralization	Bilateral
Fernandes et al. (2005): fMRI	Pinel et al. (1999): fMRI	Dehaene (1996): EEG
Fias et al. (2007): fMRI	Pinel et al. (2001): fMRI	Basso et al. (2003): fMRI
Holloway et al. (2013): fMRI	Knops et al. (2006): fMRI	Grotheer et al. (2016): fMRI
Vogel et al. (2015): fMRI	Gullick & Temple (2011): fMRI	Amalric & Dehaene (2016): fMRI
Vogel et al. (2017): fMRI	Park et al. (2012): fMRI	Grotheer et al. (2018): fMRI
Pollack & Price (2019): fMRI	Cui et al. (2013): fMRI	Bugden et al. (2019): fMRI
	Park et al. (2014): EEG	Aurtenetxe et al. (2020): MEG
	Carreiras, Monahan, et al. (2015): MEG	
	Cummine et al. (2015): fMRI	
	Abboud et al. (2015): fMRI	
	Carreiras, Quiñones, et al. (2015): fMRI	
	Park et al. (2018): EEG	
	Lochy & Schiltz (2019): EEG	
	Goffin, Vogel, et al. (2019): fMRI	
	Goffin, Sokolowski, et al. (2019): fMRI	
	Conrad et al. (2020): fMRI	

Note. EEG: electrical encephalography. MEG: magnetic encephalography. fMRI: functional magnetic resonance imaging. For EEG and MEG, vOT involvement is often characterized by event-related potentials within the first 200 milliseconds of stimulus onset (e.g., N170 component) or by frequency tagging. Intracranial recording studies were excluded as the constraint of electrode placements precluded any inference about lateralization.

Overview of Current Studies

The current state of knowledge leaves several questions unanswered. In particular, where are the most probable locations of the NFAs, if they do exist? What information do they represent? Are the left and right NFAs functionally equivalent? In this thesis, I present three studies that aimed to fill these gaps by further characterizing the spatial, functional, and representational aspects of the bilateral NFAs in neurotypical individuals using fMRI.

In Chapter 2, we asked: In the extant literature, is there quantitative evidence of a reproducibly localized NFA in the vOT in either hemisphere of neurotypical adults? We conducted a systematic review of the literature, followed by a coordinate-based meta-analysis of studies that *did* find numeral preference *somewhere* in the brain to ask whether there is spatial convergence in the vOT that would provide evidence for the existence of an NFA. We were

additionally interested in elucidating other regions outside of the vOT that could be part of a network of brain regions that supports numeral processing. To foreshadow our results, we found evidence for a reproducibly localized NFA in the right pITG.

Even though an NFA may exist, it remains unclear whether it is sensitive to or represents the visual form of numerals. The case for its role in shape processing (Hannagan et al., 2015) has been made by ruling out processing of curvilinear features, semantics, or phonology as alternative explanations (Shum et al., 2013). In other words, there is no direct evidence of shape processing thus far. Evidence for an NFA has also only been demonstrated with tasks that require attention to a character's shape in order to identify or categorize it (e.g., 1-back repetition detection, familiarity categorization, and naming). It is thus unknown whether the NFA is truly sensitive to numeral shapes, independent of decision-making processes that depend on a character's shape. Another related question is whether the NFA exhibits an experience-driven processing bias for its preferred category, like other category-selective regions in the ventral visual stream (e.g., Dehaene et al., 2001, 2004; Gauthier et al., 1999, 2000). If the NFA is functionally specialized for numerals, it should automatically process numerals (but not non-numerals) even when they are not irrelevant for the task. So, in Chapter 3, we investigated whether the NFA is sensitive to visual form and whether it distinguishes numerals from other character categories automatically during passive viewing. Using representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008), we compared the similarity of multivoxel patterns evoked by individual digits, letters and novel characters within the candidate NFA to describe how the neural representations of the individual characters were organized in a multi-dimensional stimulus space. For instance, the neural representations could be organized based on category membership with distinct clusters for digits and for non-digits, or they could be

organized based on shape similarity (e.g., 5 may be represented more similarly to S than to 4), or both.

Finally, findings of lateralization in numeral processing, as well as our findings from Chapters 2 and 3, suggest that the bilateral NFAs may not be functionally identical. A recent study by Pollack and Price (2019) found an intriguing dissociation between the left and right NFAs – the left pITG region (but not the right) showed sensitivity to digits when a digit was detected among letters, and the digit sensitivity of the right pITG region (but not the left) correlated positively with symbolic calculation skills. However, the study left us wondering how the left and right NFAs are similar or different in their functions and the information they represent, especially *within* individuals. This next step is crucial because it can inform us about whether we can generalize the inferences we make about the left NFA to the right NFA, and vice versa. So, in Chapter 4, we re-analyzed Pollack and Price's (2019) dataset with a comprehensive set of region-of-interest-based univariate and multivariate pattern analyses to probe the extent to which the bilateral NFAs are functionally dissimilar. Against the backdrop of studies finding positive associations between reading ability and responses to words in the VWFA (e.g., Dehaene et al., 2010; Feng et al., 2020), we also asked whether the functional and representational properties of the bilateral NFAs and their hemispheric asymmetries are related to individual differences in calculation skills.

In summary, Chapter 2 focused on whether there is reproducible localization of an NFA across samples and paradigms in neurotypical adults; Chapters 3 and 4 then probed the functional and representational properties of candidate bilateral NFAs when numerals are passively viewed or actively processed, respectively; and Chapter 4 further investigated the relation between numeral representations in the NFAs and symbolic calculation skills.

Taken together, findings from the studies not only address whether the NFAs satisfy certain criteria for functional specialization, but also speak to the potential of NFAs as biomarkers of Arabic numeral reading fluency. The findings will lay the groundwork for a more complete understanding of how our brains support efficient reading of numerals, when and how that neural architectural transformation begins, and what happens when the numeral reading system does not develop typically. More broadly, whether a specialized perceptual system for reading numerals exists will add to the growing knowledge base of perceptual systems specialized for reading different culturally defined symbol sets, such as words based on different writing systems and scripts (e.g., Liu et al., 2007; L. Martin et al., 2019; Nelson et al., 2009; Perfetti et al., 2007), and music notation (e.g., Bouhali et al., 2017, 2020; Mongelli et al., 2017; Y. K. Wong et al., 2014; Y. K. Wong & Gauthier, 2010). Given the youth of these symbol sets relative to the existence of modern humans, these specialized systems are unlikely to have evolved by natural selection. The collective effort in investigating these specialized systems can therefore contribute to a broader and deeper understanding of how human culture interacts with our neurobiology. In particular, they inform how our brains assimilate or accommodate the learning of multiple culturally defined symbol sets for similar or distinct representational purposes (Dehaene & Cohen, 2007), and the behavioral implications when a brain struggles to do so.

CHAPTER 2

THE SEARCH FOR THE “NUMBER FORM AREA”: A FUNCTIONAL NEUROIMAGING META-ANALYSIS

This chapter is adapted from “The search for the number form area: A functional neuroimaging meta-analysis” published in *Neuroscience & Biobehavioral Reviews*, and has been reproduced with the permission of the publisher and my co-authors, Eric D. Wilkey, and Gavin R. Price.

Yeo, D. J., Wilkey, E. D., & Price, G. R. (2017). The search for the number form area: A functional neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 78, 145–160. <https://doi.org/10.1016/j.neubiorev.2017.04.027>

Introduction

If a “Number Form Area” exists in the vOT that is specific to Arabic numerals rather than written symbols in general, particularly because of its sensitivity to the digit shapes (Hannagan et al., 2015; Shum et al., 2013), it should evidence functional specialization by a greater engagement for Arabic numerals than for other meaningful character categories, across a range of task contexts. Although recent studies suggest a candidate NFA in the pITG (Abboud et al., 2015; Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016; Roux et al., 2008; Shum et al., 2013), there have been reports of numeral-preferring regions in the fusiform gyrus as another candidate NFA location (Grotheer, Herrmann, et al., 2016; Pinel et al., 1999, 2001). Indeed, previous coordinate-based meta-analyses of neuroimaging studies of number processing in neurotypical adults have revealed convergence of functional activation in the left fusiform gyrus, but not in the ITG (Arsalidou & Taylor, 2011; Sokolowski et al., 2017).

However, there are methodological limitations to those meta-analyses that preclude associating the left fusiform gyrus region with Arabic numerals specifically. In the meta-analysis

by Arsalidou and Taylor (2011), the authors grouped both Arabic numerals and nonsymbolic (e.g., dot arrays) numerical stimuli under an umbrella term “numbers”, rendering it impossible to claim that any meta-analytic convergence in the fusiform gyrus was specific to Arabic numerals. In a more recent meta-analysis that examined symbolic and nonsymbolic numerical stimuli in separate meta-analyses, Sokolowski and colleagues (2017) did not find spatial convergence anywhere in the vOT when they meta-analyzed studies that included contrasts of only Arabic numerals. Nonetheless, they found spatial convergence across contrasts from passive-viewing paradigms that included either Arabic numerals or number words (i.e., “symbolic numbers”, collectively) in the left fusiform gyrus, but this spatial convergence could be driven by contrasts that included number words that recruit the mid-fusiform letter- and word-preferring regions.

Besides the possibility that some studies did suffer from signal dropout in the pITG, the general lack of meta-analytic convergence in the pITG may also be due to the fact that previous meta-analyses included contrasts in which the cognitive subtraction approach would have eliminated any activation associated with the perceptual processing of Arabic numerals per se. For instance, a common contrast included in prior meta-analyses is “number comparison: near distance > far distance” (e.g., 4 vs. 5 compared to 4 vs. 9), in which the activation of interest is the effect of numerical distance. The common recruitment of an NFA in both distance conditions would be excluded via cognitive subtraction.

Another potential limitation of prior meta-analyses is the inclusion of contrasts with very different cognitive demands between the numeral and non-numeral conditions, such as comparing the magnitude of a target digit to 5 versus single letter naming. These may not seem problematic if the primary aim of the meta-analysis is to uncover all brain regions that are involved in a task, regardless of whether they are specific to the task or not, and if the numeral

condition is presumed to be more cognitively demanding than the non-numeral condition. However, it is difficult to be certain that a lower-level control condition (e.g., mere fixation) always involve fewer neural resources across the whole cortex (C. J. Price et al., 2005). Moreover, of concern in any investigation of functional specialization for a particular visual stimulus category is the extent to which the cognitive subtraction involved in a contrast reflects only the difference in stimulus category per se (hereafter referred to as “contrast specificity”). Hence, a contrast such as “comparing digit to 5 > letter naming” (e.g., Chochon et al., 1999) may not reflect only a difference in stimulus category, but also the different mental computations performed in each task (e.g., working memory involved in one, but not another). In fact, low contrast specificity may include unnecessary noise (i.e., activations of non-interest) in a meta-analytical procedure and decrease its statistical sensitivity to the activations of interest (C. J. Price et al., 2005). For instance, in an object-naming meta-analysis, Price and colleagues (2005) found enhanced sensitivity to activation in regions associated with semantic processing, visual-speech integration, and response selection with high-level baseline conditions (e.g., object naming > face orientation) that controlled for speech and perceptual processing than with low-level baseline conditions that did not control for those (e.g., object naming > fixation). It is thus unclear if we would observe a meta-analytic convergence in an NFA when we consider all relevant contrasts, or only when we restrict the analyses to numeral and non-numeral conditions with comparable task demands.

In sum, whether or not the extant body of functional neuroimaging literature does in fact support a reproducible localization of an NFA in the ITG or fusiform gyrus requires further quantitative investigation with careful selection of contrasts to be included in the meta-analysis.

Current Study

We conducted two coordinate-based Activation Likelihood Estimation (ALE) meta-analyses of neuroimaging studies that contrasted the visual presentation of Arabic numerals with other meaningful written symbols across a variety of tasks in healthy, neurotypical individuals.

In the first meta-analysis, we used a maximally inclusive approach by adopting liberal inclusion criteria with varying degrees of contrast specificity as long as Arabic numerals were contrasted with at least one other written symbol category (e.g., magnitude comparison > letter naming). We predicted that a liberal set of contrasts would increase the chances of revealing a convergence of activation in an NFA, especially if its activation was weakened by signal loss.

It is however possible that the inclusion of less specific contrasts in the meta-analysis may lead to ‘noisy’ results (C. J. Price et al., 2005), potentially masking any convergence in an NFA. The inclusion of less specific contrasts may also reveal regions other than the NFA that are related to differential cognitive demands between the numeral and non-numeral conditions rather than to the stimulus category per se. This would render our interpretation of the role of these other regions in supporting numeral processing challenging. To address both of these concerns, we conducted a second meta-analysis that included only contrasts with equivalent task demands between the numeral and non-numeral conditions (e.g., digit matching versus letter matching).

Methods

Literature Search and Article Selection

We used a multi-step approach to identify relevant articles. First, a literature search was made in the following databases in October 2016: PsycINFO (1806 – 2016), PubMed (1950 – 2016), Web of Science (1965 – 2016), ScienceDirect (1823 – 2016), and Google Scholar. To identify journal articles and book chapters that have empirically examined processing of Arabic numerals relative to other meaningful symbols, or have reviewed the empirical evidence for

visual number form processing, search terms included (“number form” OR “visual Arabic” OR “visual number” OR “numeral” OR “digit” OR “symbolic number” OR “number symbol” OR “Arabic digit”) AND (“fMRI” OR “functional magnetic resonance imaging” OR “PET” OR “positron emission tomography”).

A study was included in the systematic review if it met the following criteria: (1) published in English peer-reviewed journals, (2) sample comprised healthy, neurotypical human adults, (3) conducted whole-brain, within-group analyses using fMRI or positron emission tomography to minimize any bias towards predefined regions of interest, (4) employed tasks that contrasted the visual processing of Arabic numerals with at least one other category of written symbols that are orthographically and semantically familiar to the participants (e.g., letters of the Latin alphabet for English speakers (including their usage as Roman numerals), scripts of foreign languages such as Chinese and Japanese characters for Chinese or Japanese speakers respectively, and presumably universally known non-alphanumeric symbols such as \$, %, &, etc.), and (5) reported the coordinates of activation maxima in standardized stereotaxic space such as the Talairach (Talairach & Tournoux, 1988) or Montreal Neurological Institute (MNI) templates.

For criterion (4), the control categories against which Arabic numerals were compared to may also include nonsymbolic categories (e.g., Arabic numerals > letters, false fonts and objects), as long as response profiles were provided to illustrate significantly greater activation of the regions of interest to Arabic numerals than to all other symbolic and nonsymbolic control stimuli (e.g., Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016). In other words, the control conditions could include nonsymbolic stimuli, but they could not be the only control categories in the contrast (e.g., *only* false fonts, nonsymbolic dot arrays, shapes, body parts, or

fixation/rest), otherwise those contrasts were excluded.

Moreover, to be as comprehensive as possible, identifying an NFA did not have to be one of the aims of a study to be included in the meta-analyses. As a main goal of this study was to search for a convergence of studies in an NFA, we were also liberal in selecting studies that had a non-numerical control task with cognitive demands different than the numerical task, as long as the control stimuli used were meaningful written symbols.

Next, we crosschecked the reference lists of all the relevant empirical papers, review articles, meta-analyses, and book chapters, to identify additional studies that were not captured by the database searches. We also performed forward citation searches on the relevant studies that cited Shum and colleagues (2013), and studies that those articles cited in relation to an NFA. Finally, we excluded reviews, meta-analyses, case studies, studies that only presented auditory stimuli (e.g., Abboud et al., 2015, who also used Roman numerals instead of Arabic numerals), and studies that presented auditory and visual stimuli separately, but reported only supramodal contrasts for digits versus letters (e.g., Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003). An exception are studies that presented auditory and visual stimuli simultaneously, and required conscious processing of the visual symbols presented (e.g., Fernandes, Moscovitch, Ziegler, & Grady, 2005; Holloway, van Atteveldt, Blomert, & Ansari, 2015). **Table 2-1** shows all the studies included in the systematic review.

Subsequently, for the coordinate-based meta-analyses, we further excluded studies in which no supra-threshold activation was found for Arabic numerals relative to other meaningful symbols (see **Table 2-1**) because the ALE approach assesses spatial convergence across studies given that there is *some* supra-threshold activation somewhere in the brain. For studies in which the relevant contrasts were performed, but did not report the full list of coordinates, attempts

were made to obtain the necessary data, if available; otherwise, they were excluded (e.g., Knops et al., 2006).

Lastly, to be as inclusive as possible, and to avoid further subjective thresholding, all reported activation foci were included, regardless of whether they were global peaks or subpeaks (i.e., local peaks within an activation cluster), and corrected or uncorrected for multiple comparisons. The non-independence of these foci within a study was accounted for by the use of updated ALE algorithms for the meta-analyses (Turkeltaub et al., 2012).

A final set of 30 studies (from **Table 2-1**), all of which used fMRI, met the inclusion-exclusion criteria for the ALE meta-analyses. The sample in Gullick and Temple (2011) was split into two independent studies or subject groups as a between-subjects design was used, resulting in a total of 31 studies with 50 contrasts, 388 foci, and 510 subjects. Details of those studies are presented in **Table 2-2**. All meta-analyses conducted in the current study met the recommended minimum number of studies (i.e., 20) required to avoid results that are strongly influenced by individual studies and for sufficient power to detect moderate effects (Eickhoff et al., 2016).

Table 2-1. fMRI studies that contrasted Arabic numerals with other meaningful written symbols

<u>Excluded from current meta-analyses</u>	<u>Included in current meta-analyses</u>	
No numeral-specific activation found anywhere in the brain	Numeral-specific activation within the vOT	Numeral-specific activation only outside the vOT
1. Anderson et al. (2015)	1. Amalric & Dehaene (2016) – Bilateral ITG	1. Andres et al. (2012)
2. Baker et al. (2007)	2. Basso et al. (2003) – Bilateral FG	2. Andres et al. (2011)
3. Cantlon et al. (2011)	3. Coderre et al. (2009) – Left FG	3. Attout et al. (2014)
4. Cohen Kadosh et al. (2007)	4. Cui et al. (2013) – Right ITG	4. Cappelletti et al. (2010)
5. Fulbright et al. (2003)	5. Cummine et al. (2015) – Right ITG	5. Carreiras et al. (2015)
6. James et al. (2005)	6. Fernandes et al. (2005) – Left FG	6. Chochon et al. (1999)
7. Koul et al. (2014)	7. Fias et al. (2007) – Left FG	7. Cummine et al. (2014)
8. Polk & Farah (1998)	8. Grotheer, Herrmann, et al. (2016) – Bilateral ITG and bilateral FG	8. Holloway et al. (2015)
9. Polk et al. (2002)	9. Gullick & Temple (2011) – Right ITG	9. Libertus et al. (2009)
10. Reinke et al. (2008)	10. Knops et al. (2006) ^a – Right FG	10. Park et al. (2012)
11. van der Ven et al. (2016)	11. Pinel et al. (2001) – Right FG	11. Peters et al. (2015)
	12. Pinel et al. (1999) – Right FG	12. Pinel et al. (2004)
		13. Price & Ansari (2011)
		14. Stanescu-Cosson et al. (2000)
		15. Wu et al. (2009)
		16. Yin et al. (2015)
		17. Zago et al. (2008)
		18. Zarnhofer et al. (2012)
		19. Zhang et al. (2012)

Note. vOT: ventral occipitotemporal cortex; includes inferior temporal, fusiform, and parahippocampal gyri (Grill-Spector & Malach, 2004; Grill-Spector & Weiner, 2014). FG: fusiform gyrus. ITG: inferior temporal gyrus.

^aExcluded from the meta-analyses as stereotaxic coordinates were not provided for the contrast of interest.

Data Analyses

Two separate meta-analyses were conducted with different inclusion criteria based on contrast specificity. Contrast specificity was defined by the extent to which the contrast has the potential to reveal activation specific to stimulus category after controlling for task demands. Hence, all the relevant contrasts were categorized into three groups: (1) *less specific contrasts*, in which the numerical tasks had different cognitive demands than the non-numerical control tasks (e.g., multiplication > letter naming) (47.9% of foci), (2) *suitably specific contrasts*, in which the numerical tasks and non-numerical control tasks had equivalent or closely-matched task demands (e.g., digit naming > letter naming) and the main difference lies in the stimulus category (48.2% of foci), and (3) *overly specific contrasts* that controlled for all neural activations related to visual perceptual processes pertaining to number form, leaving only the higher-order effect of interaction between stimulus category and task demands (e.g., [Arabic digit calculation – Arabic digit identification] > [Roman numeral calculation – Roman numeral identification]) (5.7% of foci). A small number of contrasts were simultaneously less specific and overly specific based on our criteria, such as [Number magnitude comparison – Number dimming detection] > [Letter ordinal comparison – Letter dimming detection] (1.8% of foci).

Meta-analysis I: All Relevant Contrasts

In the first meta-analysis, all 31 studies with 50 contrasts, 388 foci, and 510 subjects were included as we wanted to be as liberal as possible in our search for an NFA. As the bulk of the less specific contrasts had more cognitively demanding numerical tasks than non-numerical ones, their inclusion should potentially increase the probability of finding a convergence in an NFA.

Meta-analysis II: Suitably Specific Contrasts Only

In the second meta-analysis, a subset comprising 20 studies with 28 contrasts, 187 foci,

and 351 subjects was analyzed to control for task demands and degree of contrast specificity by excluding the less specific contrasts and the overly specific contrasts, respectively. Thus, only suitably specific contrasts were included.

Activation Likelihood Estimation Procedure

Before analyzing the data, all foci were converted into a common stereotaxic space. Due to the variability in the templates used by various software for spatial normalization to MNI space (e.g., SPM, FSL, etc.), and that there is no reason to favor one template over another, we opted to transform all foci reported in MNI space to Talairach space using the various best-fit Lancaster transformation options (i.e., "icbm2tal"; Laird et al., 2010; Lancaster et al., 2007) implemented for different software, such as icbm2tal for SPM or icbm2tal for FSL. Studies that performed spatial normalization in MNI space in SPM99, SPM96, and SPM2, and reported those coordinates in Talairach space were assumed to have applied the Brett transformation (i.e., "mni2tal"; Brett et al., 2002). An exception were studies by Stanescu-Cosson and colleagues (2000) and Chochon and colleagues (1999), in which the published coordinates are believed to be in MNI (SPM) space instead of Talairach space as reported (verified with BrainMap's Sleuth 2.4 database; Laird et al., 2005). For those studies that applied the Brett transformation, we adopted the recommended approach to "un-Brett" the coordinates using the "Brett: Talairach to MNI" transform followed by the Lancaster "MNI (SPM) to Talairach" transform (Fox et al., 2013; Laird et al., 2010). After the transformations were made, 3 foci (0.89%) in the dorsal, middle frontal region fell outside of the smaller, more conservative mask, which is typically used for meta-analyses of functional imaging studies (Fox et al., 2013). Typically, one would expect less than 3% of the foci to fall outside of the mask, and outlying foci that are close enough to foci within the mask still contribute to the ALE calculations (Fox et al., 2013).

To quantitatively assess the concordance of brain regions that support the functional specialization of visual processing of Arabic numerals across different tasks and subject groups, coordinate-based meta-analyses were conducted using the ALE approach with GingerALE version 2.3.6 (Eickhoff et al., 2009, 2012, 2017; Turkeltaub et al., 2012). The ALE algorithm first modeled foci from contrasts within each study as centers of three-dimensional Gaussian probability distributions. Taking into consideration between-template variances, and the relation between study sample size and inter-subject localization uncertainty, studies with larger sample sizes were weighted such that they would have a narrower full width at half maximum and higher peaks (Eickhoff et al., 2009).

For each contrast, the maximum probability associated with any one Gaussian distribution at a given voxel would be used to generate a probabilistic map of modeled activation (MA) (i.e., an MA value represents the probability of an activation in a particular voxel). Taking the maximum probability allows for the inclusion of foci that are in close proximity within a contrast (e.g., global peak and local peaks of a cluster) without inflating the MA values (Turkeltaub et al., 2012). To minimize studies with multiple non-independent contrasts (e.g., target condition versus control condition A, and target condition versus control condition B) from having a greater influence on the results compared to studies with only one contrast, we collapsed all contrasts within a study, essentially treating them as stemming from a single contrast. Therefore the ALE approach described above determines an above-chance convergence of activation probabilities between *studies*, rather than between *foci* (Eickhoff et al., 2012). It should be noted that the above approach became feasible only with a revised ALE algorithm in 2012 (Turkeltaub et al., 2012). Prior to the availability of this revised algorithm, previous meta-analyses (Arsalidou & Taylor, 2011; Houdé, Rossi, Lubin, & Joliot, 2010; Kaufmann, Vogel,

Starke, Kremser, & Schocke, 2009) had to select only one contrast per study to ensure statistical independence between sets of foci. Not only would limiting to a single contrast per study introduce selection bias that might influence the results, but also preclude the uncovering of potentially informative findings that would have only been possible with a more inclusive approach to contrast selection.

An ALE map was then computed as the voxel-wise union of the MA maps across all studies from the meta-analysis dataset. The ALE values thus represent the likelihood that at least one study activated a given voxel. Next, a random-effects significance test was performed on the ALE scores against a null distribution of ALE scores under the assumption of spatial independence (see Eickhoff et al., 2012, for detailed descriptions of the concept and algorithm). The parametric map with associated p -values obtained from the significance test was then subjected to the recommended cluster-level family-wise error (FWE) correction for multiple comparisons, which was found to provide optimal compromise between sensitivity and specificity (Eickhoff et al., 2016). Using the Monte-Carlo based approach with 1,000 permutations, the cluster-level FWE thresholding was performed with an uncorrected, cluster-forming threshold of $p < .001$, followed by a cluster-level threshold of $p < .05$. Crucially, this cluster-thresholding step is dependent on a random distribution of all the foci in the dataset, which potentially influences the ultimate cluster-size threshold, such that a larger number of studies included in a meta-analysis is likely to result in higher p -values for the same ALE scores (Eickhoff et al., 2016). This may account for any differences in the findings between the two meta-analyses conducted here.

Other than reporting the ALE global peaks within each cluster, local peaks were also provided to illustrate the spatial extent of each cluster. To be consistent with previous meta-

analysis, anatomical labels were assigned to the ALE peak locations within each cluster using the Talairach Daemon (talairach.org) that is native to GingerALE. Supplementary anatomical labels were assigned using cytoarchitectonic maximum probability maps in the Anatomy Toolbox v2.2c (Eickhoff et al., 2005) after transforming the Talairach peak coordinates to MNI (SPM) space. Besides providing sulci labels, which Talairach Daemon does not, the Anatomy Toolbox also provides additional anatomical specificity, such as gyri and sulci subdivisions.

Table 2-2. Participant demographics, tasks, contrasts, foci of interest, and statistical threshold for the fMRI studies included in meta-analyses

Study	Reference	<i>N</i>	Task(s)	Selected contrast(s)	Foci	Statistical Threshold
<i>Included in both meta-analyses</i>						
1	Amalric & Dehaene (2016)	30	1-back same/different judgment	Decimals > Checkers, faces, bodies, tools, houses, mathematical formulas, and words	16 ^a	$p < .001$ uncorrected and $p < .05$ cluster-corrected
2	Cappelletti et al. (2010)	22	Conceptual judgment (quantity and non-quantity related) vs. Perceptual color judgment	Decimals > Object names (conceptual)	14	$p < .05$ FWE corrected, or $p < .001$ uncorrected
				Decimals > Object names (conceptual)	12 ^b	
				Decimals > Object names (quantity only)	14	
				Decimals > Object names (non-quantity only)	14	
				Decimals (Conceptual – Perceptual) > Object names (Conceptual – Perceptual) [#]	13	
3	Carreiras et al. (2015)	21	1-back same/different judgment	Digit strings > Consonant strings ^c	7	$p < .001$ uncorrected, or $p < .05$ peak/cluster-level FWE/FDR corrected
				Digit strings > Nonalphanumeric symbol strings ^c	2	
4	Coderre et al. (2009)	9	Number identification	(Arabic > Japanese Kana) and (Arabic > Japanese Kanji)	3	$p < .03$ uncorrected and $p < .05$ cluster-corrected
				Arabic > Japanese Kanji	4	

5	Chochon et al. (1999) ^e	8	Digit naming/Magnitude comparison/Multiplication/Subtraction vs. Letter naming	Digit naming > Letter naming	2	<i>p</i> < .001 uncorrected and extent threshold, <i>k</i> = 16 voxels (432 mm ³ , <i>p</i> < .05 corrected)
				Magnitude comparison > Letter naming*	13	
				Multiplication > Letter naming*	12	
				Subtraction > Letter naming*	14	
				Numerical tasks > Letter naming*	14	
6	Cui et al. (2013)	18	Semantic distance judgment	Digits > Chinese characters (Numeral classifiers)	9	<i>p</i> < .001 uncorrected and <i>k</i> = 10 voxels (270 mm ³)
				Digits > Chinese characters (tool nouns)	18	
7	Cummine et al. (2015)	15	Rapid automatized naming	Digits > Words	5	<i>p</i> < .001 uncorrected and <i>p</i> < .05 cluster-corrected
				Digits > Nonwords	2	
8	Cummine et al. (2014)	15	Rapid automatized naming	Digits > Letters	9	<i>p</i> < .001 uncorrected and <i>p</i> < .05 cluster-corrected
9	Grotheer, Herrmann, et al. (2016)	22	1-back same/different judgment	Digits > False digits, noise digits, letters, false letters, noise letters, and objects	6	<i>p</i> < .05 FWE corrected and <i>k</i> = 20 voxels (20 mm ³)
10	Holloway et al. (2015)	18	Passive simultaneous viewing and listening	Visual digits and audio number word > Visual letters and audio letter sounds/names	1	<i>p</i> < .005 uncorrected and <i>p</i> < .05 cluster-corrected
11	Libertus et al.	14 ^d	2-back same/different judgment	Digits > Letters and faces	6	<i>p</i> < .01 uncorrected,

	(2009)					$p < .05$ cluster-corrected, and $k = 8$ voxels (64 mm ³)
12	Park et al. (2012)	20	Simultaneous same/different judgment task	Digit strings > Consonant strings	1	$p < .005$ uncorrected and $k = 20$ voxels (540 mm ³ , $p = .01$ cluster-corrected)
13	Peters et al. (2015)	12	Subtraction	Digits > Number words	2	$p < .05$ FDR-corrected
14	Pinel et al. (2001)	9	Magnitude comparison	Digit strings > Number words	6	$p < .05$ corrected
15	Pinel et al. (1999)	11	Magnitude comparison	Digits > Number words	1	$p < .001$ uncorrected
16	Pinel et al. (2004)	15	Physical size comparison	Digits > Letters	1	$p < .01$ uncorrected and $p < .05$ cluster-corrected
17	Price & Ansari (2011)	19	Oddball target detection (passive viewing)	(Digits > Letters) and (Digits > False digits)	1	$p < .005$ uncorrected and $p < .05$ cluster-corrected
18	Yin et al. (2015)	11 ^f	Memory (ordinality)	Digit/digit strings > Letters (encoding)	1	$p < .001$ uncorrected and $k = 100$ voxels (normalized voxel size not reported)
				Digit/digit strings > Letters (recall)	3	
19	Zarnhofer et al. (2012)	42	Arithmetic verification	Digits > Number words	5	$p < .001$ uncorrected and $p < .05$ FWE cluster-corrected
20	Zhang et al. (2012)	20	Semantic distance judgment	Numerals > Chinese characters	22	$p < .008$ uncorrected and $k = 50$ voxels

(1350 mm³)

Excluded from Meta-analysis II

21	Andres et al. (2012)	18	Arithmetic vs. Letter naming	Arithmetic > Letter naming*	5	$p < .05$ FDR-corrected and $k = 100$ voxels (800 mm ³)
22	Andres et al. (2011)	10	Arithmetic vs. Letter naming	(Subtraction > Letter naming) and (Multiplication > Letter naming)*	8	$p < .001$ uncorrected, or $p < .05$ FWE cluster-corrected, $k = 150$ voxels (1200 mm ³)
23	Attout et al. (2014)	26	Numerical order judgment vs. Letter luminance judgment	Numerical order > Letter luminance*	5	$p < .05$ FWE-corrected
24	Basso et al. (2003)	5	Digit verification (working memory and temporal production) vs. Letter verification	Number working memory > Letter verification*	5	$p < .0001$ uncorrected, $k = 10$ voxels (360 mm ³), and peak-level corrected at $p = .01$
				Number temporal production > Letter verification*	8	
25	Fernandes et al. (2005)	12	Parity judgment vs. Animacy judgment [Divided attention (DA) with auditory word list recognition] Full attention (FA) auditory word list recognition/ visual number parity/ visual animacy judgment	(DA digits – FA auditory recognition) > (DA animacy – FA auditory recognition)*	17	$p < .005$ uncorrected and $k = 50$ mm ³
				(DA digits – FA digits) > (DA animacy – FA animacy)**	3	

				(DA digits – Auditory baseline) > (DA Animacy – Auditory baseline)* ^c	13	
26	Fias et al. (2007)	17	Comparison (number magnitude vs. letter ordinality) vs. Dimming detection	(Number comparison – Number dimming) > (Letter comparison – Letter dimming) ^{#*}	3	$p < .005$ uncorrected, masked with main effect results at $p < .05$ uncorrected, and $k = 5$ voxels (135 mm ³)
27	Gullick & Temple (2011) ^g	16	Ordinality comparison vs. passive viewing of nonalphanumeric symbols/nonwords	(Years as events – Symbols) > (Event Words – Nonwords)*	1	$p < .001$ uncorrected, $p < .05$ FDR cluster-corrected, and $k = 20$ voxels (540 mm ³)
				Years as events > Symbols* ^c	23	$p < .0001$ uncorrected, $p < .05$ FDR cluster-corrected, and $k = 10$ voxels (270 mm ³)
28	Gullick & Temple (2011) ^g	16	Magnitude comparison vs. passive viewing of nonalphanumeric symbols	Numbers > Symbols* ^c	8	$p < .0001$ uncorrected, $p < .05$ FDR cluster-corrected, and $k = 10$ voxels (270 mm ³)
				Years as numbers > Symbols* ^c	10	
29	Stanescu-Cosson et al. (2000) ^e	7	Exact and approximate calculation vs. Letter-matching	Digits (1 – 9) > Letters*	16	$p < .001$ uncorrected, and $p < .05$ cluster-corrected
				Small digits (1 – 5) > Letters*	7	
30	Wu et al. (2009)	18	Arithmetic verification vs. Symbol identification (contained both numerals and	(Arabic Calculation – Arabic Identification) > (Roman Calculation – Roman Identification) [#]	2	$p < .01$ uncorrected and $p < .01$ extent threshold

nonalphanumeric symbols)

31	Zago et al. (2008)	14	Manipulation (addition vs. noun generation) vs. Memory task	(Numbers Manipulation – Maintenance) > (Syllables Manipulation – Maintenance) ^{#*}	1	$p < .001$ uncorrected and $p < .05$ cluster-corrected
----	--------------------	----	---	---	---	--

Note. FWE: Familywise Error. FDR: False Discovery Rate. Unless otherwise stated, the statistical threshold reported was applied to all contrasts within a study. In some cases, a mixture of uncorrected and corrected findings may be reported for a particular contrast (e.g., $p < .05$ FWE corrected, or $p < .001$ uncorrected). [#]These contrasts were excluded from Meta-Analysis II to control for the influence of overly-specific contrasts. ^{*}These contrasts were excluded from Meta-Analysis II to control for the influence of differential task demands between the numerical and non-numerical tasks.

^aIncludes unpublished foci provided by Amalric & Dehaene (2016).

^bResponse times modeled over both number and object names were factored out from the main effect of stimulus for this specific contrast. All other contrasts had response times modeled separately for numbers and object names.

^cFrom supplementary data.

^dChild sample in this study was excluded from the meta-analysis.

^eOriginal coordinates were believed to be in MNI (SPM) space instead of the reported Talairach space.

^fA superior memorist as a case study subject ($n = 1$) was excluded from the meta-analysis.

^gThis study is split into two independent samples here as a between-subject design was used.

Results

Meta-analysis I: All Relevant Contrasts

The all-inclusive meta-analysis revealed two bilateral parietal clusters and one right frontal cluster that are functionally specialized for symbolic number processing (**Table 2-3** and **Figure 2-1**). The right parietal cluster comprised the intraparietal sulcus extending to the inferior parietal lobule, precuneus, and posterior superior parietal lobule. The left parietal cluster comprised the intraparietal sulcus extending to the inferior parietal lobule, specifically the supramarginal gyrus. The right frontal cluster comprised the anterior cingulate and superior frontal gyrus, specifically, the premotor region.

Table 2-3. ALE results for functional specialization of symbolic number processing in studies with all relevant contrasts (Meta-analysis I)

Anatomical labels		Talairach coordinates			ALE value ($\times 10^{-3}$)	Volume (mm ³)	Studies with foci within the cluster*
Talairach Daemon	Anatomy Toolbox	<i>x</i>	<i>y</i>	<i>z</i>			
R. inferior parietal lobule (BA 40)	hIP2	40	-48	44	32.75	7680	1, 2, 3, 5, 6, 8, 10, 11, 20, 21, 22, 24, 25, 28, 29
R. inferior parietal lobule (BA 40)	hIP3	34	-46	38	31.33		
R. precuneus (BA 7)		28	-64	38	25.46		
R. precuneus (BA 7)	Area 7P	14	-74	48	23.33		
R. precuneus (BA 19)	hIP3	28	-58	38	22.26		
L. supramarginal gyrus (BA 40)	hIP1	-40	-44	36	34.44	3832	1, 2, 5, 6, 11, 22, 23, 24, 25, 28, 29, 31
L. inferior parietal lobule (BA 40)	hIP1	-36	-54	42	23.61		
R. anterior cingulate (BA 24)	-	18	6	46	16.91	744	1, 6, 7, 8, 18
R. superior frontal gyrus/premotor (BA 6)	-	22	6	60	16.37		

Note. BA: Brodmann Area. L: Left. R: Right. hIP: Horizontal segment of the intraparietal sulcus (with subdivisions 1, 2, and 3). Area 7P: Posterior subdivision of superior parietal lobule. Global peak of each cluster is in bold, and local peaks are not in bold. *Although studies with foci that fall within the boundary of each cluster are listed, it is important to note that studies not listed could also contribute to the ALE scores if they lie on or close to the cluster boundary (Fox et al., 2013).

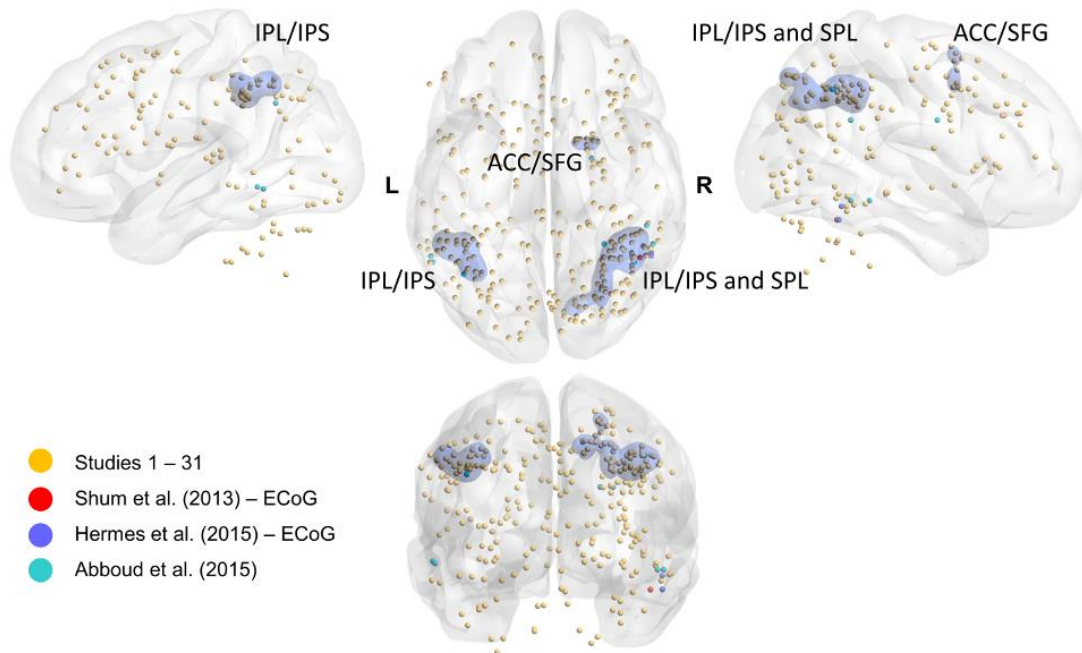


Figure 2-1. Visualization of ALE results in meta-analysis I

ALE concordant clusters (blue blobs) and foci (dots) from all relevant contrasts in meta-analysis I (cf. **Table 2-3**), and from relevant studies excluded from the meta-analysis, overlaid on the Colin27 brain template. Top panel: left lateral hemispheric, top axial, and right lateral hemispheric views. Bottom panel: posterior coronal view. L: Left. R: Right. IPL: Inferior parietal lobule. IPS: Intraparietal sulcus. SPL: Superior parietal lobule. ACC: Anterior cingulate cortex. SFG: Superior frontal gyrus. ECoG: Electrocorticography.

Note. BrainNet Viewer (Xia et al., 2013) was used to visualize the clusters and foci. While some foci appeared to fall outside the template used above, all but 3 foci in the middle frontal region were within the boundaries of the mask used in the meta-analysis.

Meta-analysis II: Suitably Specific Contrasts Only

Controlling for task demands and contrast specificity, we found in Meta-analysis II a frontoparietal network that overlaps with that observed in Meta-analysis I (**Table 2-4** and **Figure 2-2**): A right parietal cluster that comprised the precuneus, anterior superior parietal lobule, and intraparietal sulcus extending to the inferior parietal lobule; a smaller left parietal cluster that comprised the intraparietal sulcus extending to the inferior parietal lobule, specifically the supramarginal gyrus; and a right frontal cluster that comprised the superior frontal (specifically, the premotor region) and anterior cingulate gyri.

Meta-analysis II additionally revealed a right lateral prefrontal cluster at the intersection of the precentral and inferior frontal gyri, specifically, the pars opercularis, and a right temporal cluster comprising the posterior inferior and middle temporal gyri (**Table 2-4** and **Figure 2-2**). The peak of the inferior temporal cluster (TAL 50, -48, -10) is slightly superior to the NFA locations reported by Shum and colleagues (2013) (TAL² 47, -51, -21), as well as Hermes and colleagues (2015) (TAL² 52, -48, -15, and 52, -50, -21), but extremely close to those foci (≤ 3 mm) along the left-right and anterior-posterior dimensions. However, the peak of our cluster is as superior to those found by Abboud and colleagues (2015) (TAL 53, -44, -12, and 50, -35, -12).

To rule out the possibility that the convergence in the inferior temporal cluster was driven mainly by two recent studies (Studies 1 and 9) that employed advanced imaging and data preprocessing techniques to overcome signal loss (Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016), we re-ran the meta-analysis by omitting each of the seven studies that had numeral-specific activation in the fusiform and inferior temporal gyri (see **Table 2-1** and

² Using the icbm2tal_spm.m transform (brainmap.org/icbm2tal) (Lancaster et al., 2007).

Table 2-2). We confirmed that the omission of each of the five studies listed in **Table 2-4** eliminated the convergence in the inferior temporal cluster, which suggests that they were all critical contributors to the convergence.

Table 2-4. ALE results for functional specialization of symbolic number processing in studies with suitably specific contrasts only (Meta-analysis II)

Anatomical Labels		Talairach coordinates			ALE value ($\times 10^{-3}$)	Volume (mm ³)	Studies with foci within the cluster*
Talairach Daemon	Anatomy Toolbox	X	y	z			
R. precuneus (BA 7)	-	16	-74	48	22.23	5536	1, 2, 3, 6, 8, 10, 11, 14, 20
R. inferior parietal lobule (BA 40)	hIP3	34	-46	38	21.63		
R. inferior parietal lobule (BA 40)	hIP2	40	-50	44	21.58		
R. superior parietal lobule (BA 7)	Area 7A	28	-68	46	18.20		
R. inferior parietal lobule (BA 40)	hIP2	46	-40	40	17.05		
R. inferior parietal lobule (BA 39)	hIP3	28	-56	38	16.15		
L. supramarginal gyrus (BA 40)	hIP3	-38	-44	36	22.93	1400	1, 2, 6, 11
L. inferior parietal lobule (BA 40)	hIP3	-36	-50	42	15.21		
L. inferior parietal lobule (BA 40)	hIP2	-48	-48	42	13.04		
L. inferior parietal lobule (BA 40)	hIP2	-50	-44	42	11.94		
R. superior frontal gyrus/premotor (BA 6)	-	22	6	60	16.13	1104	1, 6, 7, 8, 18
R. anterior cingulate (BA 24)	-	16	6	44	14.84		
R. inferior temporal gyrus (BA 37)	-	50	-48	-10	13.95	728	1, 6, 7, 9, 15
R. middle temporal gyrus (BA 20)	-	56	-40	-14	13.86		
R. precentral gyrus (BA 6)	-	42	2	34	16.07	696	1, 6, 20
R. inferior frontal gyrus (BA 9)	pars opercularis	50	4	28	14.70		

	(Area 44)				
R. inferior frontal gyrus (BA 9)	pars opercularis (Area 44)	48	4	24	12.75

Note. BA: Brodmann Area. L: Left. R: Right. hIP: Horizontal segment of the intraparietal sulcus (with subdivisions 1, 2, and 3). Area 7A: Anterior subdivision of superior parietal lobule. Global peak of each cluster is in bold, and local peaks are not in bold. *Although studies with foci that fall within the boundary of each cluster are listed, it is important to note that studies not listed could also contribute to the ALE scores if they lie on or close to the cluster boundary (Fox et al., 2013).

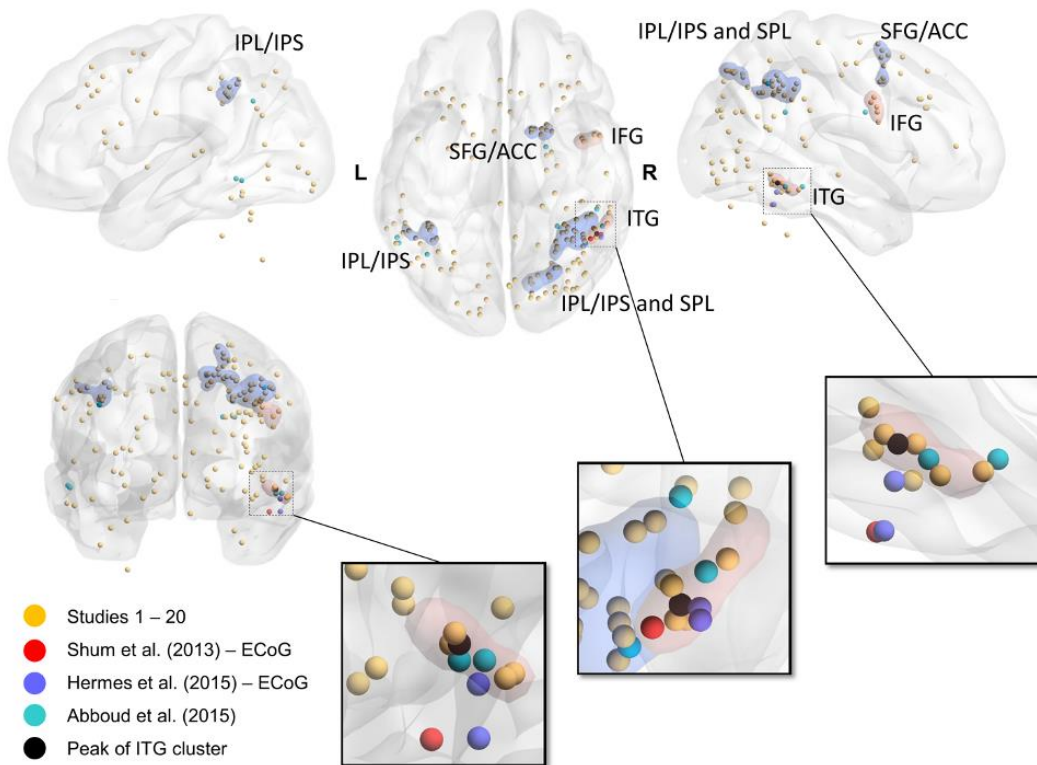


Figure 2-2. Visualization of ALE results in meta-analysis II

ALE concordant clusters (blue and red blobs) and foci (dots) from suitably specific contrasts in meta-analysis II (cf. **Table 2-4**), and from relevant studies excluded from the meta-analysis, overlaid on the Colin27 brain template. The blue clusters overlapped with those observed in Meta-analysis I, and the red clusters are unique to Meta-analysis II. Top panel: left lateral hemispheric, top axial, and right lateral hemispheric views. Bottom panel: posterior coronal view. Inserts (left to right): Magnified coronal, axial, and sagittal views of the right inferior temporal lobe. L: Left. R: Right. IPL: Inferior parietal lobule. IPS: Intraparietal sulcus. SPL: Superior parietal lobule. ACC: Anterior cingulate cortex. SFG: Superior frontal gyrus. IFG: Inferior frontal gyrus. ITG: Inferior temporal gyrus. ECoG: Electrocochography.

Discussion

The current study examined whether the extant neuroimaging literature supports the existence of a reproducibly localized “Number Form Area” in the vOT. By being maximally inclusive in Meta-analysis I, we predicted that we would find a convergence of activation in an NFA in the pITG, especially if its activation was weak due to signal loss complications. In Meta-analysis II, we tested the hypothesis that the exclusion of less specific and overly specific contrasts in the meta-analysis may increase the statistical sensitivity of the convergence in an NFA. Only in Meta-analysis II did we observe such a numeral-preferring region in the right pITG. In addition to a candidate NFA in the right pITG, the bilateral inferior parietal regions, and a right-lateralized network of superior parietal, and superior and inferior frontal regions also appear to be functionally specialized for processing Arabic numerals. This “numeral processing network” possibly supports perceptual and cognitive processes engaged specifically by Arabic numeral processing, and not by differences in task demands between the numerical and non-numerical conditions. To our knowledge, this is the first meta-analysis to quantitatively review the existence of a reproducibly localized NFA, and its associated network of brain regions.

Reproducible Localization of an NFA in the Posterior Inferior Temporal Gyrus

Previous meta-analyses (Arsalidou & Taylor, 2011; Sokolowski et al., 2017) suggest the involvement of a region in the fusiform gyrus for processing of numerical symbols, but the convergence in the fusiform gyrus may be driven not by Arabic numerals exclusively, but also by written number words or nonsymbolic quantities (e.g., dot arrays) included in the contrasts meta-analyzed. On the contrary, our findings are specific to Arabic numerals and thus provide some degree of support for the existence of a reproducibly localized NFA in the pITG that is spatially distinct from the VWFA, whose canonical location is more medial in the mid-fusiform

gyrus and along the occipitotemporal sulcus (Hannagan et al., 2015). This is consistent with a series of recent reports of pITG as a more plausible location of an NFA that were not included in the current meta-analyses (Abboud et al., 2015; Daitch et al., 2016; Hermes et al., 2017; Roux et al., 2008; Shum et al., 2013).

Notably, only five out of 20 studies in Meta-analysis II contributed to the convergence in the pITG cluster. Out of those five, only two (Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016) employed advanced imaging and data preprocessing techniques to overcome signal loss. We also found that the meta-analytic convergence in the pITG was not driven solely by those two studies. Lastly, scrutiny of recent studies (Abboud et al., 2015; Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016) suggest a lack of strong and direct evidence that an NFA in the pITG cannot be functionally localized. Hence, detection of an NFA in the pITG is not impossible with standard fMRI acquisition and data preprocessing. Our findings therefore suggest that while signal dropout might have contributed to some previous null findings using fMRI, it is not a reliable contributor and perhaps not the only factor influencing a study's ability to detect an NFA in the pITG.

A factor worth considering is the role that task contexts play in the detection of numeral-specific activation. The fact that we observed a convergence of activation in the pITG only after controlling for task demands and contrast specificity suggests that the cognitive demands of the numerical condition and/or the non-numerical condition, as well as the choice of contrasts that researchers choose to analyze may influence whether an NFA can be detected at the study level.

Alternatively, the statistical thresholding criteria used could also partly account for the mixed findings of an NFA. Among the 12 studies in **Table 2-1** that observed numeral-specific activity in the vOT, equal proportions of studies did not employ correction for multiple

comparisons or apply an arbitrary spatial extent threshold (33.3%), or employed voxelwise familywise error (FWE) or false discovery rate (FDR) correction (33.3%), or employed cluster-level correction (33.3%). However, among 19 studies that only observed numeral-specific activity outside the vOT in **Table 2-1**, there is a trend that more stringent cluster-level (52.6%) and voxelwise FWE/FDR correction techniques (31.6%) were employed.

It is important to note that while the concordant activation in the right pITG in the present work suggests that there is reproducible localization of a region that shows a preference for numerals, it is still insufficient evidence to demonstrate that it is truly an NFA. Its validity as an NFA hinges on the fulfilment of the additional criteria of partial regional selectivity and functional specialization (L. Cohen & Dehaene, 2004; Dehaene & Cohen, 2011). Based on the hypothesis of the “Visual Number Form” being an input stage common to any task involving numerals (L. Cohen & Dehaene, 1991), it should also be task-independent.

Partial Regional Selectivity

There is evidence that the numeral-preferring region responds more to Arabic digits than to Roman letters and novel characters including scrambled digits and characters from a foreign script, and to double-digit numerals than to words (e.g., Grotheer, Herrmann, et al., 2016; Shum et al., 2013). However, the sampling space of character types as control categories is currently still limited, and it is not clear whether it also responds preferentially to other types of characters used in the mathematical domain. For instance, does it care about other non-alphanumeric symbols such as \$, +, =, and % that are typically seen together with Arabic numerals? Polk and Farah (1995, 1998) hypothesized that letter and digit processing could be neural segregated merely on the basis of the frequency of their co-occurrence in our natural environments. That is, letters occur more frequently than digits (e.g., in this text), and a letter is seen more often with

other letters than with digits, both spatially and temporally. Hence, the frequency of co-occurrence may underlie neural segregation via a Hebbian learning mechanism (Polk & Farah, 1995, 1998). If so, it is possible that the so-called NFA could respond to other mathematical symbols as much as it does to Arabic numerals. For instance, Amalric and Dehaene (2016) found that the left NFA in professional mathematicians showed preferential response to mathematical formulas and expressions (e.g., $\zeta(z, w) = \sum_{n=1}^{\infty} \frac{1}{(w+n)^z}$) in addition to Arabic numerals compared to non-mathematicians.

Alternatively, the so-called NFA could even be a general “Symbol Form Area” that happens to prefer numerals more than letters, though not specialized for numerals. This is supported by a recent transcranial magnetic stimulation study by Grotheer, Ambrus, and Kovács (2016) in which they found that stimulation of the right NFA impaired the identification of both digits and letters, but not scrambled digits and letters. This led the authors to suggest that while the NFA plays a causal role in numeral processing, it might be more of “a flexible processing module for the early visual encoding of learned characters” than a numeral-specific one (Grotheer, Ambrus, et al., 2016, p. 317). Given the limited range of control stimuli considered within each study in our meta-analyses, and that four of five studies contributing to the pITG cluster only included letters as the only other learned character category, we believe that partial regional selectivity has not been adequately demonstrated across existing fMRI studies.

Functional Specialization

Originally conceived as analogous to the “Visual Word Form” (L. Cohen & Dehaene, 1991), the “Visual Number Form” is thought to represent numerals at the level of character strings and “encodes several visuo-spatial characteristics of the stimulus, including the shape and the relative positions of the digits, presumably in a stimulus-centred co-ordinate system” (p. 54).

Therefore, similar to the tuning of the VWFA to the orthographic computational requirements of reading in a given script (L. Cohen & Dehaene, 2004; Dehaene & Cohen, 2011), an NFA should be attuned to the computational requirements of the Arabic numeral system. Beyond merely distinguishing numerals from non-numerals, an NFA should be sensitive to individual digit shapes, digit identities, numeral length, order of digits, and syntactic markers including 0 as a placeholder and 1 for teens (L. Cohen & Dehaene, 1991). A complete characterization of the NFA's functional specialization for Arabic numerals has not been systematically investigated.

Task-independent Recruitment

The hypothesis that an NFA is tuned to the visual form or shape of Arabic numerals (Hannagan et al., 2015; Shum et al., 2013) suggest that the localization of an NFA should be task-independent, even when numerals are passively viewed. However, our results suggest otherwise. Hence, another issue which requires further empirical investigation is the role of task-dependent, top-down, goal-directed processes in the preferential engagement of the NFA.

For instance, using intracranial electrophysiology, Hermes and colleagues (2017) found that responses in the NFA were heightened by increased cognitive demands during active calculation relative to numeral reading and an oddball color detection task involving Arabic digits in the background. Interestingly, the modulation of the NFA responses was not numeral-specific, but extended to calculation using number words, suggesting that cognitive demands of calculation, rather than visual perception alone, influenced the NFA activity (Hermes et al., 2017). More recently, Daitch and colleagues (2016) found feedback-based functional coupling between the PITG and the anterior intraparietal sulcus (IPS). Specifically, they found that the NFA was involved in bottom-up coupling with the IPS during passive processing of Arabic numerals, and the IPS was involved in top-down coupling with the NFA during arithmetic

(Daitch et al., 2016). These findings suggest that the bottom-up, sensory processes of an NFA may be highly modulated by top-down, semantically related processes specific to the tasks.

In sum, further research is necessary to characterize the regional selectivity, functional specialization, and the degree of task-independence of an NFA. The pITG cluster found here could serve as an a priori candidate region for attempts at such characterizations, especially if the functional localization of an NFA remains challenging due to reasons other than signal dropout.

Beyond An NFA: The “Numeral Processing Network”

In addition to a candidate NFA in the pITG, other brain regions also appear to exhibit preference for Arabic numeral processing, including the bilateral inferior parietal lobules (IPL) and intraparietal sulci (IPS), a right-lateralized collection of regions in the precuneus, superior parietal lobule (SPL), premotor region of the superior frontal gyrus (SFG), anterior cingulate, and inferior frontal gyrus (IFG). Given that we controlled for task demands in Meta-analysis II, and that digits are inherently easier to identify than letters even when perceptual information is limited (Starrfelt & Behrmann, 2011), our findings suggest that the frontoparietal network is possibly involved in additional semantic, syntactic, and lexical processing that single- and multi-digit Arabic numerals may require beyond what other symbol sets do.

As noted by Price and Ansari (2011), and Starrfelt and Behrmann (2011), single Arabic digits contain richer semantic associations than most other single characters do. Not only can single digits be associated with a nominal identity (e.g., jersey number “7”), they can also be associated with magnitude (e.g., “4” represents [::]) and ordinality (e.g., “3” is after “2”, and before “4”). In contrast, alphabetic characters (e.g., letters) are mostly only meaningful in strings. Furthermore, processing multi-digit Arabic numerals may involve a higher-order interaction of syntactic or ordinal, semantic, and lexical processing, possibly accounting for the greater

activation of a frontoparietal network. For instance, as noted by Pinel and colleagues (2001), the place-value rules of the base-10 decimal system render the order of digits in a digit string to be particularly important for the extraction of its semantic and lexical content (e.g., the digit “2” in “2”, “12”, and “21” require mappings onto different number words [“two”, “twelve”, or “twenty”] and magnitude representations [two ones or two tens]). Although alphabetic/syllabic number words (e.g., English and Japanese Kana), and logographic number words (e.g., Chinese and Japanese Kanji) may seem similar to Arabic numerals in their semantic, syntactic, and lexical content, each number word is less conceptually rich than Arabic digits. For instance, “23” is written in Chinese as “二十三”, which reads “two-ten-three”. Hence, “二” is always associated with just “two” regardless of its place value. Likewise, the word “twenty” is always associated with “two tens”. In other words, Arabic digits have one-to-many mappings, whereas number words mostly have one-to-one or one-to-few mappings. Moreover, recent behavioral findings by Hurst, Anderson, and Cordes (2016) suggest that preschool children may first map verbal number words onto nonsymbolic quantities, then map Arabic numerals onto verbal number words, such that the association between Arabic numerals and nonsymbolic quantities may be indirect. Such indirect mappings may inevitably require greater cognitive processing. Taken together, Arabic numerals may possibly require a greater degree of syntactic, semantic, and lexical processing than non-numerals do, and in doing so recruit frontal and parietal regions collectively to subserve various domain-specific and domain-general cognitive processes. Below, we speculate what specific processes might be subserved by the parietal and frontal regions.

Parietal Regions

The localization of an NFA in the pITG has been hypothesized to be a result of biased intrinsic anatomical and functional connectivity between the bilateral NFAs and magnitude

processing regions in the bilateral IPS (Abboud et al., 2015; Hannagan et al., 2015). As more than half of the studies that contributed to each IPS/IPL cluster involved numerical tasks that did not require explicit magnitude processing (e.g., repetition detection or naming), the convergence of activation in the bilateral IPS may reflect the possibility that Arabic numerals automatically activate magnitude representations more than non-numerical symbols do (Cohen Kadosh, Cohen Kadosh, Schuhmann, et al., 2007; Girelli et al., 2000; Henik & Tzelgov, 1982; Rubinsten et al., 2002; although see Cohen, 2009; Naparstek & Henik, 2010; Price & Ansari, 2011 for caveats).

In addition to magnitude processing, the bilateral IPS, particularly the anterior portion, may be involved in ordinal processing (Ansari, 2008). In fact, the other parietal and frontal regions in the “numeral processing network” have also been implicated in ordinal processing of Arabic digits (Fias et al., 2007; Fulbright et al., 2003; Ischebeck et al., 2008; see Lyons, Vogel, & Ansari, 2016, for a review).

The left IPL, specifically the supramarginal gyrus, supports a myriad of cognitive processes such as orthographic-to-phonological conversion (Price, 1998), phonological working memory (Church et al., 2011; Paulesu et al., 1993), symbol-referent semantic associations (Grabner et al., 2013; K. K. Kim et al., 2011), as well as temporal order processing (Ortuño et al., 2002; Wiener, Hamilton, et al., 2010; Wiener, Turkeltaub, et al., 2010). The adjacent left angular gyrus is also often implicated in verbal and semantic processing of symbolic numbers (Dehaene et al., 2003; Price & Ansari, 2011; Seghier, 2012). All in all, the concordant activation in the IPL in the current study may reflect a confluence of lexical, semantic, and ordinal processing of Arabic numerals.

The bilateral posterior SPL, including the precuneus, are possibly involved in attention and visuo-spatial orientation on an abstract “mental number line” (Cavanna & Trimble, 2006;

Dehaene et al., 2003; Hubbard et al., 2005; Lyons & Ansari, 2009). It is conceivable that similar mechanisms are involved in attending to the length of a digit string to extract the syntactic verbal frame, and then to the specific order of digits to fill the verbal frame (e.g., “243” → “two hundred forty-three”). Hence, the convergence of activation in the right SPL may reflect visuo-spatial attention mechanisms related to ordinal and syntactic processing of Arabic numerals.

Frontal Regions

Along with the superior parietal regions, the SFG and IFG may also play a role in selective visuo-spatial attention (Anderson et al., 2007). The convergence of activation in the frontal regions may thus also reflect visuo-spatial attention mechanisms related to ordinal and syntactic processing of Arabic numerals.

Alternatively, converging evidence from human and non-human primate studies, has led to the proposal that the IFG may be directly involved in number processing (Nieder, 2009). For instance, Diester and Nieder (2007) found that neurons in the prefrontal cortex of non-human primates (presumably homologous to Brodmann Area 44/45 in humans) are tuned to both symbolic and nonsymbolic numerical magnitudes after learning to associate Arabic digits with their corresponding quantities. These findings led the authors to hypothesized that the IFG may facilitate the associations between arbitrary shapes (e.g., Arabic digits) and quantities. Findings that the IFG is activated to a greater extent in children than in adults during notation-independent numerical magnitude processing also support its role during the formative stage of the symbol-referent associations (Ansari et al., 2005; Cantlon et al., 2009; Kaufmann et al., 2006; Rivera et al., 2005). There is also growing evidence that the IFG may play a more fundamental role in magnitude processing (Damarla et al., 2016; Sokolowski et al., 2017; Tang et al., 2006). For instance, using a multi-voxel pattern analysis, Damarla, Cherkassky, and Just (2016)

demonstrated that several parietal and frontal regions, including the right IFG, support cross-modality (dots and tones) classification of magnitude representations, which suggest the role of the IFG in an abstract code for magnitude. Additionally, the IFG, or prefrontal cortex more generally, may be critical for establishing syntactic representations in human symbol systems, such as grammar, ordinality, and place-value notational rules (Friederici, Bahlmann, Heim, Schubotz, & Anwender, 2006; Meyer, Obleser, Anwender, & Friederici, 2012; see Nieder, 2009, for a review). Therefore, in the current study, the observed convergence in the IFG may reflect its multiple roles in mapping Arabic digits to their magnitude representations, magnitude processing, and higher-order integration of the semantic, syntactic and lexical information.

In summary, in combination with the extant literature, our current findings suggest that the bilateral parietal regions (IPS and IPL), and a right-lateralized collection of regions comprising the SFG, IFG, and pITG form a potential “numeral processing network”. We hypothesize that the pITG may be involved in asemantic shape processing of numerals; the bilateral IPS and IFG in magnitude processing; bilateral IPL in lexical, semantic, and ordinal processing; right SPL and SFG in visuo-spatial processing of the perceptual, syntactic and abstract ordinal relations among the symbols; and the right IFG in higher-order integration of all task-relevant information. Further empirical investigations are necessary to systematically dissociate the specific cognitive processes involved, their underlying mechanisms, the neural subdivisions supporting each function, and their interactions.

Finally, although we use the term “numeral processing network” here, the extent to which this collection of regions constitutes a true interdependent *network*, as opposed to simply a group of simultaneously activated but functionally independent regions, is yet to be determined. Thus far, some degree of temporoparietal (Abboud et al., 2015; Daitch et al., 2016; Park et al., 2012)

and frontoparietal coupling have been demonstrated during numerical symbol processing (Diester & Nieder, 2007). However, it is still unclear how independent or interdependent those two networks are. Moreover, as shown in **Table 2-1**, the findings for functional specialization for Arabic numerals at large have been very inconsistent, and it is unclear why many studies did not find a single region that is numeral-selective. In sum, given the hypothesized complex interaction of semantic, syntactic, and lexical processing engaged by Arabic numerals, the careful selection of numerical and non-numerical tasks may be critical for characterizing the functional mechanisms underlying the different nodes of the “numeral processing network”.

Limitations

As with any ALE meta-analysis, the current study lacked the means to statistically control for differences in statistical thresholding methods across studies, spatial extent and magnitude of the activation foci, or confounding variables such as age. It is also important to note that unlike a conventional behavioral meta-analysis that addresses the question of whether there is an effect at all, ALE addresses the question of whether there is convergence across studies given the premise that there is some condition-specific activation somewhere in the brain. Hence, studies that did not find any numeral-specific activity anywhere in the brain (11 studies in **Table 2-1**) had to be excluded. This bias therefore warrants caution in interpreting our findings as definitive about the existence of an NFA.

One limitation specific to this study is the inclusion of uncorrected foci (16% of total number of foci). However, this was deliberate as the primary goal of the study was to detect an NFA, which has supposedly been plagued by signal loss complications. With signal attenuation, any activation may be less likely to survive correction for multiple comparisons, and so liberal inclusion criteria were essential.

Conclusions

The present findings suggest that if an NFA does exist, the most reproducible localization is in the right pITG. However, such meta-analytic convergence was only evident when we considered contrasts with tasks demands between the numerical and non-numerical control conditions that were appropriately controlled. Importantly, several of the studies contributing to this region did not employ methods designed to overcome signal dropout, suggesting that signal dropout was not the only factor underlying previous null fMRI findings. Given that only five studies out of more than 40 studies have foci within the pITG cluster boundary, more evidence is necessary to characterize the functional specialization, regional selectivity, and task independence of an NFA. In addition to an NFA, the current study revealed a candidate “numeral processing network” consistent with the extant literature. It comprises the bilateral parietal regions, and right-lateralized superior and inferior frontal regions. While they may be involved in numerical magnitude processing, and domain-general processes related to ordinality, syntax, lexicon, and symbol-referent associations, their roles specific to Arabic numeral processing require further empirical investigation. The present work thus provides insights for understanding the neurocognitive mechanisms that support the processing of Arabic numerals. Such insights are critical for future research as we seek to understand how such systems develop, and the role they play in the typical and atypical development of mathematical skills.

Authorship Contributions

Darren J. Yeo: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Eric D. Wilkey:** Writing - review & editing. **Gavin R. Price:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision.

CHAPTER 3

PROBING THE REPRESENTATIONAL CONTENT IN THE “NUMBER FORM AREA” DURING PASSIVE VIEWING: A REPRESENTATIONAL SIMILARITY ANALYSIS

This chapter is adapted from “The “Inferior Temporal Numeral Area” distinguishes numerals from other character categories during passive viewing: A representational similarity analysis” published in *NeuroImage*, and has been reproduced with permission of the publisher and my co-authors, Courtney Pollack, Rebecca Merkley, Daniel Ansari, and Gavin R. Price.

Yeo, D. J., Pollack, C., Merkley, R., Ansari, D., & Price, G. R. (2020). The “Inferior Temporal Numeral Area” distinguishes numerals from other character categories during passive viewing: A representational similarity analysis. *NeuroImage*, 214, 116716.
<https://doi.org/10.1016/j.neuroimage.2020.116716>

Introduction

Recent fMRI studies following the publication of the Chapter 2 (Yeo et al., 2017) provided support for our conjectures regarding the lack of strong evidence that signal dropout in the ITG underlies the lack of reliable localization of a putative NFA. Other contributing factors such as task demands may also underlie the mixed findings of an NFA.

First, Grotheer and colleagues (2018) managed to localize the left and/or right numeral-preferring pITG in most individual subjects in their study (means of 13 out of 15 subject-specific peaks: MNI -54, -59, -12, and 57, -54, -14, compared to the meta-analytically identified NFA peak in Chapter 2 (Yeo et al., 2017): MNI³ 55, -50, -12) even after abandoning all of the advanced acquisition parameters⁴ that were believed to have contributed to the successful

³ Using the tal2icbm_spm.m transform (brainmap.org/icbm2tal) (Lancaster et al., 2007).

⁴ Compared to a 64-channel head coil with 1-mm isotropic voxel resolution, localized shimming, and liberal spatial smoothing 8 times the voxel resolution employed in Grotheer, Herrmann, et al. (2016), the more recent study employed a 32-channel head coil with 2.4 mm isotropic voxel resolution, no localized shimming, and no spatial smoothing.

localization in an earlier study by the same lead author (Grotheer, Herrmann, et al., 2016). They further showed that the NFA consistently falls posterior to, not within, the signal dropout zone in all of their participants. More recently, although Merkley and colleagues (2019) was unsuccessful in localizing a group-level NFA with a passive-viewing paradigm, they found minimal or no attenuation in temporal signal-to-noise ratio in the meta-analytically identified NFA (Chapter 2; Yeo et al., 2017) in individual subjects. In sum, in cases where there was signal attenuation in the ITG, they often lay anterior to the candidate NFA, so standard fMRI sequences should allow an NFA to be localized.

If not signal dropout, what could underlie the previous lack of localization of an NFA in the pITG? Task demands for both the numeral and non-numeral control conditions appear to be a likely factor. Grotheer and colleagues (2018) replicated previous findings (Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016) that, during a repetition detection (i.e., one-back) task, regions in the bilateral pITG demonstrated numeral preference. They probed the function of these regions and showed that, during an addition task using numerals (e.g., $2 + 3 = 5?$), dice patterns, or finger representations, these very same regions were not more engaged during addition with numerals than addition with dice patterns and finger representations (Grotheer et al., 2018). The authors concluded that the “numeral-preferring” pITG regions initially identified are not involved in processing the visual form of the numerals because they should otherwise show consistent numeral preference regardless of the task. Instead, Grotheer and colleagues (2018) hypothesize that the neuronal populations in the pITG predominantly “ascribe numerical content to the visual input” (p. 188) (see also Abboud et al., 2015). However, there is also evidence that an NFA is involved in non-quantitative contexts, such as whether a character is familiar or novel (Grotheer, Ambrus, et al., 2016), whether a character is read aloud (Shum et al.,

2013), or whether a character is repeated (Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016). So, associating numerical content with the stimulus input cannot fully account for the recruitment of a numeral-preferring pITG region.

In another fMRI study, Pollack and Price (2019) found that a region in the left pITG (group-level peak: MNI -57, -52, -11) was preferentially engaged for numerals when participants had to detect a digit amongst a string of letters, but the same region showed no numeral preference when participants had to detect a letter amongst a string of digits (i.e., when the digits were task-irrelevant). Taken together, whether the candidate region for an NFA shows greater engagement for numerals than other visual object categories may be highly dependent on attention to the stimulus identity and/or category, as well as task contexts. Thus, the exact computational mechanisms subserved by an NFA remain opaque. Hereafter in this chapter, I will use the term “numeral-preferring pITG” instead of NFA to avoid making an assumption about its role in processing “number form”.

In light of these new insights on the sensitivity of the numeral-preferring pITG region to task demands, we questioned whether some previous null findings for numeral preference in the pITG were a consequence of employing a passive-viewing paradigm, as was used in the first fMRI study to explicitly investigate the existence of an NFA (Price & Ansari, 2011). Specifically, Price and Ansari (2011) used a fixation-color change-detection task, in which participants were asked to respond when a white hash sign (#), turned red, but not when it turned to another character such as letters, digits, and novel characters (see **Figure 3-1**). Such a passive-viewing paradigm is in some ways ideal, because it disentangles automatic, stimulus-driven sensory processing from any effortful, task-driven conceptual or semantic processing of the numerals (Kay & Yeatman, 2017). This study, as well as two replication attempts by the same

and additional authors (Merkley, Conrad, Price, & Ansari, 2019; Price & Ansari, unpublished data), found no evidence of a numeral-preferring region anywhere in the vOT using univariate activation analyses. Although it is possible that a passive-viewing paradigm may not be optimal for investigating the NFA's role in sensory processing of numerals, it has been used successfully to reveal letter- and word-preferring regions in the vOT (e.g., Cantlon et al., 2011; Dehaene-Lambertz et al., 2018; Glezer et al., 2009; Karipidis et al., 2017; Kay & Yeatman, 2017; Parviainen, 2006; Pleisch et al., 2019; Polk et al., 2002; Vinckier et al., 2007; B. Zhang et al., 2018), and is recommended for understanding models of experience-driven neural coding in the vOT (Dehaene & Cohen, 2011). Moreover, having an explicit task requires participants to attend specifically to the visual form of the characters (e.g., repetition detection task) or to its category (e.g., whether a character is familiar or can be named, whether a digit or letter is present). This may bias the neural responses towards visual form and symbol category respectively through goal-directed modulation (Kay & Yeatman, 2017). Hence, stimulus type is confounded with task goal, rendering any interpretation of the neural representation of a stimulus difficult. In other words, an active task would not decisively inform us whether the region represents digit shapes. Despite the merits of a passive-viewing paradigm, neural responses to task-irrelevant characters may not be highly discriminable in terms of their categorical membership simply by examining the voxel-wise activation or response strength averaged across exemplars. Even if the overall response strengths across digits, letters, and novel characters do not differ in a candidate numeral-preferring pITG region, they may show distributed activation patterns that reveal categorical distinctions. Hence, multivariate pattern analyses may be more sensitive than univariate mean response analyses for examining the neural representations of task-irrelevant characters.

Current Study

In this study, we amassed three passive-viewing datasets mentioned above (Merkley et al., 2019; Price & Ansari, 2011; Price & Ansari, unpublished data), and used multivariate representational similarity analysis (RSA) to probe the spontaneous (i.e., task-irrelevant) organization of neural representations of single digits, letters, and novel characters in a candidate numeral-preferring pITG region. This candidate region (hereafter, “pITG-numerals”) is derived from the meta-analysis of studies contrasting numerals and other symbols in Chapter 2 (Yeo et al., 2017). By examining how similar or dissimilar the neural representations of individual characters within and between categories are, we can characterize the organization of the representations, or the “representational geometry”, in a particular neural region, and assess whether the observed representational geometry can be described by hypothetical functional models (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte & Kievit, 2013; Nili et al., 2014). If pITG-numerals is specialized for numeral processing, the representations of digits in the region should be similar to one another, but not to letters and novel characters, which ought to be similar amongst themselves (i.e., {digits} vs. {letters and novel characters}). If pITG-numerals is specialized for visual form processing, its representational geometry should be biased towards similarities in shape without any categorical distinctions. For instance, “5” and “S” may be represented similarly in this region, with “5” being more similar to “S” than it is to “4”. Alternatively, representations of shape and category may not be mutually exclusive in the vOT (Bracci, Ritchie, et al., 2017; Bracci & Op de Beeck, 2016), and both types of information may be coded in pITG-numerals. To synthesize findings from the three datasets, we also performed small-scale meta-analyses on the effect sizes. Finally, to distinguish between *evidence of absence* of an effect and the *absence of evidence* for an effect, we performed complementary Bayesian

analyses for the individual datasets as well as the meta-analyses.

It is possible that pITG-numerals may not show greater numeral sensitivity, but it is still important to understand if the region is at least sensitive to some other distinctions between the character categories (e.g., distinguishing between numerals and novel characters, or numerals and letters) using RSA. If this region distinguishes numerals from novel characters, but not numerals from letters, the region is possibly sensitive to familiarity of the characters. If it is also capable of distinguishing numerals from letters, it suffices as evidence that this region responds to numerals and letters differentially even though prior univariate activation analyses had been unable to detect that. Hence, we also explored more nuanced representational geometries in the region (e.g., familiar vs. novel).

Methods

Task and Datasets

Task

In each study, participants completed an identical fixation-color change-detection task (see symbol sets and example trials in **Figure 3-1**). They were instructed to fixate on a centrally positioned white hash symbol (#) on a black background, and to press a button whenever the hash changed from white to red. Participants were also informed that the white hash could change into another character, which was always white, but no response was required for those changes. The order of the task-irrelevant characters and the change target (red hash) was randomized or pseudorandomized within each run. In each run, depending on the dataset, each character was presented either 2 or 4 times, and the target was presented either 6 or 8 times (see **Table A-1** for more details). There are substantial differences in scan acquisition protocols and design parameters (e.g., additional factorial conditions examined) (see “Differences in

Experimental Contexts and Neuroimaging Acquisition Parameters” Section, and **Table A-1**).

Dataset 1 (Price & Ansari, 2011)

Participants were 19 right-handed adults (6 females, $M_{Age} = 22.2$ years, $SD = 1.7$, range = 20.5 – 27.2). All participants gave informed consent and the research procedures were approved by the Health Sciences Research Ethics Board at the University of Western Ontario.

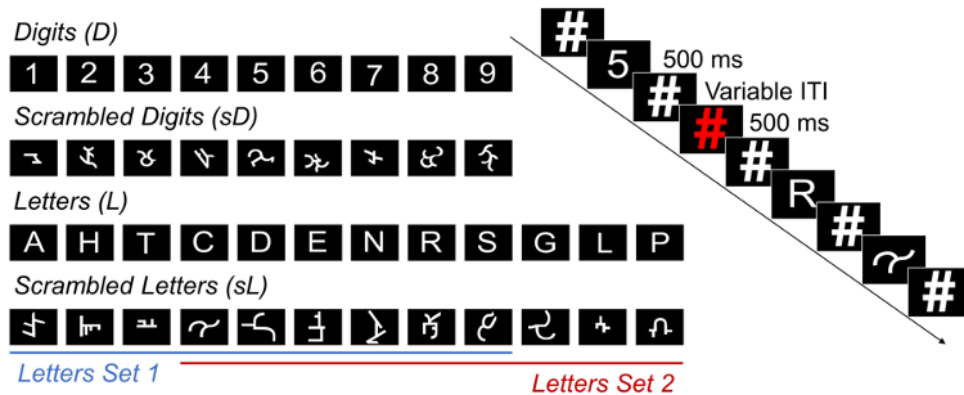


Figure 3-1. Stimulus sets and a schematic of the fixation-color change-detection task. The current study analyzed only the trials with these sets of stimuli that were only presented for 500 ms. ITI: Inter-stimulus interval. Letters Set 1 was used in Datasets 1 and 2. Letters Set 2 was used in Dataset 3.

Dataset 2 (Price & Ansari, unpublished data)

Participants were recruited from a large-scale longitudinal study of mathematical development (Mazzocco & Myers, 2003). When the cohort reached Grade 12, a subset of the cohort was recruited to participate in a neuroimaging study that included the fixation change-detection task reported here. Other tasks such as magnitude comparison and arithmetic

verification that were also conducted during this study have been reported elsewhere (Price, Mazzocco, & Ansari, 2013; Wilkey, Barone, Mazzocco, Vogel, & Price, 2017). A total of 32 participants had usable data for the task reported here (13 females, $M_{Age} = 17.8$ years, $SD = 0.4$ years, range = 17.1 – 18.8 years, handedness: 3 left, 28 right, 1 unknown). Seven additional participants were excluded due to head motion (see “Data Exclusion” Section). All participants gave informed consent and the research procedures were approved by the Institutional Review Board at Johns Hopkins University.

Dataset 3 (Merkley et al., 2019)

Participants were 37 right-handed adults (26 females, $M_{Age} = 25.1$ years, $SD = 5.9$, range = 18 – 39). All participants gave informed consent and the research procedures were approved by the Health Sciences Research Ethics Board at the University of Western Ontario. Based on an a priori right-handedness requirement for study eligibility – to be consistent with Price and Ansari (2011) – three participants were not included as they did not disclose in advance that they were left-handed. Three additional participants were excluded due to a lack of information about task performance as no button responses were recorded. Of the 37 participants in the final sample, data of one run each from two participants were excluded due to poor task performance, and data of one run each from three participants were excluded due to head motion (see “Data Exclusion” Section). In other words, 5 of 37 participants had usable data from only three runs.

Data Exclusion

Data were excluded based on two criteria – behavioral performance and head motion – and were applied uniformly to all 3 datasets. We excluded runs with less than 50% task accuracy based on errors of omission, which served as an indication of task engagement. Given our interest in the activation patterns evoked by each character, we also excluded runs with more

commission errors (i.e., making a button response to a non-target) than there were targets (e.g., in Dataset 3, two participants made 18 commission errors in one run even though there were only six targets). **Table A-2** summarizes the frequency of omission and commission errors in each dataset. In all datasets, each character of interest had at least one usable trial for the estimation of beta weights. Runs in which the participant's head movement exceeded a displacement of 3 mm over the course of the run and/or a volume-to-volume displacement of 1 mm were excluded from analyses.

Stimulus Sets

The stimuli were grayscale images with single white characters against a black background and were presented using E-Prime 2 (Psychology Software Tools, Inc., Pittsburgh, PA, USA) (**Figure 3-1**). There were four categories of characters, with nine exemplars each: (1) Digits: Arabic digits (1 – 9), (2) Letters: uppercase Roman letters (A, C, D, E, H, N, R, S, and T in Dataset 1 and Dataset 2 [hereafter, Letters Set 1]; C, D, E, G, L, N, P, R, and S in Dataset 3 [hereafter, Letters Set 2]), (3) Scrambled Digits: scrambled counterparts of the digits set, and (4) Scrambled Letters: scrambled counterparts of the letters set. The intact symbol sets were in Arial font (size 40), and the scrambled digits and letters were obtained by segmenting and rearranging the parts into a unified, but novel curvilinear shape. The average visual angles for each condition are reported in **Table A-3**. Below, we provide further characterization of the stimulus sets so as to rule out other low-level visual differences between any of these categories that the pITG may be sensitive to.

Based on previous work by Schubert (2017), we focused on two low-level visual parameters that may underlie any categorical differences between digits and letters: luminance and perimetric complexity. Luminance was chosen because Arial font is a proportional-width

font – its characters do not take up the same horizontal space. Hence, the digits set takes up less horizontal space than the letters set. Luminance was computed by summing the intensity values of all pixels in each grayscale image. As the scrambled characters appeared to be more visually complex than their intact counterparts, we wanted to quantify their complexity. Perimetric complexity is commonly used to measure the size-invariant visual complexity of individual characters (Pelli et al., 2006; Schubert, 2017; Shovman & Ahissar, 2006; Ziegler et al., 2010), and has been shown to be highly correlated with the efficiency of character identification and is mediated by the number of features (e.g., lines, curves, terminations, etc.) (Pelli et al., 2006). We computed the perimetric complexity of each character using the approach described by Pelli and colleagues (2006) with a custom MATLAB script: squared length of the inner and outer perimeter divided by “inked” area of each shape traced from the binarized version of the image. Pairwise comparisons of luminance and perimetric complexity showed that digits, letters and their scrambled counterparts did not differ substantially in their perimetric complexity, however, digits on average had lower luminance than letters in both Letters Sets 1 and 2 (see **Table A-3** and **Table A-4**). Given the difference in luminance between the digits and letters sets, we directly assessed whether the representational geometry in a region can be described by differences in luminance.

Differences in Experimental Contexts and Neuroimaging Acquisition Parameters

Besides practical differences in MRI acquisition parameters with different scanner models (see **Table A-1**), there are notable differences in the amount and nature of exposure to the stimuli.

Additional Conditions Within Each Run

Within each run in Dataset 1 (Price & Ansari, 2011), each character was presented twice

for a duration of 500 ms each, and twice for a duration of 50 ms. As the 50-ms condition evoked a much smaller signal change above a fixation baseline compared to the 500-ms condition across the character categories in Price and Ansari (2011), this condition was not included in the replication Dataset 2 and Dataset 3 (Merkley et al., 2019). In Dataset 3, the authors replaced that condition with a mirrored image condition, in which the intact digits and letters were flipped horizontally, also presented for a duration of 500 ms. In Dataset 2, the 50-ms condition was not replaced with a different condition, hence the run was the shortest among the three datasets. Analyses in this study were restricted to the 500-ms condition for intact digits and letters, and their scrambled counterparts, which were common to all three datasets.

Number of Runs

Dataset 1 had two runs, Dataset 2 had one run, and Dataset 3 had four runs.

Inter-trial Interval

The inter-trial interval (ITI) in Dataset 3 (1 – 3 s) was substantially shorter than that in Datasets 1 and 2 (4 – 8 s) due to the shorter acquisition time per volume (Merkley et al., 2019). Perceptually and cognitively, the task might appear very different between short and long ITIs. In terms of the analysis of neural responses, there is some evidence that ITIs less than 6s are sub-optimal for modeling single-trial responses in multivariate pattern analyses (Abdulrahman & Henson, 2016; Visser et al., 2016; Zeithamova et al., 2017). Single-trial responses are more commonly analyzed for classification analyses, but less so for RSA, in which exemplar-level responses (modeled across multiple trials featuring the same exemplar) are more commonly analyzed. Moreover, it is not uncommon for multivariate pattern analyses to be applied successfully to event-related designs with ITIs shorter than 2 s (1.7 - 1.9 s in Borghesani et al., 2016; 1.5 s in Bracci, Daniels, & Op de Beeck, 2017, and Bracci & Op de Beeck, 2016). To

mitigate this concern and to yield more reliable estimates of response patterns, we modeled each character with a single regressor across the repeated presentations *across runs* in a general linear model for all datasets (see Zeithamova et al., 2017). In other words, we modeled our stimuli at the exemplar level across all runs rather than the trial level.

Preprocessing and Modeling of Neuroimaging Data

Preprocessing of the structural and functional data from all three datasets was performed using the same preprocessing pipeline in BrainVoyager 20.4 (Brain Innovation, Inc., Maastricht, the Netherlands). Functional images were corrected for differences in slice time acquisition (cubic spline interpolation), head motion (trilinear-sinc interpolation), and high-pass filtered (Fourier basis, 2 cycles) to remove linear and non-linear trends. Functional data were then co-registered to the structural data using boundary-based registration, normalized into Talairach space, and re-sampled to 3-mm isotropic voxels. Functional data were not spatially smoothed. For each participant, all included runs were modeled with a two-gamma hemodynamic response function and analyzed simultaneously using a single univariate General Linear Model (GLM), corrected for serial correlations with a second-order autoregressive method. The GLM includes one predictor for each condition (8 categories \times 9 exemplars, 4 categories \times 9 exemplars, and 6 categories \times 9 exemplars in Datasets 1 – 3 respectively; **Table A-1**), one predictor for the target (red hash) condition (with or without button presses, as well as non-target (e.g., digit) trials that were responded to similarly as to a target trial), six predictors of motion parameters (translational and rotational in x , y , and z axes) for each run, and one constant predictor for each run. Although there are different number of predictors across datasets, we focused only on the beta estimates and corresponding t statistics derived from 36 predictors (9 digits, 9 letters, 9 scrambled digits, and 9 scrambled letters) for the multivariate pattern analyses.

Regions of Interest

Regions of interest (ROIs) were obtained from the meta-analysis in Chapter 2 (Yeo et al., 2017) in which preferential activity to Arabic numerals relative to other familiar symbols (e.g., Roman letters for English speakers or Chinese characters for Chinese speakers) was found to be convergent across 20 studies. Numeral preference was localized in the right ITG (55 3-mm isotropic voxels), as well as bilateral parietal and right frontal regions (see **Figure 3-2(a)** and Appendix A for more details of the ROIs). For our a priori hypotheses, we focused on the cluster in the right ITG, as well as the left homologue region because the left ITG also exhibits numeral preference (Amalric & Dehaene, 2016; Bugden et al., 2019; Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016; Pollack & Price, 2019; Roux et al., 2008), but is possibly less robust to varying task contexts or is spatially more variable. Moreover, Pollack and Price (2019) found that although the left (but not right) ITG showed, on average across participants, numeral preference when detecting digits among letters, individual differences in the activation of the right (but not left) ITG during digit detection correlated with calculation competence. To assess the specificity of the ITG findings independent of correlated signal and/or noise across regions, as well as for completeness, we also analyzed the representational geometries in the parietal and frontal regions and reported these exploratory findings in Appendix A. Individual differences and group means of the size and temporal signal-to-noise ratio in each ROI for each dataset are reported in **Figure A-2**.

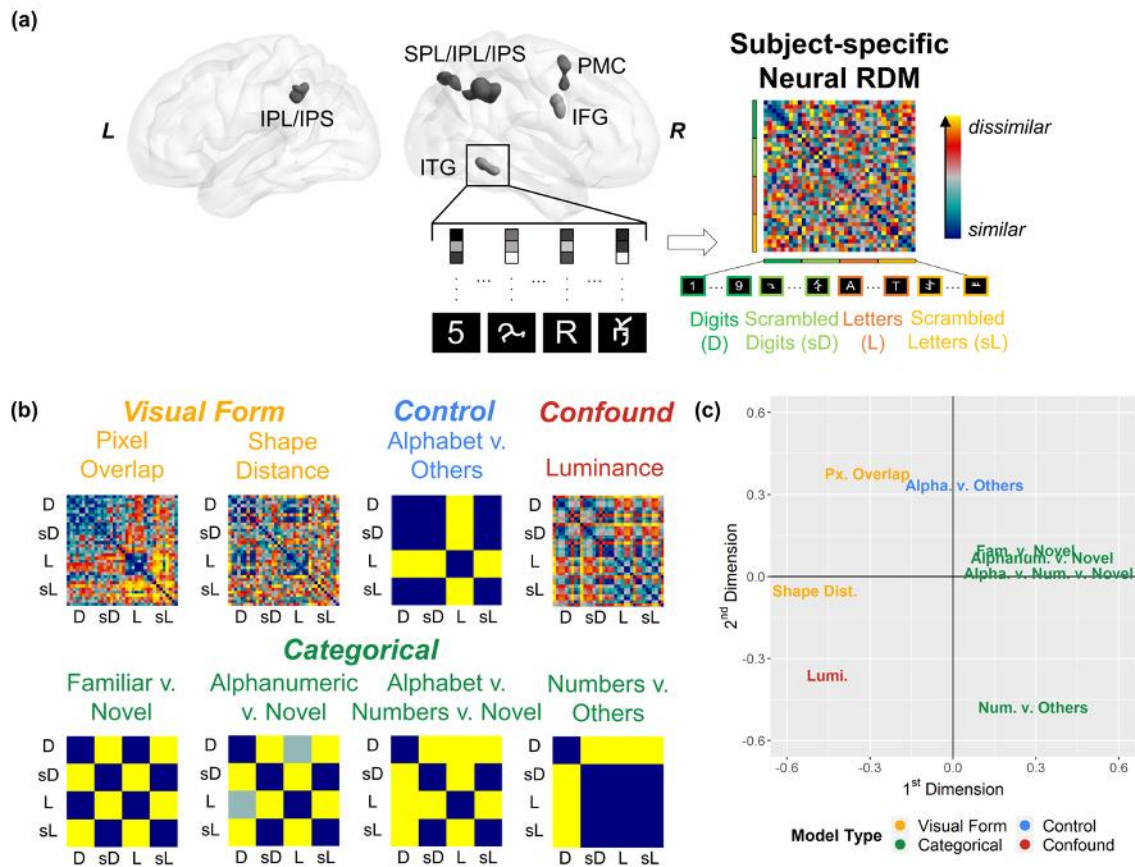


Figure 3-2. Regions of interest (ROIs), and neural and candidate representational dissimilarity models (RDMs)

(a) Numeral-preferring ROIs from the meta-analysis in Chapter 2 (Yeo et al., 2017), and an example neural RDM (using correlational distance) constructed from the activation patterns evoked by 9 Digits (D), 9 Scrambled Digits (sD), 9 Letters (L), and 9 Scrambled Letters (sL) within the right ITG. IPL: inferior parietal lobule. IPS: intraparietal sulcus. SPL: superior parietal lobule. PMC: premotor cortex. IFG: inferior frontal gyrus. ITG: inferior temporal gyrus. (b) Candidate RDMs (using Letters Set 1 from Datasets 1 and 2; see **Figure A-5** for Letters Set 2 from Dataset 3). All models presented were rescaled to [0, 1] for comparative visualization. (c) Multidimensional scaling plot of the correlational distance among the candidate models using Letters Set 1 (see **Figure A-5** for Letters Set 2 from Dataset 3).

Representational Similarity Analyses

Figure 3-2(a) shows a schematic overview of the approach taken for the RSA.

Neural Representational Dissimilarity Matrices (RDMs)

Within each ROI, activation patterns evoked by each of the 36 exemplars were characterized by the spatial distribution of t -values (Misaki et al., 2010) from exemplar vs. baseline contrasts, since t -values take into account the noise in the voxels (Misaki et al., 2010) and thus mitigate any differences in temporal signal-to-noise ratios across datasets. Subsequent analyses were performed in MATLAB using the Representational Similarity Toolbox (Nili et al., 2014) and in-house scripts. For each ROI within each participant, we first excluded voxels that had no functional coverage or signal across all exemplars using intensity-based thresholding (100 arbitrary units as a default threshold in BrainVoyager, and 1800 arbitrary units as a modified threshold for 15 participants in Dataset 3 whose raw intensities were about 15 – 20 times as high as a typical functional dataset). The activation patterns were then scaled by subtracting the mean activation pattern (across exemplars) from the exemplar-specific activation pattern (Diedrichsen & Kriegeskorte, 2017; Misaki et al., 2010; Op de Beeck, 2010; Walther et al., 2016). Finally, for each ROI, participant-specific dissimilarities between all 36 exemplar-evoked activation patterns, computed using correlational distance $1 - \text{Pearson's } r$, were summarized in a 36×36 matrix (**Figure 3-2(a)**).

Candidate Representational Models

We constructed eight candidate model RDMs, two that characterize representational similarity based on visual form of the characters, four that characterize conceptual categories, a control model that characterizes letter sensitivity, and a confound model that characterizes luminance differences between digits and letters (**Figure 3-2(b)**).

Visual Form Models. We focused on two different measures to quantify lower-level and higher-level visual form similarity between each pair of characters. The Pixel Overlap model is based on the commonly used pixel-wise Euclidean distance between each pair of grayscale images. It is defined by $d_{jk} = \frac{1}{\sqrt{N}} \sqrt{\sum_{x=1}^N [I_j(x) - I_k(x)]^2}$, where N is the number of pixels in the image, and $I_j(x)$ and $I_k(x)$ are the pixel intensities at location x in images I_j and I_k (e.g., Chouinard, Morrissey, Köhler, & Goodale, 2008; Grill-Spector et al., 1999; Op de Beeck, Torfs, & Wagemans, 2008). The larger this index is, the greater the physical (retina) difference between each pair of characters. This model thus assesses whether the representational geometry in a region retains lower-level retinotopic overlap in the shape envelope of the characters. It is not invariant to font, size, and position.

In contrast, although “5” and “S” may not have high pixel-to-pixel overlap, human observers may consider their abstract shapes to be highly similar. The Shape Distance model overcomes the limitation of the Pixel Overlap model by considering higher-level shape similarity based on a computational algorithm that relies on the “context” of a sampled point on a shape (i.e., how one point on a shape relates to all other points on the shape) (Belongie et al., 2002) (see Appendix A for computational details of this measure). Compared to the pixel-based measure above, the shape distance measure is invariant to translation and scaling (but not rotation, otherwise “6” and “9” will be highly confusable), and has been shown to outperform the pixel-based measure in recognition of several categories of objects, including handwritten digits (hence font-invariant too) (Belongie et al., 2002). Several studies have employed this measure in investigations of the role of abstract shape similarity in neural representations of object recognition (Bracci et al., 2015; Fairhall et al., 2011; Gotts et al., 2011; Mahon et al., 2007). Multidimensional scaling plots illustrating the dissimilarities of the 36 characters based on pixel

overlap and shape distance are shown in **Figure A-3** and **Figure A-4**.

Categorical Models. Four categorical models were constructed. Unless otherwise noted, description of high similarity between each pair of characters was coded as having a correlational distance ($1 - \text{Pearson's } r$) of 0, and high dissimilarity was coded as having a correlational distance of 1.

Figure 3-3 illustrates these four categorical models. The Familiar v. Novel model and Alphanumeric v. Novel model are based on the hypothesis that a region responds to all familiar characters (digits and letters) in a manner that is different from how it responds to novel characters (scrambled digits and letters). In the Familiar v. Novel model, digits and letters are indistinguishable. In the Alphanumeric v. Novel model, digits and letters are somewhat distinguishable, but are still more similar to one another than to novel characters (here we coded 0 for high similarity, 2 for high dissimilarity, and 1 for in-between). The Alphabet v. Numbers v. Novel model is based on the hypothesis that digits, letters, and novel characters are equally distinguishable, and that one category is no more similar to any one of the other two categories. Although the Alphanumeric v. Novel and Alphabet v. Numbers v. Novel models suggest that digits are represented as a distinct category from letters and novel characters (i.e., a region is sensitive to the three character categories, but shows no greater sensitivity for any one category over the others), they do not indicate that a region is specialized for processing numerals. Numbers v. Others model is the strongest test for numeral sensitivity in pITG-numerals. It is based on the hypothesis that a region responds to digits in a manner that is different from how it responds to letters and novel characters, and importantly, it does not distinguish letters from novel characters.

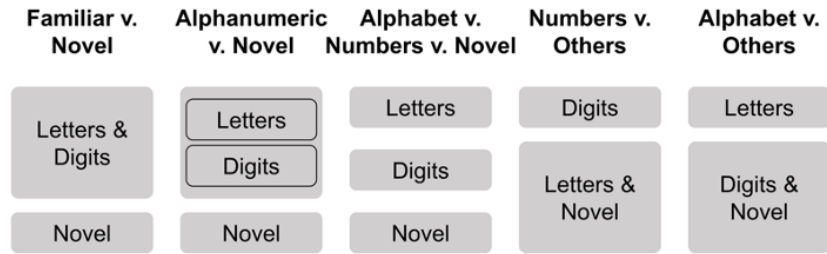


Figure 3-3. Schematic of the categorical and control models

Control Model. To rule out the possibility that the representational geometry in the pITG-numerals is simply categorical in nature, and non-specific, we also tested an Alphabet v. Others control model (i.e., shows greater letter sensitivity) (see **Figure 3-3**). Given the spatial dissociation found in previous work, this control model is highly unlikely to describe the representational geometry in pITG-numerals, and thus provides a strong test for the specificity of the other more plausible categorical models above.

Confound Model. As the Letters Set has greater luminance than the Digits Set, the Luminance model was included to directly assess whether pairwise differences in luminance suffice to describe the representational geometry of the characters in a region. Pairwise dissimilarity in luminance was computed by taking the absolute difference in luminance between two images and rescaled to [0, 1].

It is critical to note that all eight models are neither mutually exclusive nor fully orthogonalized. In particular, the Familiar v. Novel, Alphanumeric v. Novel, and Alphabet v. Numbers v. Novel models show very subtle differences and are highly correlated with one another. These three highly correlated categorical models were included primarily to explore whether one model may be better than another in describing the representational geometry in a

region. Other than this family of highly correlated models, **Figure 3-2(c)** and **Figure A-5(b)** show that the models are sufficiently different from one another, with each group of models roughly occupying separate quadrants in a two-dimensional representational space (see **Figure A-5(c)** for the pairwise rank correlations between the models).

Similarity Between Neural RDMs and Model RDMs

To quantify the extent to which the representational geometry in an ROI is similar to that described by a candidate model, we compared the neural RDM with each model RDM (one-half of each symmetric matrix) using Kendall's tau-a (τ_a) rank correlation (Nili et al., 2014). This was performed for each participant, and the participant-specific correlational coefficients were subjected to a one-sided Wilcoxon signed-rank test across participants to assess whether the mean neural-model similarity was significantly greater than 0. The use of ranked measures at both the participant and group levels ensures that our findings are robust to any outlying data points, but it necessarily comes with a loss of sensitivity to distinguish between models within participants because it does not exploit the continuous nature of the values in the neural RDMs (Diedrichsen & Kriegeskorte, 2017). For this and all other hypothesis tests, we used $\alpha < .05$ as our threshold for false positives. Multiple comparisons across models within each ROI were accounted for by controlling for false-discovery rate (FDR) at $q < .05$ (Benjamini & Hochberg, 1995). Given that some models were of no theoretical interest (e.g., luminance model) and that some candidate models tested are highly similar and their inclusion was primarily exploratory, FDR-correction might be too conservative. Hence, although we reported statistics that were corrected for FDR, inferences were made jointly from the uncorrected frequentist and Bayesian statistics (see below for details of Bayesian tests).

To quantify the degree of between-participant variability in each dataset, we estimated

the mean correlation between the participant-specific neural RDM and an unknown true model RDM. This is indicated by the noise ceilings in **Figure 3-4**. The ceiling upper bound was computed by correlating participant-specific neural RDMs with a “central” neural RDM (that maximizes its correlation to the participant-specific neural RDMs), and averaged across participants (see Nili et al., 2014, for details). Hence, assuming that the experimental paradigm was meant to yield robust effects across participants with low measurement error, this upper bound is the highest correlation that any model RDM can achieve in a given dataset. The lower bound was computed by a “leave-one-participant-out” approach, such that each participant’s neural RDM was correlated with that of the remaining participants, and averaged across participants (Nili et al., 2014). The noise ceilings not only provide information of between-participant variability across datasets to account for potential differences in our findings, but also allows us to examine whether the task was sensitive in detecting the effects of interest at the group level.

For cases in which there is evidence that at least one categorical model described the representational geometry in a region, we probed the “unique” similarity of each categorical model using a semipartial correlation approach (i.e., controlling for visual form and luminance confound models only from the neural RDM). For the semipartial correlations computed using the *ppcor* R package (S. Kim, 2015), Kendall’s τ_b was used instead of τ_a because there is no statistical software to the best of our knowledge that implements the τ_a variant for semipartial correlations. Although τ_a has been found to favor simplified models (e.g., categorical) over detailed models less often than τ_b (Nili et al., 2014), we focused on comparing only among categorical models, so the bias is less critical here. Pairwise differences were also performed, and multiple comparisons across models within each ROI were FDR-corrected.

Visualization of Representational Geometry Within ROI

To visualize the mean representational geometry within each ROI in two- and three-dimensional spaces, we applied multi-dimensional scaling (MDS) to the group-averaged neural RDM using *cmdscale* function in R. All plots are made available at <https://osf.io/jwgk8/>.

Complementary Frequentist and Bayesian Analyses

Statistical inferences were made jointly based on both frequentist Type I error control of $\alpha < .05$ (uncorrected for multiple comparisons) as well as Bayes factors as a more continuous measure of evidence in support of one hypothesis over another. Non-parametric frequentist analyses and Bayesian analyses were conducted in MATLAB (Nili et al., 2014), R (R Core Team, 2018) and JASP 0.10.0 (JASP Team, 2019). As the availability of Bayesian equivalent of non-parametric tests is currently limited, and to accommodate the assumptions of traditional parametric tests that also apply to Bayesian analyses, we first transformed Kendall's τ to Pearson's r using the formula $r = \sin(.5\pi\tau)$ (Gilpin, 1993; Walker, 2003) (e.g., see Martin, Douglas, Newsome, Man, & Barense, 2018), and then performed Fisher's z -transformation on Pearson's r . These z -transformed r values (r_z) were then used to estimate the Bayes factors (BF). In summary, we performed frequentist tests on raw Kendall's τ values, and performed complementary Bayesian analyses on z -transformed r values.

For all Bayesian analyses, we used the default “objective” priors (correlation: stretched beta prior width = 1; one-sample and paired-samples t -test: “medium” Cauchy prior width of 0.707) because of a lack of literature to specify “informed” priors. Nonetheless, as Bayes factors are dependent on the priors used, we also conducted sensitivity analyses of the robustness of the BF s to different specifications of prior (“wide” and “ultrawide” Cauchy priors 1 and 1.414 respectively), and any evidence that a specific finding may not be robust to the choice of the

priors was noted as a caveat. In general, BFs tend to decrease with wider Cauchy priors, hence, all reported BFs using the default prior (.707) were close to the highest attainable.

For all one-sample t -tests, we report BF_{+0} that expresses the likelihood of the data given $H+$ (one-sided, $r_z > 0$) relative to $H0$ ($r_z = 0$) assuming that $H+$ and $H0$ are equally likely, to complement one-sided p -values. For post-hoc paired-samples tests, we report BF_{10} that expresses the likelihood of the data given $H1$ (two-sided, r_z difference $\neq 0$) relative to $H0$ ($r_z = 0$), to complement two-sided p -values. Although we note that BFs provide continuous measure of evidence, we used BF_{+0} or $BF_{10} > 3$ in support of the alternative hypothesis, and BF_{+0} or $BF_{10} < 1/3$ in support of the null hypothesis as thresholds for deciding whether the evidence for either hypothesis was conclusive (Dienes, 2014; Dienes & Mclatchie, 2017).

Small-scale Meta-analyses of Effect Sizes

To provide a summary effect size of the three datasets for each model in each ROI, we performed a classical fixed-effects meta-analysis. It is valid and recommended to conduct a small-scale meta-analysis on a minimum of two studies to provide a quantitative summary of studies with similar methodology (Goh et al., 2016; Lakens & Etz, 2017; Valentine et al., 2010). These meta-analyses were conducted using JASP 0.10.0 (JASP Team, 2019) on the mean Fisher's z -transformed Kendall's τ_a values (r_z) from each dataset as effect sizes of the similarity between the neural RDM and a model RDM, weighted by their inverse squared standard errors. In other words, each meta-analytic effect size is a weighted mean of the three datasets. A fixed-effects approach assumes a common true effect size across studies, and that its variance is solely due to sampling variation. Here, we do not aim to generalize the findings from our specific task and stimulus sets to other studies, so a fixed-effects approach is appropriate. Tests of heterogeneity in the residuals in 45 out of 48 meta-analyses indicated no significant

heterogeneity in effect sizes across the datasets (all $ps > .053$), and that a fixed-effects model was justified in most cases. Multiple comparisons across models within each ROI were accounted for by controlling for FDR at $q < .05$. Finally, we also performed complementary Bayesian fixed-effects meta-analyses with Cauchy prior width of 0.707 using the *BayesFactor* package (Morey & Rouder, 2018) as described in Rouder and Morey (2011). Specifically, a summative Bayes factor is computed using the t -statistics of each dataset (derived from a one-sided one-sample t -test on the Fisher's z -transformed Kendall's τ_a values) and weighted by their sample sizes. The Fisher's z values were then transformed back to Pearson's r for presentation (Goh et al., 2016).

Data and Code Availability

Raw behavioral and MRI data from Datasets 1 and 2 are available upon direct request from the corresponding author. Dataset 3 is publicly available at OpenNeuro (<https://openneuro.org/datasets/ds002033>; doi: 10.18112/openneuro.ds002033.v1.0.0) (Merkley et al., 2019). The stimuli, model RDMs, neural RDMs from all datasets, RSA data and code necessary to reproduce all results reported are publicly available at Open Science Framework (<https://osf.io/jwgk8/>; doi: 10.17605/OSF.IO/JWGK8).

Results

Representational Geometry in Candidate Numeral-Preferring Regions in ITG

Given the large number of tests conducted across all datasets, we summarize the data-specific and meta-analytic findings visually in **Figure 3-4**, and provide the detailed statistics only for the meta-analyses in **Table 3-1**. We also report the statistics and describe the findings only for dataset-specific positive evidence from frequentist and/or Bayesian tests, but invite readers to refer to the complete results output in the format of JASP files at <https://osf.io/jwgk8/> for all other statistical details.

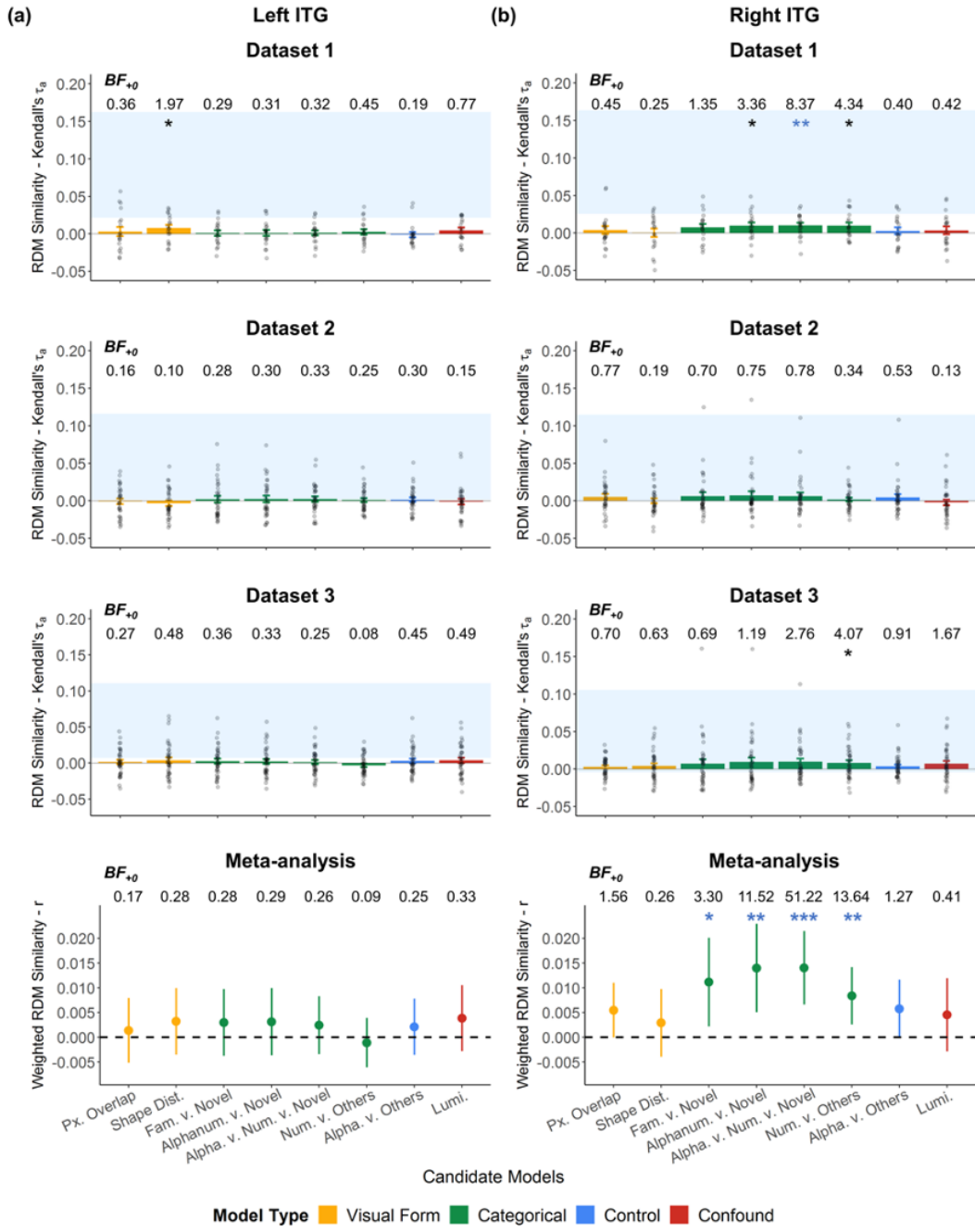


Figure 3-4. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-preferring regions in (a) left and (b) right inferior temporal gyrus (ITG). Blue bars indicate the estimated noise ceiling. Group means and standard errors of the similarity are indicated by the bar plots with error bars. Individual data points are shown as grey dots. Evidence of similarity is indicated by black asterisk: * $p < .05$, ** $p < .01$, *** $p < .001$, one-sided, uncorrected. Blue asterisks indicated results that remained significant with FDR correction. BF_{+0} = Bayes Factor [$r_z > 0$ vs. $r_z = 0$]. Lines in meta-analytic plots indicate the 95% confidence interval around the overall weighted r .

Table 3-1. Meta-analyses for the degree of similarity between each model RDM and neural RDMs in left and right ITG

Model	Left ITG				Right ITG			
	<i>r</i>	95% CI	<i>p</i>	<i>BF</i> ₊₀	<i>r</i>	95% CI	<i>p</i>	<i>BF</i> ₊₀
Pixel Overlap	.0014	[-.0052, .0079]	.680	0.17	.0055	[-.00005, .0110]	.052	1.56
Shape Distance	.0032	[-.0035, .0099]	.347	0.28	.0029	[-.0039, .0097]	.406	0.26
Familiar v. Novel	.0030	[-.0038, .0097]	.386	0.28	.0112	[.0022, .0201]	.015	3.30
Alphanumeric v. Novel	.0031	[-.0036, .0099]	.365	0.29	.0140	[.0050, .0229]	.002	11.52
Alphabet v. Numbers v. Novel	.0024	[-.0034, .0083]	.415	0.26	.0140	[.0066, .0214]	.0002	51.22
Numbers v. Others	-.0011	[-.0061, .0039]	.667	0.09	.0084	[.0026, .0141]	.005	13.64
Alphabet v. Others	.0021	[-.0035, .0078]	.463	0.25	.0058	[-.0001, .0116]	.055	1.27
Luminance	.0038	[-.0028, .0105]	.260	0.33	.0045	[-.0029, .0119]	.230	0.41

Note. 95% CI: 95% confidence interval. *BF*₊₀ = Bayes Factor (*r* > 0 vs. *r* = 0)

Left ITG

Overall, there was conclusive meta-analytic evidence of a lack of similarity between the neural RDMs and any model RDM (**Figure 3-4(a)** and **Table 3-1**). Below we report whether the meta-analytic findings were also observed in each dataset.

Visual Form Models. There was no evidence of similarity between the neural RDMs and Pixel Overlap or Shape Distance model RDM in any of the datasets, except for some weak evidence for the Shape Distance model for Dataset 1 ($\tau_a = .0078$, $p = .044$, $BF_{+0} = 1.97$).

Categorical, Control, and Confound Models. Across the three datasets, there was no evidence of similarity between the neural RDMs and any of the categorical, Alphabet v. Others model, and Luminance model RDMs.

Right ITG

Overall, there was conclusive meta-analytic evidence of no similarity between the neural RDMs and Shape Distance model RDM, and conclusive meta-analytic evidence of similarity between the neural RDMs and models that distinguish numerals from letters (Alphanumeric v. Novel, Alphabet v. Numbers v. Novel, and Numbers v. Others) (**Figure 3-4(b)** and **Table 3-1**). The Bayes factor for the Familiar v. Novel model was not robust to varied priors as it decreased below 3 with a wide prior. The Alphanumeric v. Novel, Alphabet v. Numbers v. Novel, and Numbers v. Others models were, on average, at least 3 times more likely than the Familiar v. Novel model. Importantly, the Bayes factor for the Numbers v. Others model was 7.83 even with an ultrawide prior, suggesting that the similarity between the neural and the Numbers v. Others model RDMs was 7 to 13 times more likely under the hypothesis of a positive correlation than that of a null correlation. Moreover, for the Numbers v. Others model, a “Fail-safe N ” analysis estimated that 8 studies with an effect size of 0 would have to be added to the meta-analysis to reduce the meta-analytic effect size to one with a false positive rate $\geq .05$. The Numbers v. Others model still remained statistically significant when we controlled for false positives for the Numbers v. Others model across the three datasets (FDR-corrected $ps < .05$ in Datasets 1 and 3).

Although evidence for the Alphabet v. Others control model RDM was inconclusive, it was much less likely to describe the neural RDMs than the Numbers v. Others model RDM. A comparison of the BFs for the Alphabet v. Others ($BF_{+0} = 1.27$) and the Numbers v. Others models ($BF_{+0} = 13.64$) indicates that the Numbers v. Others model was 10 times more likely to describe the neural RDMs than the Alphabet v. Others model, suggesting that novel characters are more similar to letters than to digits.

Taken together, there was conclusive evidence that the candidate numeral-preferring ITG region processed digits and letters differently, and the fact that the Numbers v. Others model

could describe its representational geometry suffices as evidence supporting some degree of greater numeral sensitivity relative to the other categories. Below we report whether the meta-analytic findings were also observed in each dataset. The dataset-specific results below are summarized in **Table A-5**.

Visual Form Models. There was no evidence of similarity between the neural RDMs and Pixel Overlap or Shape Distance model RDM in any of the datasets.

Categorical Models. For Dataset 1, there was evidence of similarity between the neural RDMs and the three categorical model RDMs that distinguish numbers as a distinct category from letters and novel characters (Alphanumeric v. Novel: $\tau_a = .0096$, $p = .016$, $BF_{+0} = 3.36$; Alphabet v. Numbers v. Novel: $\tau_a = .0100$, $p = .005$, $BF_{+0} = 8.37$; and Numbers v. Others: $\tau_a = .0098$, $p = .033$, $BF_{+0} = 4.34$). Although the evidence for the Alphabet v. Numbers v. Novel model was the strongest amongst the three models, post-hoc pairwise comparisons revealed no evidence of within-participant differences between these three categorical model RDMs in their similarity to the neural RDMs (all $ps > .828$, $BF_{10s} < 0.25$). There was also still evidence of similarity between the neural RDMs and these categorical model RDMs even after controlling for the visual form and confound model RDMs (Alphanumeric v. Novel: $\tau_a = .0121$, $p = .020$, $BF_{+0} = 3.28$; Alphabet v. Numbers v. Novel: $\tau_a = .0143$, $p = .006$, $BF_{+0} = 7.90$; and Numbers v. Others: $\tau_a = .0145$, $p = .033$, $BF_{+0} = 4.49$). Similarly, post-hoc pairwise comparisons revealed no evidence of within-participant differences between these three categorical model RDMs in their unique similarity to the neural RDMs (all $ps > .651$, $BF_{10s} < 0.31$). Finally, it is important to note that the Bayes factors for the Alphanumeric v. Novel model in both zero-order and semipartial correlations were not robust to varied priors as they decreased below 3 with a wide Cauchy prior (≥ 1), whereas the Bayes factors for the Numbers v. Others model remained

relatively robust, and decrease to 2.95 (zero-order correlation) and 3.07 (semipartial correlation) only with an ultrawide prior ($\geq \sqrt{2}$).

For Dataset 2, we found no evidence of similarity between the neural RDMs and any categorical model RDMs. Importantly, there was no evidence of null correlations either ($BF_{+0} > 1/3$), suggesting that these results were not inconsistent with those of Dataset 1.

For Dataset 3, there was evidence of similarity between the neural RDMs and only the Numbers v. Others model ($\tau_a = .0083, p = .026, BF_{+0} = 4.07$). There was still evidence of similarity between the neural RDMs and Numbers v. Others RDM after controlling for the visual form and confound model RDMs ($\tau_b = .0126, p = .021, BF_{+0} = 4.97$). The Bayes factors for the Numbers v. Others model in both zero-order and semipartial correlations were somewhat robust to varied priors and decreased to 2.47 and 3.05 respectively only with an ultrawide prior.

Figure 3-5 illustrates the group-averaged dissimilarity matrix and representational geometry of the 36 characters in this region for each dataset. To further assess whether the three-way distinction (numerals, letters, and novel characters) observed using a model-driven approach could also be observed using a data-driven approach, we performed a k -medoids clustering analysis (Kaufman & Rousseeuw, 1990) for each dataset. Overall, evidence for a three-cluster structure was not strong in all datasets, but consistent with our findings above, there exists a cluster that showed a slight dominance of numeral representations in both Datasets 1 and 3 (see **Table A-6 – Table A-8** and **Figure A-6 – Figure A-8**).

Control Model. There was no evidence of similarity between the neural RDMs and Alphabet v. Others model RDM in any of the datasets.

Confound Model. There was no evidence of similarity between the neural RDMs and Luminance model RDM in any of the datasets.

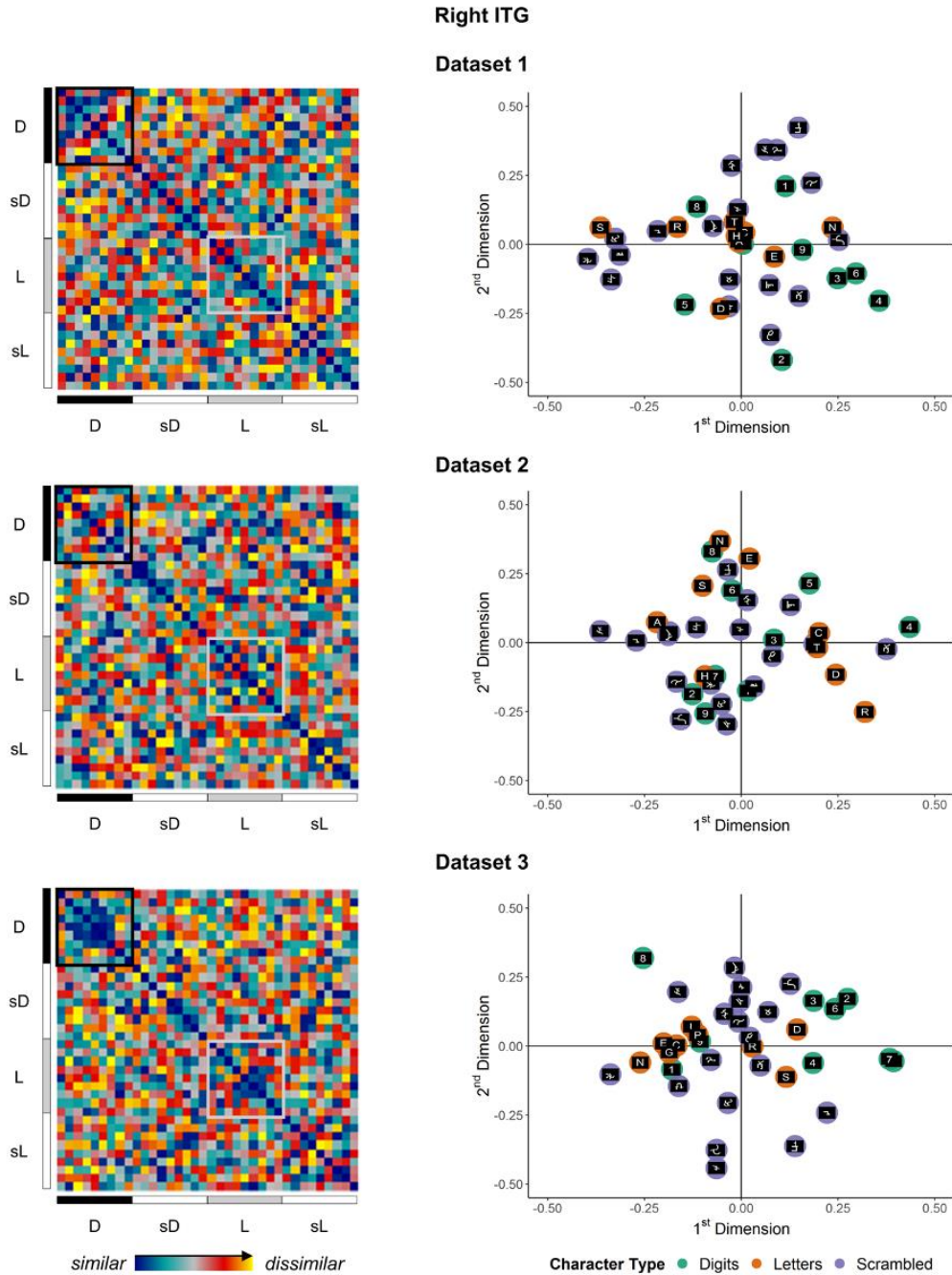


Figure 3-5. Group-averaged representational dissimilarity matrices and representational geometry of exemplars in right ITG
 Group-averaged representational dissimilarity matrices and representational geometry of the 36 exemplars (D: Digits, sD: Scrambled Digits, L: Letters, sL: Scrambled Letters) in two-dimensional space in the numeral-preferring right inferior temporal gyrus (ITG) in each dataset. Three-dimensional interactive plots are available at <https://osf.io/jwqk8/wiki/home>.

Representational Geometry in Candidate Numeral-Preferring Parietal and Frontal Regions

To assess how specific the above findings are to the ITG regions, as well as for completeness in exploring other candidate numeral-preferring regions, we performed identical analyses for the bilateral parietal and right frontal regions from the meta-analysis by Yeo and colleagues (2017) (**Figure 3-2(a)**). We found meta-analytic evidence of similarity between the Numbers v. Others model RDM and the neural RDM in the right parietal region, as well as similarity between all other category-sensitive categorical model RDMs and the neural RDMs in both left and right parietal regions (**Figure A-9 – Figure A-10** and **Table A-9**). There was also meta-analytic evidence of similarity between the category-sensitive categorical model RDMs and the neural RDMs in the right inferior frontal region (**Figure A-11 – Figure A-12** and **Table A-10**).

Discussion

The ventral occipitotemporal cortex (vOT) is known to include distinct neuronal populations that are tuned to different perceptual categories such as faces, body parts, scenes, and written words (for a review, see Kanwisher, 2010). Although it has long been known that regions in the left vOT show preference for single letters and letter strings relative to other character types including digits (L. Cohen & Dehaene, 2004; Flowers et al., 2004; James et al., 2005; Park et al., 2012; Polk et al., 2002; Polk & Farah, 1998; Vinckier et al., 2007), letters are no longer that special. There is now a growing body of evidence that the vOT also has a region that shows preference for Arabic numerals, known as the “Number Form Area” (NFA) in the posterior inferior temporal gyrus (pITG) (Amalric & Dehaene, 2016; Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016). In this study, we probed the organization of the neural

responses to task-irrelevant individual digits, letters and novel characters to better understand the functional boundaries of a candidate numeral-preferring region in the pITG (“pITG-numerals”).

Evidence of Numeral Sensitivity in Right pITG-numerals During Passive Viewing

Using multivariate representational similarity analyses (RSA), our results suggest that the pITG-numerals in the right hemisphere does distinguish between digits and letters in its distributed response patterns even when the characters are task-irrelevant. This is contrary to the univariate findings previously reported for the same datasets. Moreover, the right pITG-numerals was more likely to represent digits in its own category, and letters and novel characters indistinguishably in another category (Numbers v. Others model) than to represent letters in its own category, and digits and novel characters indistinguishably in another category (Alphabet v. Others model). Complementary to our model-driven approach, data-driven clustering analyses also support the presence of a digit-dominated cluster in Datasets 1 and 3, albeit weakly. These findings suggest the possibility of a greater numeral sensitivity in the right pITG-numerals. This is not surprising given that the region examined here is a region defined a priori from a meta-analysis of numeral-preferring regions in Chapter 2 (Yeo et al., 2017). However, it resolves the crucial concern that passive-viewing paradigms may be ill-suited for the investigation of the putative NFA, and clarifies the need for different analytical approaches that go beyond mean activation levels and that are more sensitive to effects evoked by mere passive viewing. While we did find support for functional specialization using a multivariate approach, active tasks may still be better for investigating the function of this region given a recent finding that mathematical tasks with visually dissimilar stimuli (e.g., Arabic numerals, dice patterns, and finger representations) engage the pITG more consistently than the mere presence of Arabic numerals (e.g., Grotheer et al., 2018; see also Pollack & Price, 2019). Here we demonstrate that

digits and letters may evoke distinct distributed response patterns even though their overall response strengths may be similar. Hence, multivariate approaches may have greater sensitivity than univariate approaches when analyzing the processing of task-irrelevant characters.

Although we found conclusive evidence of categorical distinctions in the right pITG-numerals in Datasets 1 and 3, evidence was inconclusive in Dataset 2. One possibility for the inconclusive findings in Dataset 2 is the fewer trials per exemplar that were available for the estimation of the activation patterns in Dataset 2 (two trials per exemplar compared to four trials in Dataset 1 and eight trials in Dataset 3). This factor should not be specific to any ROI. Yet, we found conclusive findings in the parietal ROIs in Dataset 2 and inconclusive findings in the parietal ROIs in Dataset 3. Hence, number of trials seems unlikely to fully account for the difference in results. Although the mean temporal signal-to-noise ratio (tSNR) in the right pITG-numerals was much higher in Dataset 3 than in Datasets 1 and 2, there was inconclusive evidence that the mean tSNRs differed between Datasets 1 and 2. This suggests that differences in tSNR also do not fully account for the differential results. Other possible factors may include a younger sample in Dataset 2 that has fewer years of experience with processing numerals and math instruction, but the small number of studies included here preclude any analysis of moderators in the meta-analyses. In any case, the inconclusive findings for Dataset 2 do not provide support for the null hypotheses either, and thus do not undermine the positive findings observed in the other two datasets.

These results are also likely to be specific to the right pITG-numerals. We did not observe identical findings in most of the other candidate numeral-preferring parietal and frontal region or the left homologue of the pITG-numerals within each dataset, and thus the greater numeral sensitivity observed in the right pITG-numerals cannot be purely driven by noise in the

activation patterns or by intrinsic connectivity across these regions. The absence of a greater numeral sensitivity in the left homologue is also consistent with previous findings that the left pITG is involved in numeral processing, but does not show a preference for numerals when they are irrelevant for the task (Pollack & Price, 2019). Although a numeral-preferring region in the left pITG has been observed in several studies (Amalric & Dehaene, 2016; Bugden et al., 2019; Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016), its specific role is still unclear, and may be engaged under different circumstances from its right counterpart or have different functional and structural connections to other brain regions. For instance, it has long been proposed that the bilateral NFA have connections to magnitude processing regions in the parietal cortex, but only the left NFA has connections to frontal language regions for number word transcoding (L. Cohen & Dehaene, 1995, 1996, 2000; Dehaene & Cohen, 1995). Findings from several lines of research support the hemispheric asymmetry account. Several event-related potential studies found right-lateralization of digit-specific processing, in contrast to left-lateralization of letter-specific processing, in children in first grade through adolescence as well as in adults (Lochy & Schiltz, 2019; Park et al., 2014, 2018). In a study comparing mathematicians and non-mathematicians (Amalric & Dehaene, 2016), a right numeral-preferring pITG region in non-mathematicians responded more to numerals than to words and mathematical formulas, but the left numeral-preferring pITG region showed an attenuated preference for numerals. In mathematicians, however, both left and right numeral-preferring pITG regions responded to formulas and numerals to a similar degree, but only the left numeral-preferring pITG region was modulated by mathematical expertise. Recently, it was also found that individual differences in digit-specific activation in the right, but not left, pITG correlated positively with calculation competence (Pollack & Price, 2019). Alternatively, there may be a left numeral-preferring pITG region, but it

does not overlap with our candidate region-of-interest, possibly due to greater inter-individual variability in its localization as a function of other symbol-preferring regions, such as letter- and word-preferring regions (Glezer & Riesenhuber, 2013). Taken together, it is likely that the functional specialization for numeral processing in the right pITG may be more robust than its left counterpart for reasons yet to be known.

To the extent that digits and letters are highly similar in their curvilinear features, it is conceivable that the ability of neuronal populations to categorize “S” as a letter and “8” as a digit is due mainly to the conceptual knowledge that “8” has a quantitative referent, but “S” does not (i.e., task-driven conceptual processing). Given that the characters are irrelevant for the task, there was no need for participants to distinguish digits from letters, or their individual identities. Hence, observing some degree of greater numeral sensitivity in the pITG-numerals in the current datasets suggests that there are automatic, stimulus-driven processing biases. Considering the broader question of how different perceptual categories seem to occupy different regions in the vOT, Gauthier (2000) proposed a “process-map” model, in which automatic processing biases arise from our (and the brain’s) experience in associating different recognition and computational goals with different categories of objects (for a recent review, see Op de Beeck, Pillet, & Ritchie, 2019). It is therefore likely that literacy and numeracy lead to the association of letters and numerals with habitually different goals (numerical and mathematical relevance or not), which in turn lead to divergent neural processing pathways in the vOT that have a preparatory or biased response for stimuli that potentially have numerical relevance or not. From this perspective, not only does pITG-numerals encode subtle differences between visually similar objects, such as numerals and letters, it could even encode similarities between objects that are visually dissimilar (Gauthier, 2000), such as numerals, dice patterns, finger representations, mathematical formulas

with Greek and Roman letters (Amalric & Dehaene, 2016; Grotheer et al., 2018), or even from a different sensory input, such as soundscapes associated with numerical content (Abboud et al., 2015), and auditory mathematical statements (Amalric & Dehaene, 2016, 2018, 2019). In other words, it is likely that pITG-numerals is recruited whenever the brain “predicts” that the stimulus has numerical relevance, through automatic feedforward connections from posterior ventral (occipital cortex) regions and/or feedback connections from parietal and frontal regions. This is also consistent with the interactive account that the analogous “Visual Word Form Area” is involved in predictive coding through an experience-driven automatic interaction of forward and backward connections, rather than word form detection per se (Price & Devlin, 2011). In the current study, we are unable to disentangle the automatic, stimulus-driven feedforward and feedback contributions, but only seek to exclude any modulatory contribution of effortful, task-driven processing that may bias or amplify the representations along a particular dimension (e.g., shape or conceptual domain). This exclusion is important because contemporary computational models of category selectivity in the vOT suggest that at least for faces versus words, categorically distinct representations can already be observed in category-selective vOT regions during passive viewing, and those representations are further amplified by task-driven conceptual processing (Kay & Yeatman, 2017).

Relatedly, it is noteworthy that the candidate numeral-preferring region in the right parietal lobule is the only other region that showed evidence of greater numeral sensitivity. This observation is not only consistent with Price and Devlin's (2011) interactive account, but is also consistent with the hypothesis that the specific localization of the pITG-numerals is due to its biased connectivity with parietal regions thought to be involved in numerical magnitude processing (Abboud et al., 2015; Daitch et al., 2016; Hannagan et al., 2015; Nemmi et al., 2018).

Although intracranial electrophysiological recordings have begun probing the extent to which the numeral preference observed in the right pITG-numerals and right parietal lobe are dependent on each other, and in which direction, many of the findings are situated within an arithmetic context, which do not allow the dissociation of the contributions of sensory and conceptual processing (Baek et al., 2018; Daitch et al., 2016; Pinheiro-Chagas et al., 2018). With the surge in findings of relatively more robust pITG involvement during arithmetic and high-level mathematical tasks (Amalric & Dehaene, 2016, 2019; Baek et al., 2018; Bugden et al., 2019; Daitch et al., 2016; Grotheer et al., 2018; Hermes et al., 2017; Pinheiro-Chagas et al., 2018), there seems to be a shift in focus from a numeral-preferring ITG region to the surrounding “math-preferring” ITG region (Grotheer et al., 2018). However, even within the math-preferring ITG region, there is evidence for preference to Arabic numerals than to number words during an addition task (Baek et al., 2018), which suggests that there is non-trivial neural specialization for numeral processing. Moreover, individual differences in digit-specific activation during a digit detection task (i.e., whether a digit is present in a letter string) in the right pITG correlated positively with calculation competence (Pollack & Price, 2019). The stimulus-driven specialization of the pITG region must therefore be a product of learning, and may have bidirectional relations with the development of math competence. Hence, processing of numerals in the pITG as a distinct perceptual object category should also be an active area of investigation that is complementary to the investigation of pITG in mathematical tasks.

No Evidence of Abstract Shape Processing in Right pITG-numerals

In addition to the biased connectivity hypothesis, it has been argued that the lateral localization of pITG-numerals is partly, but necessarily accounted for by its role in detecting gross shapes of objects (e.g., relative to faces and houses) (Hannagan et al., 2015). In the shape

hypothesis, shape is defined as “a representation of the adjacency of the component parts of an object, that is at least partially invariant to translation, reflection, rotation, distance, and other variations in the stimulus” (Hannagan et al., 2015, p. 379). So, is the spatial dissociation of numeral-preferring and letter-preferring regions simply due to clustering of digit-shape and letter-shape “neural detectors”? Direct evidence for numeral visual form processing specifically has been lacking because a preference for digit shapes has been inferred from their category rather than shape per se (Shum et al., 2013). In fact, both visual form models examined in the current study revealed that digits and letters tend to look more alike within than across categories despite sharing same curvilinear features. Univariate contrast analyses clearly cannot dissociate shape from character category, or examine subtle differences in the configuration of features. Multivariate RSA is therefore most suited for examining the shape of characters independent of their categories, because it allows us to consider both within- and between-category similarities in shape in describing the empirical representational geometry of a region. Although we found an absence of evidence that the right pITG-numerals discriminates low-level visual features (Pixel Overlap model), there is conclusive evidence that it does not discriminate abstract shapes (Shape Distance model). Therefore, the current study provides the first direct evidence against shape processing as a primary role of the right pITG-numerals, and that it likely encodes the abstract digit identity and/or category.

Taking into account prior findings that the sensitivity of the right pITG-numerals may not be specific to Arabic digits – because it also respond more to soundscapes that represent “T”, “V”, and “X” as Roman numerals than as soundscapes that represent those same shapes as Roman letters (Abboud et al., 2015), and that it is equally responsive to Arabic digits, dice patterns, and finger representations (Grotheer et al., 2018) – we propose that this region is not

driven by visual form of Arabic numerals per se. In other words, in agreement with Grotheer and colleagues (2018), the selectivity observed appears to be to a numeral regardless of form, which, according to the Oxford Dictionary, is “a figure, symbol, or group of figures or symbols denoting a number”. Given that the region’s function is not constrained by visual form of numerals per se, and that its anatomical localization in the pITG is highly consistent across individuals and studies, we propose that researchers should refer to such a region as the “Inferior Temporal Numeral Area”.

Limitations

The datasets examined in this study were not designed specifically with multivariate pattern analyses of *individual characters* in mind, but rather the univariate mean response to an *entire character category*. Hence, the number of instances per exemplar in each run was limited. Response pattern estimates tend to be less reliable if they are estimated with fewer trials of the same exemplar and/or when the inter-trial intervals (ITI) are short (< 6 s) (Visser et al., 2016; Zeithamova et al., 2017). To overcome these issues, we modeled across repetitions of an exemplar within and across runs to enhance the estimation of exemplar-level representations. Response pattern estimation directly by combining runs is not uncommon, especially for RSA (e.g., Kriegeskorte, Mur, Ruff, et al., 2008; Rothlein & Rapp, 2014). Compared to Dataset 2, we had up to four and eight instances of an exemplar in Datasets 1 and 3 respectively, which may partially explain why we found evidence of the Numbers v. Others model in Datasets 1 and 3, but inconclusive evidence in Dataset 2. It is also possible that an ITI of 1s in Dataset 3 might only have allowed for shallow encoding of the characters. Even if that were true, the fact that we found evidence of the greater numeral sensitivity suggests that the effect in the pITG-numerals is robust enough to be detected.

Despite finding conclusive evidence of categorical distinction in the right pITG-numerals, it is evident that the effect sizes estimated for the ITG were all very low ($\tau_{\alpha}s < .02$; overall weighted $r = .008$) (**Figure 3-4**). The small effect sizes may suggest that there is still substantial variance within each participant's RDM that is not accounted for by all models tested. Moreover, the estimated “noise ceiling,” which is a measure of the inter-individual variability in participants' neural RDMs, was also low ($\tau_{\alpha}s < .17$) for the right pITG-numerals. This is not unexpected given that participants could have processed the task-irrelevant characters to varying extents (e.g., whether a character is attended to, and processed asemantically or semantically). Given that numeral-preferring pITG regions are intrinsically connected to parietal regions thought to subserve magnitude processing (Nemmi et al., 2018), future research may want to assess the contribution of semantic models (e.g., Lyons & Beilock, 2018). Yet, even with high inter-individual variability, it is remarkable that group-level numeral sensitivity was observed. It is also apparent that the noise ceiling in Dataset 1 was much higher (i.e., lower inter-individual variability) than those in Datasets 2 and 3. Even though Dataset 1 had half as many trials contributing to the estimated response pattern of each exemplar as Dataset 3, it had a longer ITI of 4 – 8 s compared to an ITI of 1 – 3 s in Dataset 3. This suggests that ITIs rather than number of repetitions may reduce inter-individual variability in the neural RDMs, presumably by the indirect benefit of improving the deconvolution of the hemodynamic responses, and/or the direct benefit of providing more time to attend to and encode the task-irrelevant characters. Future studies should aim for greater number of repetitions of each exemplar and/or have longer ITIs, especially if group-level effects such as those examined here are of interest.

Lastly, we used an a priori meta-analytic ROI, but there could be variability in the localization of the numeral-preferring region (e.g., see Glezer & Riesenhuber, 2013, for

variability in the localization of the “Visual Word Form Area”). Future work could therefore also investigate the representations in participant-specific ROIs.

Conclusions

Univariate analyses of task-irrelevant processing of numerals, letters, and novel characters have thus far not revealed evidence of any region in the vOT that showed a preference for numerals. In this study, we showed that a multivariate pattern analytic approach is more sensitive for uncovering categorical distinctions among written characters during a passive viewing task. Specifically, in a candidate numeral-preferring region in the pITG, we found that the organization of neural representations evoked by numerals, letters, and novel characters can be described by models that distinguish numerals and letters, and even a model that characterizes greater sensitivity for numerals. It is also less likely to be described by a model that characterizes greater sensitivity for letters, and unlikely by differences in abstract shapes (i.e., not visual form detection per se). It is likely that literacy and numeracy experiences may associate letters and digits with distinct processing goals (e.g., numerical relevance), and that the numeral-preferring pITG is part of a neural pathway that has been developed with automatic processing biases for stimuli with potential numerical relevance. In other words, “2” recruits the region because the brain predicts based on past experiences that this character is likely to be numerically relevant.

Authorship Contributions

Darren J. Yeo: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Courtney Pollack:** Conceptualization, Methodology, Formal analysis, Data curation, Writing - review & editing. **Rebecca Merkley:** Methodology, Formal analysis, Investigation, Data curation, Writing - review & editing, Funding acquisition. **Daniel Ansari:** Investigation, Resources, Data curation, Funding acquisition. **Gavin R. Price:** Conceptualization, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

CHAPTER 4

PROBING THE HEMISPHERIC ASYMMETRY OF FUNCTION AND REPRESENTATIONS IN THE BILATERAL “INFERIOR TEMPORAL NUMERAL AREAS”

Introduction

The findings in Chapters 2 and 3 suggest that the right “Inferior Temporal Numeral Area” (ITNA) is more reliable – spatially and/or functionally – in distinguishing numerals from other character types than its left counterpart. These findings of lateralization are not surprising against the backdrop of evidence of laterality in the literature (**Table 1-1**). This chapter aims to understand how the bilateral ITNAs may be functionally different, and explore whether functional and representational properties in each ITNA, and their asymmetry, are related to symbolic calculation skills.

In a study by Amalric and Dehaene (2016), the authors examined the response profile of each ITNA using separate regions-of-interest analyses. Professional mathematicians, compared to non-mathematicians, had an enhanced sensitivity to well-known mathematical constants in Arabic numeral format (e.g., 3.14159 [π]) relative to non-symbolic object categories in the left ITNA (Amalric & Dehaene, 2016). Mathematicians also had an enhanced sensitivity to mathematical formulas in the bilateral ITNAs (Amalric & Dehaene, 2016). Although the authors did not speculate how mathematical expertise might underlie the left lateralization, the frequent use of well-known constants and formulas could have led to their lexicalization in mathematicians. Such lexicalization may rely on a left-lateralized verbal pathway. It is not clear, however, whether the left lateralization can be observed within individuals. Besides, professional

mathematicians typically work with algebraic mathematical proofs that involve Roman and Greek letters more than they do with Arabic numerals because the former can be decontextualized from their numerical referents. Hence, it is also unclear whether hemispheric asymmetry exists in non-mathematicians, although a trend for greater sensitivity and selectivity to Arabic numerals in the right ITNA relative to the left ITNA was observed (see Figure 8E in Amalric & Dehaene, 2016).

Recently, an intriguing hemispheric asymmetry in the bilateral ITNAs was observed in a study by Pollack and Price (2019). In that study, adults performed a visual search task in which they had to detect whether a digit was present among a string of letters (e.g., ‘T S N 2 R’) or not (e.g., ‘A H T N R’). In a whole-brain localization analysis, they found that a region in the left (but not the right) ITG (MNI -57, -52, -11) was more engaged when a digit was present than when a digit was absent (i.e., [Digit Present > Digit Absent]; hereafter, we refer to this differential response as “digit sensitivity”) (**Figure 4-1(a)**). A brain-behavior correlational analysis revealed a homologous region in the right ITG (but not the left) (MNI 54, -52, -14) in which individuals with higher symbolic calculation skills showed greater digit sensitivity (**Figure 4-1(a)**). Both of these regions (with the left region mirrored in the right hemisphere) contained the peak coordinates of a meta-analytically identified right ITNA (MNI 55, -50, -12) from Chapter 2 (Yeo et al., 2017), suggesting that the regions could be considered ITNAs. Nonetheless, as with the findings by Amalric and Dehaene (2016), the hemispheric asymmetry of the bilateral ITNAs found by Pollack and Price (2019) was also based on separate group-level analyses – one localized by a contrast of two conditions, and another by a brain-behavior correlation. Whether the functional and representational properties of the bilateral ITNAs differ *within an individual* remains unexplored.

Using the triple-code model (Dehaene & Cohen, 1995) as a framework (see **Figure 1-1**), we speculated that the following three explanations might account for such a hemispheric asymmetry during visual search for digits observed by Pollack and Price (2019). The first possible explanation is that the left-hemispheric pathway may be recruited in most participants due to a reliance on the verbal system, possibly in retrieving the character names or identities as one scans the character string. Indeed, evidence from split-brain patients suggest a slight left-hemispheric advantage for same-different judgements of digits (L. Cohen & Dehaene, 1996; Corballis, 1994; Seymour et al., 1994).

The second possible explanation is that the right-hemispheric pathway may be less obligatory for a categorization task that does not require distinguishing between digits precisely (i.e., a digit was detected regardless of whether it was a 2 or a 7), but is almost always recruited in tasks in which the approximate quantitative meanings of numerals are necessary. It is therefore plausible that individuals who are more skilled in calculation would tend to automatically engage the right-hemispheric pathway to a greater extent. The latter hypothesis is plausible for several reasons. First, quantitative meanings of digits can be represented automatically even when they are task-irrelevant (e.g., during same-different or physical size judgements) (Dehaene & Akhavein, 1995; Henik & Tzelgov, 1982), and such automaticity develops with age, experience, and proficiency with a numerical symbol set (Girelli et al., 2000; Hochman Cohen et al., 2014; Rubinsten et al., 2002; Rubinsten & Henik, 2005). Second, the right hemisphere may be more efficient and reliable in representing analog or approximate quantities that are non-verbal in nature than the left hemisphere, which in turn may be more efficient and reliable in representing discrete or categorical processing of quantities that are more verbal in nature (Chassy & Grodd, 2012; Kimura, 1966; Kosslyn et al., 1989; Piazza et al.,

2006). Parietal activation engaged by number comparison also tends to be higher in the right hemisphere than the left hemisphere (Chochon et al., 1999; Dehaene, 1996; Pinel et al., 1999), which suggests a right-hemispheric bias for abstract number representations (see also Dickson & Federmeier, 2017; Jang & Hyde, 2020 for numerical comparison in the context of arithmetic verification). A right-hemispheric bias is also supported by developmental studies showing right parietal lateralization during numerical magnitude processing in infants, and the emergence of bilateral parietal involvement later in development after substantial experience with symbolic representations of numbers (number words and Arabic numerals) (Cantlon et al., 2006; Edwards et al., 2016; Emerson & Cantlon, 2015; Hyde et al., 2010; Izard et al., 2008; Libertus, Pruitt, et al., 2009; Vogel et al., 2015). Using large-scale resting-state datasets, Nemmi and colleagues (2018) recently showed that the right ITNA has intrinsic functional connections with the right intraparietal sulcus as early as age 3.5, whereas connectivity between these regions in the left hemisphere only appear around age 10. These developmental trends are also consistent with evidence of a genetic basis for quantity representation in the right parietal lobe, but not in the left (Pinel & Dehaene, 2013). Third, the hypothesis of a right-lateralized automatic activation of magnitude representations is also consistent with the finding in Chapter 3 that both meta-analytically defined right ITNA and right parietal region discriminate numerals from letters and novel characters during passive viewing, but evidence of a similar cross-regional pattern was not observed in the left hemisphere (Yeo et al., 2020). In sum, even though the involvement of the left hemisphere increases with experience in representing numbers symbolically, it may be related to the integration of the verbal code with the visual and magnitude codes. The evidence based on magnitude processing suggests that calculation skills independent of verbal processing may be associated with a right-hemispheric bias.

The third possible explanation is that intra-hemispheric interactions may be stronger than inter-hemispheric interactions, perhaps to increase efficiency of processing (Karolis et al., 2019; Ringo et al., 1994). Even though there could be inter-hemispheric informational exchange between the bilateral ITNAs, there is evidence suggesting that inter-hemispheric interaction impacts the efficiency of isolated hemispheres in identifying alphanumeric characters (Teng & Sperry, 1973). Such interference costs may underlie the need for separate numeral identification systems that Cohen and Dehaene (1995) proposed, rather than a unified one.

Current Study

In this study, we re-analyzed Pollack and Price's (2019) data with two aims. While Pollack and Price (2019) focused on *localization* of the ITNAs using mass-univariate voxel-wise analyses, our first aim was to use region-of-interest (ROI) based univariate analyses and within-participant comparisons to further characterize the hemispheric asymmetries of the bilateral ITNAs in their digit sensitivities and their relation to symbolic calculation skills. Our second aim was to characterize any hemispheric asymmetries in the representational properties of the ITNAs indexed by multivoxel pattern analyses, and also to relate them to calculation skills. This in-depth investigation of the functional and representational properties allows us to test and further inform models of numerical processing, such as the triple-code model.

Is There Within-Individual Hemispheric Asymmetry in the ITNAs' Digit Sensitivity?

Due to the original univariate findings (Pollack & Price, 2019), we predicted that, on average, digit sensitivity would be greater in the left than in the right ITNA. We also predicted that higher calculation skills would be associated with less left lateralization⁵, or greater right

⁵ Conditioned on the observed positive correlation with digit sensitivity in the right IT reported by Pollack and Price (2019), an association between calculation skills and greater left lateralization would imply a stronger relation between calculation skills and digit sensitivity of the left ITNA than with the digit sensitivity of the right ITNA. This

lateralization.

Do the Multivoxel Patterns in the ITNAs Discriminate Between Digits and Letters?

Although the right ITNA did not appear to be sensitive to digits in terms of its regional mean response amplitude, an analysis of multivoxel response patterns may reveal category discriminability (Yeo et al., 2020; Chapter 3). Specifically, the response patterns in the ITNAs evoked by each single target allowed us to examine the multi-dimensional organization of exemplar-level neural representations – commonly referred to as “representational geometry” within the representational similarity analytic (RSA) framework (Kriegeskorte, Mur, & Bandettini, 2008). The neural representations of detected digit and letter targets could form separate clusters. Based on the original univariate finding on digit sensitivity, we predicted that the left ITNA would show greater category discriminability than the right ITNA. However, based on prior meta-analytic and multivoxel pattern findings in Chapters 2 and 3 (Yeo et al., 2017, 2020), category discriminability could also be greater in the right than in the left ITNA. We also predicted that higher calculation skills would be associated with greater category discriminability in both ITNAs, but we had no specific prediction for an effect of laterality.

Do the Multivoxel Patterns in the ITNAs Discriminate Between Digit Exemplars?

Regardless of whether category discriminability was evident in the ITNAs, it would be informative to assess whether digit discriminability was evident. This is because the identity and category of a character are thought to be represented in parallel rather than serially (McCloskey & Schubert, 2014; Taylor, 1978). We predicted that digit discriminability would be observed in both ITNAs, but it would be greater in the left than in the right ITNA. We also predicted that higher calculation skills would be associated with greater digit discriminability in both ITNAs,

was, however, not the case.

but we had no specific prediction for an effect of laterality.

Are Digit Representations Organized Similarly Between the ITNAs?

Even if digit discriminability was not evident on average across participants, the relative pairwise dissimilarities between digits within participants could still reveal a meaningful organization⁶. Hence, it would be informative to explore whether a discernable representational geometry can be uncovered. Given the poor temporal resolution of fMRI and that the numeral identification systems function as an integrative intra-hemispheric network, the representational geometries being studied in the ITNAs would likely reflect both bottom-up and top-down influences from the visual, verbal, and magnitude codes within each hemispheric pathway (cf. Bar et al., 2006; Gwilliams & King, 2020; Kay & Yeatman, 2017; C. J. Price & Devlin, 2011). For instance, if response patterns in the left ITNA are influenced by phonological representations from the verbal code, we would expect characters with similar phonological form to evoke similar response patterns in the left ITNA. Likewise, if the response patterns in both ITNAs are influenced by the magnitude code (Grotheer et al., 2018), we would expect digits that are numerically closer (e.g., 8 vs. 9) to evoke more similar response patterns than digits that are numerically distant (e.g., 2 vs. 9) (Piazza et al., 2007; Vogel et al., 2015, 2017). Hence, at a coarse level, we predicted that the representational geometries of digits would be different between the bilateral ITNAs. We also predicted that higher calculation skills would be associated with greater dissimilarity in the representational geometries of digits in the ITNAs.

Finally, we explored whether we could describe the representational geometries using

⁶ Exemplar discriminability is typically indexed by the mean correlational distance between exemplar pairs (a single summary measure), whereas the representational geometry takes into account the rank order of the correlational distances of the exemplar pairs (a multivariate measure). In other words, even if the mean correlational distance were zero, the correlational distances could still be rank ordered to uncover an underlying structure.

hypothetical representational models. We predicted that the representational geometries in both ITNAs would not be adequately described by visual form similarity (Yeo et al., 2020; Chapter 3). Based on the triple-code model, we predicted that the representational geometry of digits in the left ITNA would reflect an organization based on phonological similarities of the verbal code more than the numerical distances of the magnitude code, whereas the representational geometry of digits in the right ITNA would reflect an organization based on similarities in the magnitude code more than the verbal code.

Methods

Participants

Thirty-two neurologically typical and right-handed adults ($M_{Age} = 19.38$, $SD = 1.50$, 21 females) were included in the current analyses. These are the exact same data that were analyzed in the initial univariate functional localization study by Pollack and Price (2019). The study was approved by the university's Institutional Review Board, and all participants gave written informed consent.

As the current aims precluded a priori power analyses, we performed sensitivity analyses using G*Power 3.1.9.6 (Faul et al., 2009) to determine the minimal effect sizes we could detect with 80% power and $N = 32$. For $\alpha = .05$, we expected to detect effect sizes that exceed the following critical values: Cohen's $|d|$ or $|d_z| = 0.449$ and 0.511 for one- and two-tailed (one-sample or paired) t -tests, respectively; $|r| = .296$ and $.349$ for one- and two-tailed correlation analyses, respectively. The critical d s are higher than the average $d \approx .4$ expected in psychological studies (see Brysbaert, 2019 for a review), so some true effects might be undetectable with the current data. However, the critical r s indicate that we would be able to detect effect sizes typical of non-circular, univariate brain-behavior correlations (modal $r \approx .55$)

(Vul et al., 2009), as well as of relations between vOT measures and perceptual expertise (e.g., $r_s \geq .49$ for word inversion sensitivity and multivariate "Visual Word Form Area" laterality measure in Carlos et al., 2019; $r_s \geq .49$ for reading performance and bilateral fusiform gyri activation in Feng et al., 2020; $r_s \geq .38$ for car expertise and right "Fusiform Face Area" activation in McGugin et al., 2014, 2015). To distinguish whether non-significant effects were true negatives or due to a lack of statistical power, we also reported Bayes factors.

Tasks

fMRI Tasks

To localize digit-related vOT regions, participants completed visual search tasks involving alphanumeric characters in the MRI scanner. During digit detection, participants determined whether a digit was present among a string of letters (**Figure 4-1(a)**) by pressing one of two assigned buttons. During letter detection, participants determined whether a letter was present among a string of digits (**Figure 4-1(b)**). The single target digit or letter, which could be digits 1 – 9 and letters A, C, D, E, H, R, N, S, and T, was presented in either the 2nd, 3rd, or 4th position of a 5-character string. Each target digit/letter exemplar was presented in three unique strings per run (see **Table B-1** for stimulus list). Each run comprised 16 s of fixation baseline at the start and the end of the run, and 54 trials (27 Target Present trials across all 9 digit/letter exemplars, and 27 Target Absent trials). On each trial, the character string was presented for 1 s, and the inter-stimulus interval was 2, 4, or 6 s ($M = 4$ s). Justifications for the design of the stimulus sets can be found in Pollack and Price (2019).

All participants completed four runs each of digit detection and letter detection, hence, across all four runs, each target digit/letter exemplar was presented a total of 12 times. Based on pre-determined criteria for excessive motion (> 3 mm maximum displacement and/or three

degrees of volume-to-volume displacement), one digit run and one letter run from one participant, and one letter run from another participant, were excluded from the analyses in Pollack and Price (2019). All analyses reported here thus included at least three runs each of digit detection and letter detection for every participant. Mean accuracies for the conditions critical to our key analyses (Digit Present and Letter Present) were at least 92%. See **Table B-3** for full descriptive statistics of the task performance, and Pollack and Price (2019) for analyses on the behavioral measures.

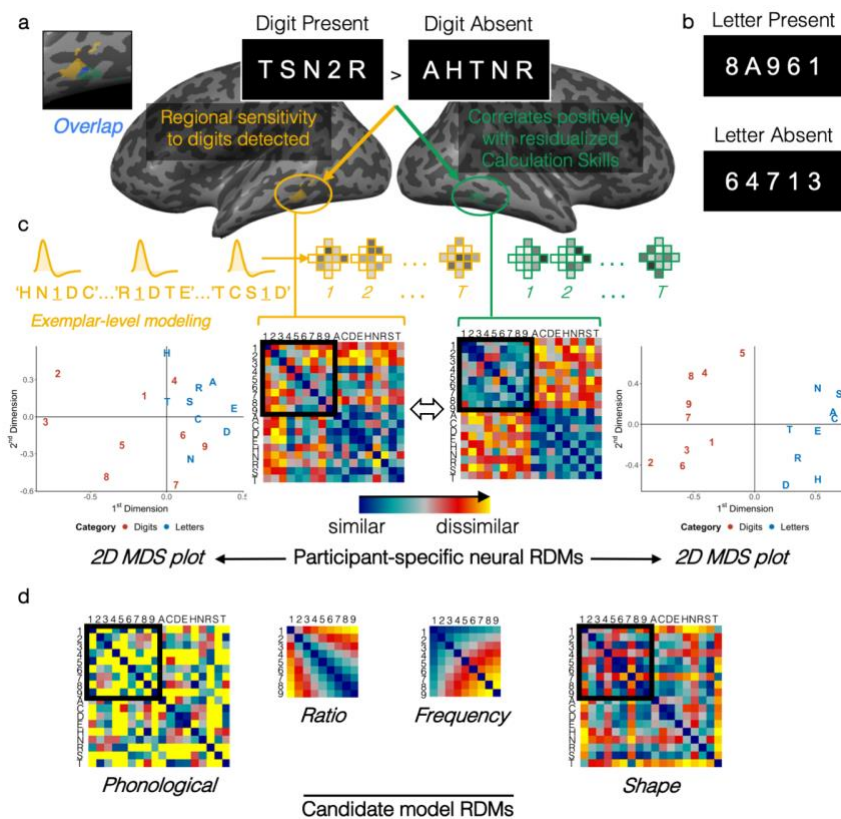


Figure 4-1. Regions of interest (ROIs) and a schematic of the key components of the representational similarity analyses (RSA) (a) Left and right IT ROIs from Pollack and Price (2019) derived from a random-effects analysis of [Digits Present > Digits Absent] contrast maps (yellow), and a correlational analysis of those contrast maps with calculations skills (green), respectively, and their overlap (blue). (b) Example

stimuli for letter detection task. (c) For each ROI, response patterns reliably evoked by correctly detected target across different character strings [Digit/Letter Present > Fixation] were correlated in a pairwise manner to construct participant-specific representational dissimilarity matrices (RDMs). 2D plots of the representational geometries in each ROI obtained by multidimensional scaling. Hemispheric asymmetry was assessed by correlating the (full or subset) RDMs of the left and right IT ROIs. (d) Candidate model RDMs.

Standardized Cognitive Assessments

Calculations skills were measured using the Calculation and Math Fluency subtests of the Woodcock-Johnson III Tests of Achievement (WJ III ACH; Woodcock et al., 2001). The Calculation subtest is an untimed test that assesses arithmetic (with natural and rational numbers), algebra, trigonometry, and calculus. The Math Fluency subtest assesses the ability to solve as many simple addition, subtraction, and multiplication problems with the numerals 0–10 as possible within three minutes. A Calculation Skills cluster score was computed from a composite of Calculation and Math Fluency measures. As a proxy for domain-general symbol decoding, the Letter-Word Identification (ID) subtest of the WJ III ACH was used. The Letter-Word ID subtest is an untimed test that assesses the ability to read aloud a list of letters and words accurately. See **Table B-3** for descriptive statistics of these measures. For consistency with the original study, in all analyses involving the Calculation Skills cluster measure, we used scores that were residualized for Letter-Word ID scores (i.e., the residuals after regressing the Calculation Skills standard scores on Letter-Word ID standard scores).

Neuroimaging Data Acquisition

Structural and functional brain images were acquired using a 3T Philips Intera Achieva scanner with a 32-channel head coil. High-resolution 3D anatomical scans were collected over approximately 6 min with TR/TE = 8.1/3.8 ms, flip angle = 5°, field of view (FOV) = 256 mm,

and 1 mm isotropic voxels. T2*-weighted single-shot echo-planar imaging sequence functional images were acquired with TE = 25 ms, TR = 2000 ms, flip angle = 90°, FOV = 240 mm, matrix size = 96 × 96 mm, 2.5 × 2.5 × 3 mm³ voxels, with 0.25 mm gap between the 3-mm thick slices, 40 slices, and 151 volumes per run. Five additional dummy volumes acquired at the start of each run to allow for steady-state magnetization were discarded.

fMRI Data Preprocessing

Structural and functional images were preprocessed and analyzed using BrainVoyager 20.4 (Brain Innovation, Inc., Maastricht, the Netherlands). Functional images were corrected for differences in slice time acquisition (cubic spline interpolation), head motion (trilinear-sinc interpolation), and high-pass filtered (GLM approach with Fourier basis set, 2 cycles) to remove linear and non-linear trends. Functional data were co-registered to the structural data using boundary-based registration, normalized to MNI space, and re-sampled to 3-mm isotropic voxels. Univariate analyses were conducted on spatially-smoothed data with a Gaussian kernel of 6 mm at full-width half-maximum. Multivariate analyses were conducted on spatially unsmoothed data.

Neuroimaging Statistical Modeling

Univariate Analyses

For each participant, all included runs were modeled with a two-gamma hemodynamic response function. The data were analyzed simultaneously using a random-effects multi-subject General Linear Model (GLM), corrected for serial correlations with a second-order autoregressive method. The GLM included a regressor for Digit Present (correct only), a regressor for Digit Absent (correct only), a regressor for Letter Present (correct only), a regressor for Letter Absent (correct only), one regressor for errors of commission and omission, and six

regressors of motion parameters (translational and rotational in x , y , and z axes) for each run.

Multivariate Analyses

Each of the nine target digit/letter exemplars were presented up to 12 times across all digit or letter detection runs. Given the presence of characters from the non-target category (e.g., letters in ‘T S N 2 R’ for digit detection) on each Target Present trial, we attempted to minimize the influence of the non-target characters in our analyses by modeling all instances of a target exemplar that was detected correctly across all runs (e.g., [‘H N 1 D C’, ‘R 1 D T E’, ‘T C S 1 D’] \times 3 or 4 runs to estimate the voxel-wise response to a detected ‘1’). This approach ensured that the voxel-wise responses estimated from the [Target Present - Fixation] contrast would be reliably specific to the target exemplar common to the modeled trials (see **Figure 4-1(c)**). However, as the original study was not designed with exemplar-level representations in mind, some characters from the non-target category always co-occurred with some of the target exemplars, such as ‘D’ was always present in all Digit Present trials with the target ‘1’. As passive-viewing paradigms revealed that there could be bottom-up representations of unattended characters in our regions of interest (Yeo et al., 2020; Chapter 3), it is possible that what we assumed to be a reliable representation of a detected ‘1’ could in fact also be a reliable representation of an undetected ‘D’. Yet, the neural representations of attended stimuli or dimensions of a stimulus tend to be amplified or expanded, whereas unattended stimuli or dimensions tend to be attenuated or compressed (e.g., Castaldi et al., 2019; Çukur et al., 2013; Nastase et al., 2017). This led us to assume that the representation of the co-occurring non-target characters would not be reliably captured in our response estimates. Supplemental analyses (**Table B-4**) confirmed that co-occurring characters from the non-target category were not as reliably represented in our regions of interest as when they were from the target category.

To further ensure that we could reliably estimate the response patterns at the exemplar level for each participant, we used an arbitrary cut-off of 50% accuracy (i.e., at least 6 correct Target Present trials) per exemplar as an inclusion criterion for multivariate analyses. All participants but one had at least six correct trials per digit/letter exemplar to reliably estimate an exemplar-level response pattern. That one participant had only 1 to 5 correct trials for 7 out of 9 letters. Hence, for all analyses involving letter exemplar representations, we excluded that participant. The mean number of remaining correct Target Present trials per exemplar was 11 for both digits and letters.

For each participant, all included runs were modeled with a two-gamma hemodynamic response function and analyzed simultaneously using a fixed-effects single-subject GLM, corrected for serial correlations with a second-order autoregressive method. The GLM included a regressor each for the nine target digits (correct only), a regressor for Digit Absent (correct only), a regressor for each of the nine target letters (correct only), a regressor for Letter Absent (correct only), four regressors for errors of commission and for omission (separately modeled for digit and letter detection), and six regressors of motion parameters for each run.

Regions of Interest (ROIs)

Figure 4-1(a) shows the left and right IT ROIs from Pollack and Price (2019). A *t*-test of [Digit Present > Digit Absent] contrast maps revealed a left IT cluster, and a brain-behavior correlation of the same contrast maps and residualized calculations skills scores revealed a right IT cluster. The statistical thresholds used for both analyses were identical: voxel-level threshold of $p < .005$, and cluster-level threshold of $p < .05$ via Monte Carlo simulations. The peaks of the ROIs are within ± 3 mm (i.e., one functional voxel) along each dimension (left: MNI -57, -52, -11 and right: MNI 54, -52, -14), but the spatial extents of the ROIs (left: 728 mm³; right: 670

mm³) are non-homotopic. Flipping one ROI onto the other hemisphere revealed that the spatial overlap is about 15% (105 mm³), which contains the meta-analytic peak of the ITNA from Chapter 2 (Yeo et al., 2017) (MNI 55, -50, -12) (see **Figure 4-1(a)**). Nonetheless, in terms of 3-mm isotropic functional voxels, both ROIs have 59 voxels as features for the multivariate analyses.

Statistical Analyses

Univariate Analyses

To fully characterize the IT ROIs in terms of their regional mean digit sensitivity (i.e., response amplitudes evoked by detected digits) beyond the findings reported by Pollack and Price (2019), we conducted the following post hoc analyses⁷: (1) We tested whether, on average, there was within-individual asymmetry in the left and right ITs' digit sensitivity (i.e., mean beta values from the [Digit Present – Digit Absent] contrast) (paired samples *t*-test, two-tailed; Bayesian prior: Cauchy distribution, scale = 0.707); (2) We also probed whether the degree of right lateralization in the digit sensitivity was positively correlated with calculation skills. To compare the degree of lateralization across individuals normalized for individual differences in digit sensitivity, we computed a dissimilarity-like⁸ laterality index, $LI = \frac{L - R}{\max(|L|, |R|)}$, where L and R are the mean digit sensitivity of the left and right ROIs, respectively (Seghier, 2019). A positive LI indicates left lateralization and a negative LI indicates right lateralization. The participant-specific LI scores were then correlated with the residualized Calculation Skills scores

⁷ Although these analyses are post hoc and therefore non-independent from the analyses reported by Pollack and Price (2019), the authors did not compute a within-participant difference score to directly assess hemispheric asymmetry.

⁸ A more widely used formula is $LI = \frac{L - R}{|L| + |R|}$ (for reviews, see Bradshaw et al., 2017; Seghier, 2008). This LI formula is typically used for classification purposes (left- or right-lateralized, or bilateral), but is problematic for analyses of individual differences because it lacked meaningful variation (Bradshaw et al., 2017; Jansen et al., 2006) and it is not a proper distance metric (Seghier, 2019).

(Pearson's correlation, two-tailed; Bayesian prior: Stretched beta prior width = 1).

As the effects for the [Digit Present – Digit Absent] contrast could be driven by the Digit Present and/or Digit Absent condition, we further explored whether these predicted effects also pertained to the condition versus baseline contrasts (i.e., [Digit Present – Fixation] and [Digit Absent – Fixation]). We also performed an identical set of analyses for the letter detection task to assess the category specificity of these findings.

Representational Similarity Analyses

Neural Representational Dissimilarity Matrices (RDMs). For each participant and each ROI, the response pattern evoked by each correctly detected digit or letter exemplar from the [Digit/Letter Present – Fixation] contrast was characterized by the spatial distribution of t -values (Misaki et al., 2010) (**Figure 4-1(c)**). Non-informative voxels (i.e., $t = 0$ across all conditions due to intensity-based thresholding) were excluded from subsequent analyses. This resulted in 53–59 voxels ($M = 58.09$) per participant for the left IT ROI, and 58–59 voxels ($M = 58.97$) per participant for the right IT ROI. We then computed the pairwise correlational distances ($1 - \text{Pearson's } r$) to construct participant-specific 18×18 representational dissimilarity matrices (Full-RDMs) for the left and right IT ROIs (**Figure 4-1(c)**). For all key analyses, we focused on the 9×9 Digits-RDMs (i.e., digits subset of the Full-RDMs; $N = 32$). To assess category specificity as well as features applicable to both categories (i.e., phonology and shape), we also conducted control analyses on the 9×9 Letters-RDMs (i.e., letters subset of the Full-RDMs; $N = 31$) and exploratory analyses on the Full-RDMs (i.e., including all digits and letters; $N = 31$), respectively.

Category Discriminability. To assess the degree of category discriminability within each ROI, we computed a participant-specific category discriminability index (CDI) using the

formula $CDI = M_{\text{between-category dissimilarities}} - M_{\text{within-category dissimilarities}}$ from the Full-RDM. A higher CDI indicates greater category discriminability. We tested whether the mean CDI was statistically greater than zero (one-sample t -test, right-tailed; Bayesian prior: Cauchy distribution, scale = 0.707). To assess whether greater category discriminability in each ROI was associated with higher calculation skills, we correlated the participant-specific CDIs with the residualized Calculation Skills scores (Pearson's correlation, right-tailed; Bayesian prior: Stretched beta prior width = 1). We also tested whether laterality of category discriminability ($LI = \frac{L-R}{\max(|L|, |R|)}$) was associated with higher calculation skills (Pearson's correlation, two-tailed; Bayesian prior: Stretched beta prior width = 1).

Although both ROIs were either directly or indirectly localized using the contrast [Digit Present > Digit Absent] (and not [Letter Present > Letter Absent]), one might argue that univariate activation differences between digits and letters detected already presupposed category discriminability in these ROIs. However, these new analyses focused on the similarity of exemplar-level multivoxel response patterns for Target Present relative to baseline (e.g., how similar the response pattern for a detected '4' was to a detected 'N'). Moreover, their similarities were assessed using correlational distance, which standardizes the response amplitudes and therefore reduces the influence of mean amplitude differences.

Exemplar Discriminability. To assess the degree of exemplar discriminability within each ROI, we first split each participant's data into two halves (i.e., odd runs and even runs), and computed the reliabilities of the response patterns between the two halves. We then performed Spearman-Brown correction (Brown, 1910; Spearman, 1910; corrected $r = 2r/(1+r)$) to estimate the reliabilities for the full dataset, and then computed the correlational distance to construct an adjusted split-data RDM. The cells along the diagonal of the split-data RDM are the within-

exemplar dissimilarity estimates (e.g., how dissimilar was the response pattern of ‘2’ in the odd runs to that in the even runs), which reflects measurement noise (Nili et al., 2020). The off-diagonal cells are the between-exemplar dissimilarity estimates. We then computed a participant-specific exemplar discriminability index (EDI) using the formula $EDI = M_{\text{between-exemplar dissimilarities}} - M_{\text{within-exemplar dissimilarities}}$ (Nili et al., 2020). A higher EDI indicates greater exemplar discriminability. We then analyzed the EDI in an approach identical to that for CDI. A *t*-test is valid in practice as it provides adequate control of the false positive rates and is robust to assumption violations (Nili et al., 2020).

We performed the EDI analyses for digits and letters separately. The letter analysis was included to assess category specificity. Two participants who did not have an equal number of even and odd runs for the digit and/or letter detection were excluded from the EDI analyses ($N = 31$ for split-data Digits-RDMs, and $N = 30$ for split-data Letters-RDMs).

Hemispheric Asymmetry of Representational Geometries. For each individual, we computed the similarity between the Digits-RDMs of the left and right IT ROIs using Spearman’s correlation, and transformed them using Fisher’s z transformation (ρ_z) to render them appropriate for parametric statistical tests. $\rho_z > 0$ indicates that the representational geometries of digits are similar between the hemispheres, whereas $\rho_z \leq 0$ indicates hemispheric asymmetry in the representational geometries. To assess our prediction that, on average, there was hemispheric asymmetry in the representational geometries, we compared the alternative hypothesis that the mean similarity > 0 to the null hypothesis that the mean similarity ≤ 0 (one-sample *t*-test, right-tailed; Bayesian prior: Cauchy distribution, scale = 0.707). Next, to assess whether greater hemispheric asymmetry (i.e., lower similarity) in the representational geometries of digits was associated with higher calculation skills, we correlated the participant-specific

similarity scores with the residualized Calculation Skills scores (Pearson's correlation, left-tailed; Bayesian prior: Stretched beta prior width = 1).

Finally, to assess category specificity, we performed an identical set of analyses on the degree of hemispheric asymmetry in representational geometries of letters.

Representational Content. To probe the representational content of each IT ROI, we constructed four candidate model RDMs that are characterized by similarity in phonology, numerical magnitude, and visual form (**Figure 4-1(d)**).

Phonological Model. We constructed an 18×18 phonological model RDM from an empirically derived character-name confusion matrix that described the perceptual confusion of participants who were asked to identify a digit or letter aurally presented in noise (Hull, 1973). We converted the asymmetric similarity-based confusion matrix that comprises the frequencies of confusions between every stimulus-response pair (e.g., responding '8' to stimulus '6', or responding 'A' to stimulus '8') into a dissimilarity matrix using the following approach: For each column (stimulus), we normalized its row values (frequencies for each response to a particular stimulus) by dividing them by the sum of the frequencies. This transformed the frequencies to proportions or the probability that one character tends to be confused for another. Next, we averaged the proportions in the upper and lower triangles of the asymmetric matrix to obtain a symmetric confusion matrix. Finally, we transformed the symmetric confusion matrix to a dissimilarity matrix by subtracting each value from the maximum similarity value.

Numerical Models. Lyons and Beilock (2018) reported that in many brain regions including the ventral temporal-occipital junction, the representational similarity space of nonsymbolic quantities (e.g., sets of dots) could be predicted by the ratio of any two quantities, and the representational similarity space of Arabic digits could be predicted by the frequency of

their co-occurrence more so than ratio. A frequency-based account of digit representations is also consistent with behavioral work (Dehaene & Mehler, 1992; Kojouharova & Krajcsi, 2019; Krajcsi et al., 2016). In order to fully explore the nature of the representational geometry of digits in the IT ROIs, we constructed two 9×9 numerical magnitude model RDMs based on ratio and frequency, identical to those used by Lyons and Beilock (2018). The Ratio model was first constructed using the ratio between the quantities represented by a pair of digits n_i and n_j as a measure of similarity, where $\text{ratio} = \frac{\min(n_i, n_j)}{\max(n_i, n_j)}$, and larger values indicate greater similarity. The RDM was derived using the inverse of the ratios such that larger values indicate greater dissimilarity.

The Frequency model is based on the frequency of co-occurrence of any given pair of digits as a measure of similarity. According to Benford's (1938) law, the probability of encountering a given digit in the leftmost position of multi-digit numerals, $P(n)$, is $\log_{10}(n+1) - \log_{10}(n)$ (also see Dehaene & Mehler, 1992). We then computed the probability of the joint frequency of each pair of digits using $P(n_i) \times P(n_j)$, where larger values indicate greater similarity (Lyons & Beilock, 2018). Likewise, the RDM was derived using the inverse of the probability of the joint frequency such that larger values indicate greater dissimilarity.

Visual Form Model. We constructed an 18×18 Shape model RDM using a computational algorithm that is based on the similarity in the “context” of sampled points on a shape (i.e., how one point relates to all other points on a shape) and the degree to which one shape has to be deformed to map onto another shape (Belongie et al., 2002). The performance of this algorithm was validated on classification of handwritten digits (Belongie et al., 2002), and an identical model was used in an RSA study by Yeo and colleagues (2020; Chapter 3, see Appendix A for computational details).

Figure B-1 – Figure B-4 are multidimensional scaling plots that illustrate the 2D representational geometry of the digit and letter exemplars in each model. The bivariate rank correlations of the models are reported in **Table B-2**. No pairs of models are highly similar ($\rho < .16$), although the Ratio and Frequency models have a strong negative correlation ($\rho = -.65$). Hence, greater support for the Ratio model would likely indicate less support for the Frequency model, and vice versa.

Similarity Between Neural and Model RDMs. The degree to which the 9×9 neural Digits-RDMs can be described by each model RDM was examined using the RSA toolbox (Nili et al., 2014). We correlated the neural and model RDMs (only one-half of each symmetric matrix) using Spearman’s rank correlation, and Fisher’s z -transformed the correlation coefficients for subsequent parametric statistical tests on those coefficients. For each model, we tested whether the mean correlation coefficient was statistically greater than zero (one-sample t -test, right-tailed; Bayesian prior: Cauchy distribution, scale = 0.707). To estimate the upper and lower bounds of the maximum similarity that any model could achieve given the degree of between-participant variability, a “noise ceiling” was computed using the approach proposed by Nili and colleagues (2014). We also tested (a) within each ROI, whether the mean correlation coefficients between any pair of models were statistically different, and (b) for each model, whether the mean correlation coefficients between the left and right ROIs were statistically different (paired-sample t -tests, two-tailed; Bayesian prior: Cauchy distribution, scale = 0.707).

As the Phonological and Shape model RDMs are not category-specific, we also examined whether the 18×18 neural Full-RDMs (i.e., the whole alphanumeric set; see **Figure 4-1(c)**) were similar to the full versions of the Phonological and Shape model RDMs (**Figure 4-1(d)**).

We corrected for multiple comparisons separately for each group of tests (a) across

models within each ROI for group-level neural RDM-model RDM similarity (i.e., 5 tests), (b) across pairs of models within each ROI for within-participant model comparisons (i.e., 6 tests), and (c) across models for within-participant comparison of left and right ROIs (i.e., 4 tests), by controlling for false-discovery rate (FDR) at $q < 0.05$ (Benjamini & Hochberg, 1995). All p -values reported in the Results section are uncorrected for multiple comparisons, and statistically significant ones were noted if they also survived an FDR-correction.

Handling of Bivariate Outliers

To assess the robustness of the correlation analyses to bivariate outliers, we used the Minimum Covariance Determinant approach to estimate the bivariate location and scatter from 75% of the data (i.e., assuming no more than 25% of outlying values), and a chi-square distribution ($df = 2$) with $\alpha = .001$ (99.9% percentile) as an outlier criterion (Leys et al., 2018, 2019). As there is no theoretical basis for deciding whether an outlier could rightfully belong to the distribution of interest, we reported the affected correlation coefficients with and without the outliers (i.e., “skipped correlation”; Rousselet & Pernet, 2012; Wilcox, 2004).

Comparison of Correlation Coefficients

As the difference between a pair of statistically significant and non-significant correlation coefficients may not be itself statistically significant (Gelman & Stern, 2006; Nieuwenhuis et al., 2011; Rousselet & Pernet, 2012), whenever necessary, we used the *R* package ‘cocor’ (Diedenhofen & Musch, 2015) to conduct a suite of statistical tests to compare whether a pair of correlation coefficients differed significantly. The default suite of 10 tests for comparing two dependent correlations (i.e., r_{jk} vs. r_{jh} , accounting for r_{kh}) was used to make inferences. By and large, the data were likely underpowered to detect a difference between effect sizes, so we caution the inferences made about the null differences in effect sizes.

Bayesian Statistical Inferences

To facilitate Bayesian inferences for each test, we report the Bayes factor in favor of the hypothesis supported using the following conventions in JASP (JASP Team, 2020): For a two-tailed test, BF_{10} indicates the number of times the observed data are more likely to occur under the alternative hypothesis than under the null hypothesis, whereas BF_{01} indicates greater support for the null hypothesis. For a one-tailed test, BF_{+0} and BF_{-0} indicate the number of times the observed data are more likely to occur under the alternative ('+' : right-tailed or positive correlation, or '-' : left-tailed or negative correlation) hypothesis than under the null hypothesis, whereas BF_{0+} and BF_{0-} indicate greater support for the null hypothesis. Whenever the evidence in support of one hypothesis relative to another is less than 3 times, we inferred that the evidence is inconclusive, and that the data are insensitive to the hypotheses tested (Dienes, 2016; Dienes & Mclatchie, 2017).

Results

Regional Mean Digit Sensitivity

Digits

In Pollack and Price (2019), the left IT ROI was localized by its high digit sensitivity (i.e., [Digit Present – Digit Absent] contrast), whereas the right IT ROI was localized separately by the relation between individual differences in digit sensitivity and calculation skills. Here, we probed whether these hemispheric differences would hold when we directly compare them within participants. On average, there was no hemispheric asymmetry in the regional mean digit sensitivity (left: $M = 0.17$, $SD = 0.22$; right: $M = 0.13$, $SD = 0.46$, difference: $M = 0.05$, $SD = 0.42$), $t(31) = 0.62$, $d_z = 0.11$, $p = .541$, $BF_{01} = 4.44$ (**Figure 4-2(a)**). However, individuals with higher calculation skills had greater right lateralization in their mean digit sensitivity, $r(30) = -$

.45, $p = .009$, $BF_{10} = 5.69$ (**Figure 4-2(b)**).

Although individual differences in the [Digit Present – Digit Absent] contrast in the left IT did not correlate significantly with calculation skills in the original whole-brain correlational analysis reported by Pollack and Price (2019), we tested whether such a relation could be observed using an ROI approach. Digit sensitivity in the left IT was also positively correlated with calculation skills, $r(30) = .42$, $p = .008$, $BF_{+0} = 6.77$. However, the correlation coefficient in the left IT did not differ significantly from that of the right IT ($r(30) = .62$), $ps > .206$.

In sum, there was, on average, no hemispheric asymmetry in digit sensitivity, and no evidence that digit sensitivity in the left and right IT differed qualitatively in their relation to calculation skills. However, consistent with our prediction, the *within-individual difference* in the digit sensitivity between hemispheres was related to calculation skills.

As the [Digit Present – Digit Absent] contrast could be driven by Digit Present and/or Digit Absent, we wanted to understand the nature of involvement of the IT ROIs for each condition. Contrary to the findings above, the condition-wise regional mean response amplitudes (i.e., condition > fixation) were strongly right-lateralized for both Digit Present (left: $M = 0.21$, $SD = 0.45$; right: $M = 0.88$, $SD = 0.78$, difference: $M = -0.67$, $SD = 0.72$) [$t(31) = -5.26$, $d_z = -0.93$, $p < .001$, $BF_{10} = 2061$] and Digit Absent (left: $M = 0.04$, $SD = 0.44$; right: $M = 0.75$, $SD = 0.64$, difference: $M = -0.72$, $SD = 0.74$) [$t(31) = -5.48$, $d_z = -0.97$, $p < .001$, $BF_{10} = 3642$] (**Figure 4-2(c)**). Individuals with higher calculation skills had greater response amplitudes for Digit Present in the right IT [$r(30) = .48$, $p = .003$, $BF_{+0} = 16.93$]. A similar relation was inconclusive in the left IT [$r(30) = .25$, $p = .084$, $BF_{0+} = 1.01$]. However, these correlation coefficients did not differ significantly ($ps > .182$). There was evidence that calculation skills were not positively correlated with the response amplitude for Digit Absent in the left IT [$r(30) = .05$, $p = .393$, BF_{0+}

= 3.64], but whether a positive correlation between calculation skills and response amplitude for Digit Absent in the right IT was inconclusive [$r(30) = .14$, $p = .224$, $BF_{0+} = 2.24$; $r_{skipped}(28) = .37$, $p = .023$, $BF_{+0} = 2.98$]. These correlation coefficients did not differ significantly between the left and right IT ($ps > .704$; after exclusion of outliers: $ps > .179$). In sum, there was only conclusive evidence of an association between calculation skills and Digit Present responses in the right IT, and a lack of association between calculation skills and Digit Absent (i.e., letters only) responses in the left IT.

These correlation coefficients also differed significantly for Digit Present and Digit Absent in both regions (left IT: $ps < .026$; right IT: $ps < .002$, after bivariate outlier exclusion, $ps < .020$). In sum, these findings suggest that the relations between calculation skills and response amplitudes were specific to the detection of digits.

Finally, there was weak to moderate evidence that individuals with higher calculation skills also had greater right lateralization in their mean response amplitudes for Digit Present [$r(30) = -.36$, $p = .042$, $BF_{10} = 1.58$; $r_{skipped}(29) = -.48$, $p = .006$, $BF_{10} = 8.03$]. There was evidence of a lack of a similar relation for Digit Absent [$r(30) = -.09$, $p = .621$, $BF_{01} = 4.05$] (**Figure 4-2(d)**). Critically, these correlation coefficients were significantly different regardless of outlier exclusion ($ps < .011$).

Taken together, although there was no hemispheric asymmetry in digit sensitivity, there was a strong right lateralization of IT activity when the conditions were considered separately. Moreover, the relations between calculation skills and right lateralization in digit sensitivity were specific to only the detection of digits.

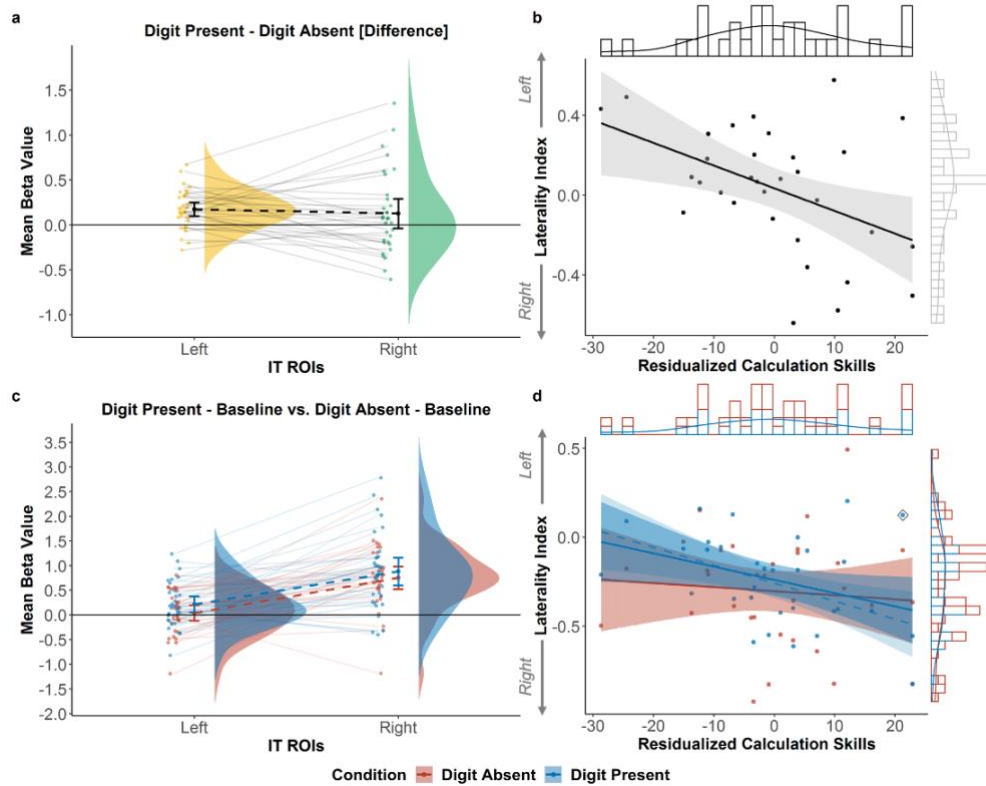


Figure 4-2. Hemispheric asymmetries of regional mean response amplitudes and their relation to calculation skills for digit detection (a, b) digit sensitivity ([Digit Present – Digit Absent] contrast), and (c, d) condition-wise activity ([Digit Absent – Fixation] and [Digit Present – Fixation]). Error bars and bands are 95% confidence intervals. Dashed regression lines exclude bivariate outliers enclosed in \diamond .

Letters

An identical set of analyses with the letter detection task (see Appendix B) revealed no lateralization for the regional mean letter sensitivity (i.e., [Letter Present – Letter Absent] contrast) ($BF_{01} = 5.00$), but right-lateralized mean condition-wise response amplitudes for both Letter Present and Letter Absent conditions ($BF_{S10} > 6689$), similar to that observed for digit detection task. There was no conclusive or robust evidence that higher calculation skills were associated with greater response amplitudes in both regions evoked by the mere presence of digits, regardless of whether a letter was present or not ($BF_{S0+} = 1.13$ to 2.73 ; after outlier

exclusion: $BF_{0+} = 11.22$ for Letter Present in left IT to $BF_{+0} = 73.22$ for Letter Absent in right IT). There was also evidence that calculation skills were not associated with lateralization for Letter Present ($r = -.14$, $BF_{01} = 3.39$), or the [Letter Present – Letter Absent] contrast ($r = .06$, $BF_{01} = 4.31$). The relation between calculation skills and lateralization of responses for Letter Absent (i.e., only digits) was inconclusive ($r = -.17$, $BF_{01} = 2.95$; after outlier exclusion: $r = -.24$, $BF_{01} = 2.03$). In sum, the relations between digit sensitivity and calculation skills were largely category-specific, although the right-lateralization of condition-wise response amplitudes appeared to be related to the detection task in general, regardless of whether one was looking for digits or letters.

Category Discriminability (Digits vs. Letters)

Category discriminability between digits and letters was evident in both the left ($M = 0.19$, $SD = 0.07$) [$t(30) = 14.58$, $d = 2.62$, $p < .001$, $BF_{+0} = 3.35 \times 10^{12}$] and right IT ($M = 0.29$, $SD = 0.10$) [$t(30) = 16.19$, $d = 2.91$, $p < .001$, $BF_{+0} = 4.90 \times 10^{13}$] (**Figure 4-3**). Moreover, category discriminability was higher in the right IT than in the left IT (difference: $M = -0.10$, $SD = 0.11$), $t(30) = -5.51$, $d_z = -0.99$, $p < .001$, $BF_{10} = 3646$.

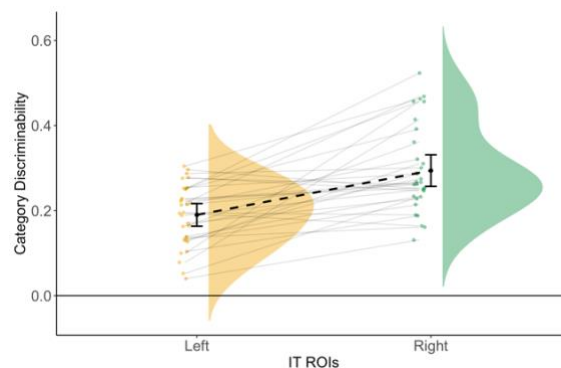


Figure 4-3. Category discriminability in the left and right IT ($N = 31$)

Error bars are 95% confidence intervals.

There was inconclusive evidence that greater category discriminability was associated with higher calculation skills in the left IT [$r(29) = .27, p = .070, BF_{+0} = 1.18$] and the right IT [$r(29) = .18, p = .166, BF_{0+} = 1.73; r_{skipped}(28) = .28, p = .069, BF_{+0} = 1.19$] (**Figure 4-4**).

There was also no relation between the degree of lateralization of category discriminability and calculation skills, $r(29) = .01, p = .939, BF_{01} = 4.47; r_{skipped}(28) = -.07, p = .697, BF_{01} = 4.10$].

Taken together, not only was category discriminability robust in both ROIs, it was greater in the right IT than in the left IT, contrary to the pattern of results for regional mean digit sensitivity. Whether greater category discriminability was associated with higher calculation skills was inconclusive. However, lateralization did not matter for the association.

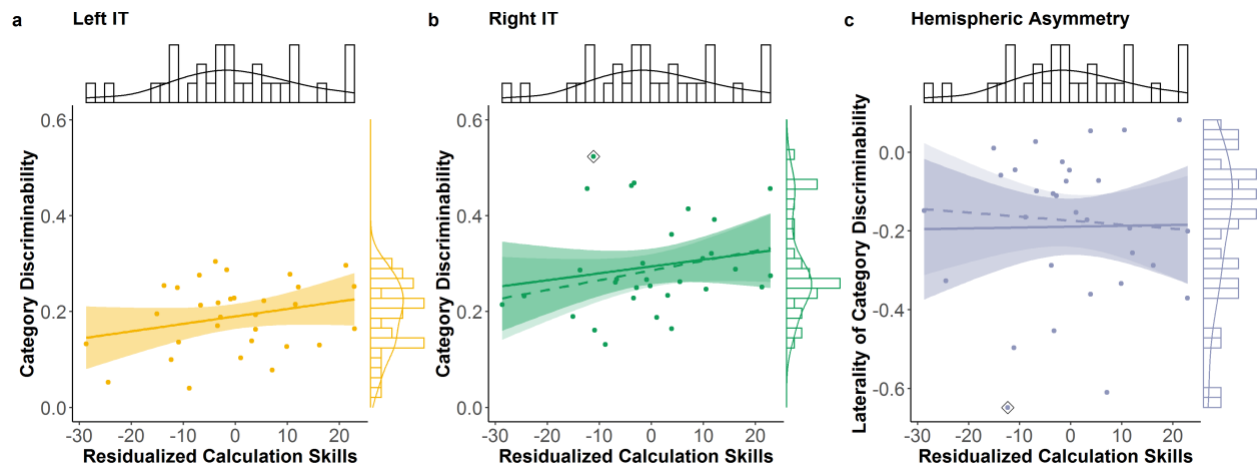


Figure 4-4. Relation between calculation skills and category discriminability (a) in left IT, (b) in right IT, and (c) its lateralization (negative: left lateralization, positive: right lateralization) ($N = 31$). Error bands are 95% confidence intervals. Dashed regression lines exclude bivariate outliers enclosed in \diamond .

Exemplar Discriminability

Digits

There was inconclusive evidence of digit discriminability in the left IT ($M = 0.02$, $SD = 0.12$) [$t(30) = 0.99$, $d = 0.18$, $p = .166$, $BF_{0+} = 2.02$], and moderate evidence of a lack of digit discriminability in right IT ($M = 0.01$, $SD = 0.11$) [$t(30) = 0.60$, $d = 0.11$, $p = .275$, $BF_{0+} = 3.07$] (**Figure 4-5**). However, there was no hemispheric asymmetry in digit discriminability (difference: $M = 0.01$, $SD = 0.17$), $t(30) = 0.32$, $d_z = 0.06$, $p = .751$, $BF_{01} = 4.98$.

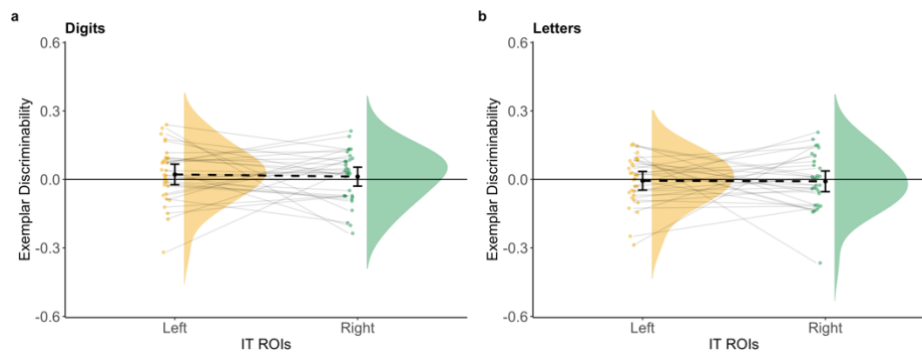


Figure 4-5. Exemplar discriminability in left and right IT (a) digits ($N = 31$) and (b) letters ($N = 30$). Error bars are 95% confidence intervals.

Greater digit discriminability was, however, associated with higher calculation skills in the left IT [$r(29) = .38$, $p = .018$, $BF_{+0} = 3.56$], but evidence of a similar relation in the right IT was inconclusive [$r(29) = .22$, $p = .115$, $BF_{0+} = 1.28$] (**Figure 4-6**). These correlation coefficients did not differ significantly (all $ps > .500$). There was also no relation between the degree of lateralization of digit discriminability and calculation skills, $r(29) = .13$, $p = .489$, $BF_{01} = 3.56$.

Even though we regressed out letter-word identification skills from calculation skills to

account for variance related to general symbol decoding, we wanted to confirm that digit discriminability was not also positively correlated with letter-word identification skills. We found no positive association between digit discriminability and letter-word identification skills (residualized for calculation skills) in the left IT [$r(29) = -.04$, $p = .595$, $BF_{0+} = 5.34$], but the evidence in the right IT was again inconclusive [$r(29) = .10$, $p = .297$, $BF_{0+} = 2.80$]. Digit discriminability had a significantly stronger positive association with calculation skills than with letter-word identification skills in the left IT [$ps = .043$ for two tests and $ps = .056 - .060$ for eight other tests], but not in the right IT [$ps > .319$].

In sum, on average across participants, the right IT did not distinguish digit exemplars, but there was inconclusive evidence whether the left IT distinguished digit exemplars. However, greater digit discriminability in the left IT ROI – and possibly the right IT ROI too – were associated with higher calculations skills, and laterality did not moderate the brain-behavior relation.

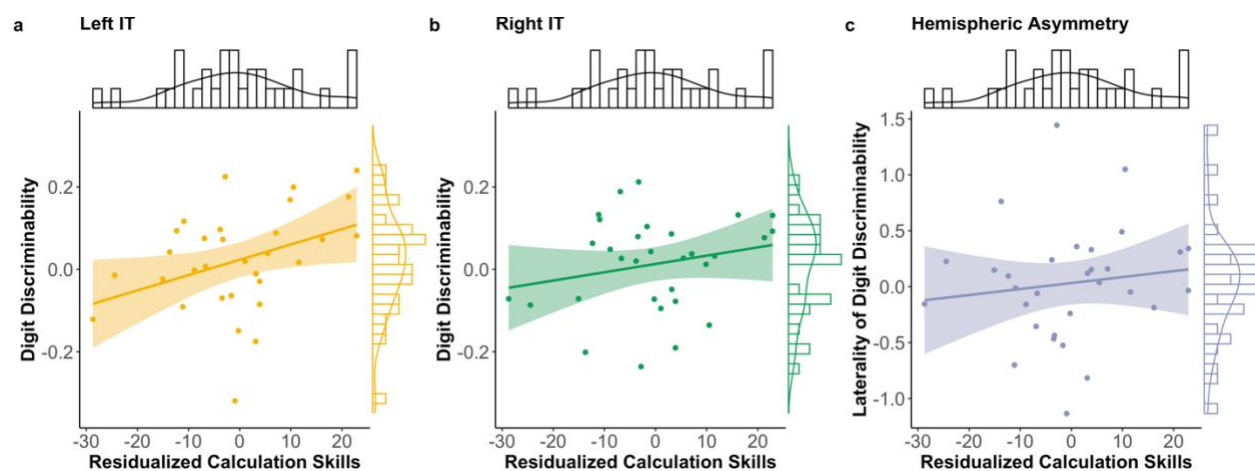


Figure 4-6. Relation between calculation skills and digit discriminability (a) in the left IT ROI, (b) in the right IT ROI, and (c) its lateralization (negative: left lateralization, positive: right lateralization) ($N = 31$). Error bands are 95% confidence intervals.

Letters

To assess category specificity, we performed an identical set of analyses pertaining to exemplar discriminability on the Letters-RDMs. There was no letter discriminability in either ROI ($BF_{S0+} > 6.34$) (**Figure 4-5(b)**), and calculation skills were not positively associated with the degree of letter discriminability in either ROI ($r_s = -.22$ to $-.003$, $BF_{S0+} = 4.46 - 8.96$) and their laterality ($r = -.13$, $BF_{01} = 3.47$) (see Appendix B). In sum, the associations between exemplar discriminability and calculation skills were specific to digits.

Hemispheric Asymmetry of Representational Geometries

On average, there was a small, but significant positive correlation between the representational geometries of digits (i.e., Digits-RDMs) in the left and right IT (mean $\rho_z = .11$, $SD = .20$), $t(31) = 3.14$, $d = 0.56$, $p = .002$, $BF_{+0} = 20.51$. There was inconclusive evidence that lower between-hemisphere similarity (i.e., greater asymmetry) in the representational geometries of digits was associated with higher calculation skills, $r(30) = -.17$, $p = .172$, $BF_{0-} = 1.80$.

A similar pattern of results was observed for the representational geometries of letters (see Appendix B).

Representational Content

Digits

Although there was no evidence of digit discriminability in both IT ROIs, there could be a discernable organization among the exemplar representations regardless of how similar each exemplar pair was on average. Hence, we explored whether the representational geometries could be described by hypothesized models of phonological, numerical, and shape similarity.

Left IT. The left IT Digits-RDMs were not similar to the RDMs of the Phonological model (Mean $\rho_z = -.01$, $SD = .14$) [$t(31) = -0.59$, $d = -0.10$, $p = .719$, $BF_{0+} = 7.84$], Ratio model

(Mean $\rho_z = -.01$, $SD = .22$) [$t(31) = -0.33$, $d = -0.06$, $p = .629$, $BF_{0+} = 6.70$], and Shape model (Mean $\rho_z = -.05$, $SD = .18$) [$t(31) = -1.38$, $d = -0.24$, $p = .910$, $BF_{0+} = 11.55$] (**Figure 4-7**). There was, however, inconclusive evidence that the left IT Digits-RDMs were similar to the Frequency model RDM (Mean $\rho_z = .09$, $SD = .29$), $t(31) = 1.69$, $d = 0.30$, $p = .051$, $BF_{+0} = 1.27$.

In terms of pairwise model comparisons, there was no evidence of within-participant differences between any pair of neural RDM-model RDM similarities, all $ps > .052$, $BF_{S10} < 1.13$ (**Table B-5**).

Right IT. The right IT Digits-RDMs were not similar to the RDMs of the Phonological model (Mean $\rho_z = -.03$, $SD = .15$) [$t(31) = -1.09$, $d = -0.19$, $p = .858$, $BF_{0+} = 10.19$], Frequency model (Mean $\rho_z = -.06$, $SD = .20$) [$t(31) = -1.63$, $d = -0.29$, $p = .943$, $BF_{0+} = 12.74$], and Shape model (Mean $\rho_z = -.04$, $SD = .14$) [$t(31) = -1.63$, $d = -0.29$, $p = .943$, $BF_{0+} = 12.76$] (**Figure 4-7**). There was, however, inconclusive evidence that the right IT Digits-RDMs were similar to the Ratio model RDM (Mean $\rho_z = .05$, $SD = .19$), $t(31) = 1.42$, $d = 0.25$, $p = .084$, $BF_{+0} = 1.18$.

In terms of pairwise model comparisons, there was some evidence that the Digits-RDMs were more similar to the Ratio model RDM than to the Shape model RDM [$t(31) = 2.61$, $d = 0.46$, $p = .014$, FDR-corrected $p = .084$, $BF_{10} = 3.36$], and no evidence of within-participant differences between any other pair of neural RDM-model RDM similarities [all $ps > .076$, $BF_{S10} < 0.84$] (**Table B-6**).

In sum, there was evidence that the models could not adequately describe the representational geometries in both ROIs. However, there was inconclusive evidence for the Frequency model in the left IT and Ratio model in the right IT. Replicating Yeo and colleagues (2020; Chapter 3), we found evidence of an absence of shape similarity in both regions.

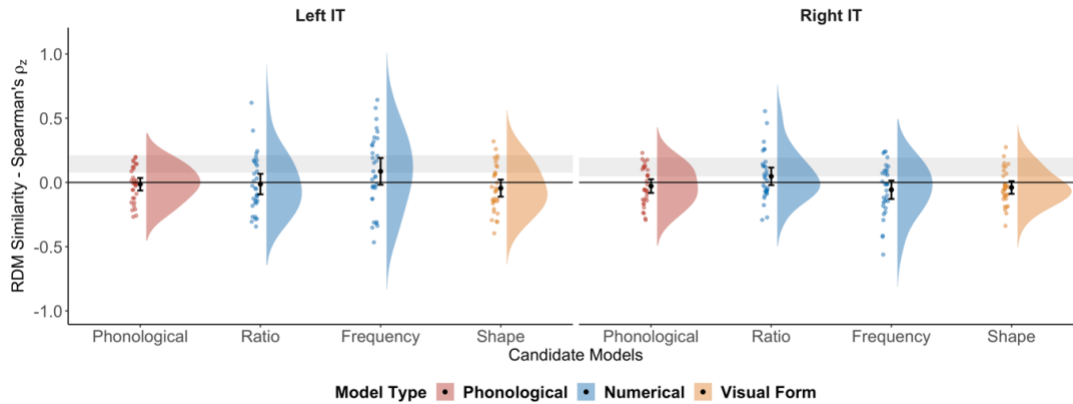


Figure 4-7. Similarity between the model RDMs and Digits-RDMs of IT ROIs ($N = 32$) Error bars are 95% confidence intervals. Grey bars indicate the estimated upper and lower bounds of the expected similarity achievable by the unknown true model given the degree of between-participant variability.

Left IT vs. Right IT. Neural RDM-model RDM similarities were not different between the left and right IT for the Phonological model [difference: $M = 0.01$, $SD = 0.16$, $t(31) = 0.50$, $d_z = 0.09$, $p = .624$, $BF_{01} = 4.73$] and Shape model [difference: $M = -0.005$, $SD = 0.22$, $t(31) = -0.13$, $d_z = -0.02$, $p = .900$, $BF_{01} = 5.26$] (**Figure 4-8**). There was inconclusive evidence that the neural RDM-model RDM similarities were different between the left and right IT for the Ratio model [difference: $M = -0.06$, $SD = 0.30$, $t(31) = -1.15$, $d_z = -0.20$, $p = .259$, $BF_{01} = 2.90$], and for the Frequency model [difference: $M = 0.14$, $SD = 0.36$, $t(31) = 2.27$, $d_z = 0.40$, $p = .030$ (FDR-corrected $p = .120$), $BF_{10} = 1.75$]. In sum, there was no hemispheric asymmetry in the degree to which the Phonological and Shape models described the representational geometries, but the evidence of hemispheric asymmetry was inconclusive for the Ratio and Frequency models.

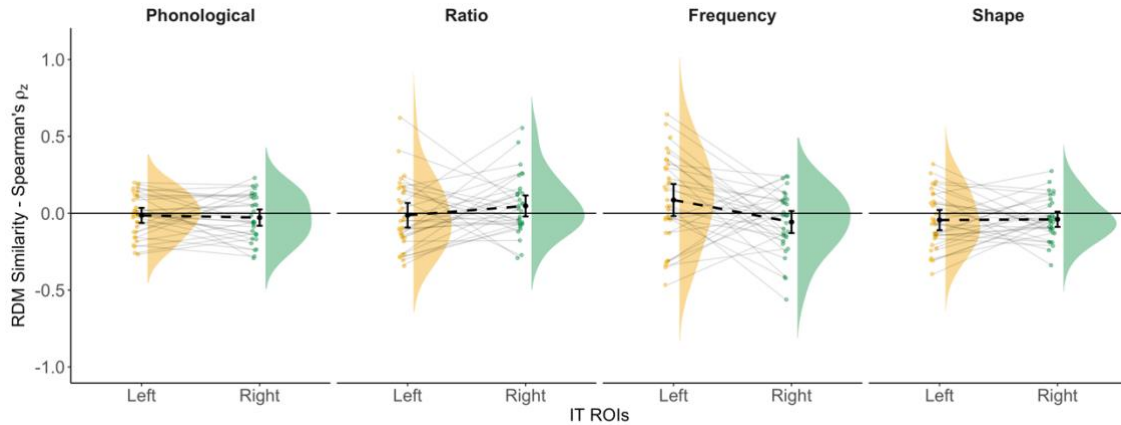


Figure 4-8. Hemispheric asymmetry in similarity between the model RDMs and neural Digits-RDMs ($N = 32$)
 Re-plotted from **Figure 4-7**. Error bars are 95% confidence intervals.

Alphanumeric Set

The Full-RDMs of the left and right IT were not similar to the Phonological and Shape model RDMs, $p_s > .410$, $BF_{S_{0+}} > 4.34$ (see Appendix B). Taken together, phonological and shape information were not represented in either ROI regardless of whether digits were considered alone or simultaneously with letters.

Discussion

The present study applied both univariate and multivariate region-of-interest analyses to Pollack and Price's (2019) data to probe the hemispheric asymmetry of various representational properties in the bilateral ITNAs during a digit detection task. We also probed the relation between those representational properties and calculation skills. Based on the findings of Pollack and Price (2019), we asked: Does the left ITNA show greater digit sensitivity than the right ITNA? Does the right ITNA relate to calculation skills in a way that the left ITNA does not? Here, we report that the left and right ITNAs did not actually differ in these properties, but they

did differ on other properties. Our findings thus suggest that the bilateral ITNAs are asymmetrically weighted in some functional responses and representations, but do not appear to be qualitatively different, at least during a digit detection task. However, until we have more convergent evidence from future studies, the field should err on the side of caution and not readily generalize findings about one ITNA to the other.

Right ITNA is More Involved in Category Discrimination Than Left ITNA

First, contrary to our prediction, we found that the right IT region showed no less digit sensitivity than the left IT region when we directly compared their sensitivities within individuals. This suggests that the successful localization of digit sensitivity in the left IT (statistically significant), but not in the right IT (statistically non-significant), reported in Pollack and Price (2019), are not statistically significantly different from each other (see Gelman & Stern, 2006; Nieuwenhuis et al., 2011; Rousselet & Pernet, 2012).

However, using multivoxel pattern analyses, we found that both IT regions showed significant category discriminability between the alphabet and numerals, and the degree of category discriminability was greater in the right IT than in the left IT. Interestingly, further probing of the univariate analyses for each condition relative to the fixation baseline (e.g., Digit Absent > Fixation) revealed that the right IT region was substantially more engaged during visual search regardless of both the target category (i.e., digit or letter detection) and the presence or absence of a target. Taken together, these findings suggest that there may be an inherent task-related hemispheric asymmetry, and that the right ITNA is no less involved, if not more, than the left ITNA in alphanumeric category discrimination. This right lateralization of alphanumeric category discrimination is consistent with the findings from Chapters 2 and 3 (Yeo et al., 2017, 2020). In Chapter 3 (Yeo et al., 2020), using similar representational similarity analyses, we

found that a meta-analytic identified right ITNA (Chapter 2; Yeo et al., 2017) discriminates digits from letters and novel characters that were passively viewed, but such evidence was absent in its left homologue. However, it was unclear if a true hemispheric asymmetry exists because a direct comparison between the left and right ITNAs was not made in that study. Moreover, applying transcranial magnetic stimulation to the right ITNA has been found to disrupt both letter and digit detection when participants were asked to categorize between alphanumeric characters and novel ones, suggesting a causal role of the right ITNA in alphanumeric categorization (Grotheer, Ambrus, et al., 2016). Nonetheless, because Grotheer, Ambrus, and colleagues (2016) did not also stimulate the left ITNA, it is unclear whether the left ITNA plays a qualitatively similar, but weaker causal role in alphanumeric categorization.

The results of the present study suggest that multivoxel pattern analyses do confer greater sensitivity than traditional univariate analyses, and they might be a powerful tool as a localization technique (Kriegeskorte et al., 2006; Kriegeskorte & Bandettini, 2007), especially for character categories that differ only by arbitrary representational purposes rather than inherent visual features. For instance, 5-7 year-old children who do not have a putative “Fusiform Face Area” based on univariate activation contrasts already show adult-like multivoxel pattern discrimination between faces and other categories in the most probable location of the “Fusiform Face Area” (M. A. Cohen et al., 2019). Hence, future studies that are unsuccessful in functionally localizing an ITNA using traditional univariate analyses may consider using a multivoxel pattern searchlight instead (e.g., see Carlos et al., 2019 for an example on localizing the "Visual Word Form Area").

Visual Search for Digits May Not Require Representations of Digit Identity in ITNAs

Even though there was evidence of category discriminability in both IT regions, we found

no conclusive evidence of digit discriminability. This suggests that category representations may be independent of digit identity representations (i.e., one does not need to identify which character it is in order to categorize it). Such a category distinction could simply be due to the spatial segregation of digit sensitivity (i.e., ITNAs) and letter sensitivity (i.e., “Visual Word Form Area” and “Letter Form Area”) (Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016; Pollack & Price, 2019). The dissociation between category identification and character identification is also consistent with existing behavioral evidence (McCloskey & Schubert, 2014; Taylor, 1978). McCloskey and Schubert (2014) showed that patient L.H.D., with alexia due to a left ventral lesion, was impaired in the ability to identify individual digits and letters, but was perfectly accurate in classifying digits and letters in mixed strings (e.g., ‘2VG5QS’). The authors concluded that “digit/letter category representations and character identity representations were computed separately but concurrently for all elements in the display, with the category representations providing the basis for present/absent judgements when the target and distractors differed in category” (McCloskey & Schubert, 2014, p. 458). Psychophysics evidence in neurotypical adults also suggest that identity and category are extracted in parallel (Taylor, 1978).

It is important to note that our results demonstrate an absence of evidence of digit discriminability in the left ITNA, and an evidence of an absence of digit discriminability in the right ITNA, but there was evidence of an absence of hemispheric asymmetry in digit discriminability. It could be that visual search tasks, at least in the version implemented here, are not robust in eliciting digit identity representations because task decisions may require only a basic level categorization (i.e., a digit or a letter) and not subordinate level categorization (i.e., which specific digit or letter). Nonetheless, it is possible that digit discriminability in the ITNAs

would be robustly evident in tasks in which digit identity is crucial. For instance, Wilkey and colleagues (2020) found above-chance decoding of the multivoxel response patterns evoked by digits in the left homologue of the meta-analytically identified right ITNA (Yeo et al., 2017; Chapter 2) across tasks involving single-digit identification (Is it a 2?) and comparison to a reference magnitude (Is it greater or less than ‘5’?)⁹. Moreover, we found that digit discriminability in the left IT region was associated with calculation skills, which suggests that the discriminability of digit representations in the ITNAs do have behavioral relevance. The numerically weaker digit discriminability in the right IT in the current study and the study by Wilkey and colleagues (2020) is compatible with the hypothesis that the magnitude representations downstream in the right parietal cortex (to which the right ITNA is connected to) are more approximate, or less discrete, in nature compared to its left counterpart (Chassy & Grodd, 2012; Kimura, 1966; Kosslyn et al., 1989; Piazza et al., 2006, 2007). Taken together, we speculate that the ITNAs may differ in how strongly they represent digit identity depending on the task context: When a task does not require discrimination between digits (e.g., category detection), digit identity is weakly represented in the ITNAs; when a task requires a digit to be discriminated from another digit (e.g., digit identification or magnitude comparison), digit identity is more strongly represented in the ITNAs, possibly by top-down modulation. The extent to which the right ITNA discriminates digits may also be slightly weaker than that in the left ITNA.

Calculation Skills Are Associated with Hemispheric Asymmetry of Some Functional and Representational Properties

⁹ Above-chance decoding was found in the left ITNA ($M = 27.8\%$, chance level = 25%), but not in the right ITNA ($M = 26.6\%$). However, paired samples t -test revealed inconclusive evidence of whether the decoding accuracies were significantly different between hemispheres, $t(38) = 1.50$, $p = .143$, $BF_{01} = 2.05$.

Using a region-of-interest analysis, we clarified that the mean digit sensitivity in the left IT region was also positively correlated with calculation skills, suggesting that the left IT region was not qualitatively different from the right IT region, despite indications to that effect in the results reported by Pollack and Price (2019). Importantly, consistent with our prediction, we found that higher calculation skills were also associated with greater right lateralization in digit sensitivity. Although our findings may not be consistent with Amalric and Dehaene's (2016) findings that the left (but not the right) ITNA's response to numerals was modulated by professional mathematical expertise, it may be possible that a right lateralization in digit sensitivity is more robust within non-mathematicians, which can be observed in their data (Figure 8E).

Even though we found inconclusive evidence of digit discriminability in the left IT region (on average across participants), we found that greater digit discriminability was associated with higher calculation skills. Hence, here we show that individual differences analyses can provide information that would otherwise be obscured by group averages (cf. Fisher et al., 2018). This relation cannot be entirely explained by general symbol decoding because we regressed out letter-word identification skills from calculation skills. We also did not find a positive association between calculation skills and letter discriminability in both ROIs. Moreover, we found some evidence of stronger associations between digit discriminability and residualized calculation skills than residualized letter-word identification skills in the left IT. Taken together, there is some degree of specificity between digit discriminability and calculation skills that is worth replicating in future research with a larger sample. Future work should also consider controlling for general object recognition ability (that is independent of intelligence) for which reliable individual differences have been found that generalize across familiar and novel

object categories (Gauthier, 2018; Richler et al., 2017, 2019).

Although evidence for a similar relation between digit discriminability and calculation skills in the right IT region was inconclusive, we found that its correlational strength did not differ statistically from that in the left IT region, and that greater hemispheric asymmetry in digit discriminability was also not associated with higher calculation skills. Hence, it is possible that digit discriminability in the right IT region is as important for calculation skills as the left IT. Our findings are in contrast with a recent study by Wilkey and colleagues (2020) that found weak to moderate evidence of a null relation between decoding accuracy of multivoxel pattern classification of digit representations and calculation skills. One explanation is that the difference in tasks may modulate the degree of inter-individual variability in the discriminability of digit-specific representations. The digit detection task used here did not require discrimination between digits (i.e., whether the digit was a 2 or 3 did not matter), whereas Wilkey and colleagues (2020) used an identification task and a magnitude comparison task, for which discrimination between digits was necessary. It is possible that the digit detection task evoked spontaneous digit-specific representations with substantial inter-individual variability in the degree of discriminability. In contrast, in identification and comparison tasks, such inter-individual variability in the degree of discriminability may be attenuated when the digit-specific representations were amplified.

Finally, given the flexibility in the recruitment of either hemispheric number identification system depending on task contexts (L. Cohen & Dehaene, 1995, 1996, 2000), it is important to note that these findings may be specific to visual search tasks, and may not apply to other numerical tasks. Hence, more research would be needed to replicate and extend the current findings.

No Conclusive Evidence of Hemispheric Asymmetry in Representational Geometries of Digits in Both ITNAs

Based on the triple-code model, we predicted that intra-hemispheric information exchange among the verbal, magnitude, and visual codes would outweigh the inter-hemispheric information exchange, resulting in dissimilar representational geometries of digits between hemispheres. On one hand, we found that the representational geometries of digits were significantly correlated between the left and right IT regions. On the other hand, the correlation was small, so it is possible that the inter-hemispheric similarity was indeed attenuated by the intra-hemispheric interactions. Because it is statistically difficult to falsify the hypothesis simply by correlating the representational dissimilarity matrices, we further explored whether there was a discernable organization in representational geometry of digits in each IT region that could be described by models that characterize phonological, numerical, or shape similarity. None of the models were adequate in describing the regions' representational geometries of digits.

The null findings of the neural-and-model comparisons can possibly be explained by the lack of exemplar discriminability, which could result in noisy representational geometries without any meaningful rank order for the neural-and-model comparisons. As argued above, the lack of exemplar discriminability could be an artifact of the visual search paradigm rather than an intrinsic property of the ITNAs. Given these caveats, we refrain from making any inferences about what the ITNAs represent.

Nonetheless, there was conclusive evidence that the representational geometry of digits could not be described by visual form, which replicates the finding in Chapter 3 (Yeo et al., 2020). Hence, the prevalent label “Number Form Area” should be avoided since, to the best of our knowledge, there is currently no direct evidence that numeral-preferring IT regions are sensitive

to visual form per se (i.e., the actual physical shape of the digit), nor that systematic differences in visual form between letters and digits even exist (Schubert, 2017).

In sum, the insensitivity of the current data to disambiguate the various models suggests that more research needs to be conducted to examine hemispheric asymmetry or lack thereof in the representational geometries of digits in the ITNAs.

Limitations

First, it is possible the presence of other characters from the non-target category could have led to noisy exemplar-level representations. Although a single-character categorization task would be ideal to minimize the influence of characters from the non-target category, the present study was meant as a case study to probe a very specific hemispheric asymmetry reported by Pollack and Price (2019), so we were necessarily constrained by their experimental design. Future studies should consider designs with a single-stimulus presentation as well as a task that requires explicit digit discrimination to better understand the representational geometries of digits in the ITNAs.

Second, the present study focused only on the ITNA rather than also on the regions subserving the verbal and magnitude codes. Lateralization has been shown to be a regional-level phenomenon (i.e, lateralization may manifest in some regions of a network, but not others; Pinel & Dehaene, 2010), so a focal analysis solely on the ITNAs is an appropriate first step for understanding the shared and distinct roles of the bilateral ITNAs. Besides, localization of regions underlying the verbal and magnitude codes is non-trivial without additional localizer tasks to isolate their respective representations. Given the hypothesized intra-hemispheric interaction between codes, future studies should also examine if hemispheric asymmetries in representational properties exist in the parietal regions subserving the magnitude code, and how

they relate to the asymmetries between the ITNAs.

Third, we used group-level ROIs because the right IT ROI was localized using a correlational approach at the group level. However, such group-level ROIs are less optimal and sensitive than subject-specific ROIs because they do not account for anatomical variability between individuals (Nieto-Castañón & Fedorenko, 2012). This may partly explain the difference in conclusions drawn here and from previous studies that utilize subject-specific ROIs (Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016) and found that the bilateral ITNAs do not have distinct functional profiles. Alternatively, it may be that task differences may underlie the extent to which the left and right ITNAs are functionally different. Future research free of such methodological constraints should consider using subject-specific ROIs instead and probe the functional and representational properties across different tasks that vary in their verbal and magnitude processing demands.

Conclusions

To explore how the bilateral ITNAs are functionally dissimilar, we probed whether hemispheric asymmetry exists in an array of functional and representational properties of the ITNAs during a visual search task. In general, the ITNAs appear differentially weighted between hemispheres in their functional responses and representations, but there is no strong evidence that they are qualitatively different. We found that the bilateral ITNAs did not differ in their sensitivity to digits, and that digit sensitivity in both ITNAs correlated positively with calculation skills. The differences between these findings and those originally reported by Pollack and Price (2019) suggest that within-individual comparisons are a necessary follow-up to infer about hemispheric asymmetries. Nonetheless, we did uncover hemispheric asymmetry in other properties, such as activity in alphanumeric categorization in general, as well as category

discriminability. We also found certain properties that were associated with calculation skills, such as right-lateralization of digit sensitivity, and digit discriminability in the left ITNA. In light of these asymmetries and their relevance to behavior, future studies should not readily generalize findings about one ITNA to another. Given the hypothesis of flexible, task-dependent engagement of the bilateral ITNAs, our results are likely specific to visual search. Other numerical tasks may uncover hemispheric asymmetries in a similar or different set of functional and representational properties of the bilateral ITNAs as we have found here. Hence, further investigation may be worthwhile to probe the individual and joint contributions of both hemispheres in processing numerals. Finally, to better understand the nature of hemispheric asymmetry of cognitive functions in general, our study highlights the need to supplement traditional univariate group-averaged analyses with within-participant comparisons, multivoxel pattern analyses, and individual differences analyses.

Authorship Contributions

Darren J. Yeo: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Courtney Pollack:** Methodology, Investigation, Formal analysis, Writing – review & editing. **Benjamin N. Conrad:** Conceptualization, Writing – review & editing. **Gavin R. Price:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

CHAPTER 5

GENERAL DISCUSSION

Regardless of the writing scripts our native languages adopt, we almost certainly also have to master the Arabic numeral system comprising the digits 0 – 9 to function successfully in modern societies. As the foundational years of formal education typically have separate numeracy and literacy classes, the learning of the Arabic numeral system and native language writing systems tends to be highly contextualized. The repeated and predictable use of Arabic numerals mostly in numeracy contexts may influence how the visual systems in our brains are organized to identify and distinguish Arabic numerals from other character categories (Hannagan et al., 2015; also see Gauthier, 2000). The collection of studies in this thesis provided findings consistent with this hypothesis in that the recognition of numerals is supported by an inferior temporal region that is spatially distinct from the word- and letter-preferring fusiform regions. As the Hindu-Arabic numerals were only introduced to the Western world as late as the 12th century, and adopted worldwide only several centuries later (Chrisomalis, 2010; Smith & Karpinski, 1911), it is unlikely that a specialized brain system for reading Arabic numerals evolved due to natural selection. Hence, the findings inform us how human culture interacts with neurobiology – particularly, how our brains assimilate or accommodate multiple culturally defined symbol sets learned for various representational purposes (e.g., words, numbers, or music) (Dehaene & Cohen, 2007). In this final chapter, I summarize the findings of the studies in this thesis, discuss the key limitations of the approaches taken here, and provide directions for further investigations of the bilateral numeral-preferring vOT nodes (so-called “Number Form Areas” (NFAs) and the

extended numeral recognition circuits.

Summary of Findings

Right Posterior Inferior Temporal Gyrus as the Most Probable Location of an NFA

In Chapter 2, we used a coordinate-based meta-analytic approach to provide evidence of a numeral-preferring neural population in the right pITG that is spatially convergent across samples and paradigms. Although recent evidence offered a qualitative sense of a reproducible location of an NFA across studies, whether it lies within the pITG or fusiform gyrus, spans both gyri, or has a different location in each hemisphere remains unclear. For example, Grotheer, Herrmann, and colleagues (2016) found numeral-preferring regions in both pITG and fusiform gyri bilaterally, but did not consider the latter regions as candidate NFAs despite prior evidence suggesting the fusiform gyri as likely candidates (Pinel et al., 1999, 2001). Hence, the meta-analysis contributes to the literature by providing the first quantitative evidence for the spatial reproducibility of a candidate NFA in the right pITG. Our finding also suggests that its spatial reproducibility is higher in the right hemisphere than in the left hemisphere. The absence of evidence of an NFA in the left pITG could be due to either spatial variability within and/or between samples, the reliability of its recruitment due to task demands, or both. The findings of group-level clusters in the left pITG in individual studies (Amalric & Dehaene, 2016; Grotheer, Herrmann, et al., 2016; Pollack & Price, 2019), however, indicate that there is some degree of spatial consistency. Hence, varying task demands likely underlie the unreliable recruitment of the left pITG region. The role of numeral-preferring fusiform regions observed in some studies remains unclear and their representational contents should be probed alongside the pITG regions when both sets of regions are localized within the same sample or individual.

Our findings also cast doubt on signal dropout as the sole factor underlying the

unsuccessful localization of an NFA in previous fMRI studies. First, some of the studies contributing to the meta-analytic convergence in the pITG did not employ advanced acquisition and post-acquisition analytic methods to mitigate signal loss. Second, a more liberal inclusion of contrasts from additional studies that ought to reveal its preferential engagement for numerals (i.e., less specific contrasts in favor of the numeral condition) did not support such a convergence. Hence, we speculated that task demands for both the numerical and non-numerical control stimuli, and the cognitive subtraction underlying the contrasts, could have played a role in the mixed findings of an NFA.

Following the publication of Chapter 2 (Yeo et al., 2017), other fMRI studies provided support for our conjectures by showing that the NFAs reliably fall posterior to the signal dropout zone (Grotheer et al., 2018; Merkle et al., 2019), and that localization of the NFAs is strongly modulated by what one is tasked to do with the Arabic numerals (Grotheer et al., 2018; Pollack & Price, 2019). Hence, the meta-analysis has already advanced the field by prompting a re-direction of resources from a mitigation of signal dropout in the ITG to understanding the recruitment of the NFAs under different task conditions. Beyond the theoretical and methodological insights contributed by the meta-analysis, the meta-analytically defined NFA also provides an a priori region-of-interest (ROI) for focal analyses in studies for which functional localization is unsuccessful (such as in Chapter 3; Yeo et al., 2020) or not feasible (e.g., lack of control stimuli for contrast analyses, or lack of task such as resting-state scans) (e.g., Abboud et al., 2015; Nemmi et al., 2018; Wilkey et al., 2020).

Besides the candidate right NFA, bilateral parietal and right frontal regions also showed convergent activation across studies and paradigms as part of the broader “numeral processing network”. As the convergence in each region could arise from different subsets of studies

included in the meta-analysis, we could not conclude that they are co-activated, much less functionally connected. However, they offer a set of candidate regions that are more involved in processing Arabic numerals than other familiar symbol sets. Resting-state functional connectivity analyses with the NFA and VWFA as seed regions showed that they are indeed functionally connected to different set of regions – NFA with parietal and frontal regions implicated in magnitude processing, and VWFA with left-lateralized regions implicated in language processing (Abboud et al., 2015; Nemmi et al., 2018).



So-called NFAs Are Numeral-Relevant, but Are Not Sensitive to Visual Form of Digits

Although it was originally thought that the “Number Form Area” is particularly sensitive to the shapes of digits – visual or not (Abboud et al., 2015; L. Cohen & Dehaene, 1991; see Hannagan et al., 2015, for a review; Shum et al., 2013), our findings from Chapters 3 and 4 suggest otherwise. In Chapter 3, using three datasets in which participants passively viewed digits, letters, and novel characters, we found that the right candidate NFA and its left mirrored homologue were not automatically sensitive to the global shape of digits. These suggest that the information readout from the NFAs by downstream regions is not digit shapes. However, the right candidate NFA still automatically distinguished digits from letters and novel characters, and even showed some degree of numeral preference, in which letters and novel characters were represented alike, but distinct from digits. We found parallel evidence of automatic categorical distinction and numeral preference also in the right parietal region that was meta-analytically identified. No other region of the meta-analytic network showed a similar pattern of results as the right candidate NFA. We interpreted these findings as suggestive that the NFAs may be part of a predictive pathway for stimuli that have potential numerical relevance for readout and feedback by downstream regions involved in quantity processing, particularly in the bilateral parietal

cortices (Abboud et al., 2015; Daitch et al., 2016). The evidence of automaticity in numeral processing is also consistent with the “process-map” hypothesis that the organization of category-selective regions in the vOT may result from each region being best suited for the task or type of processing associated with each object category (Gauthier, 2000). Although the left parietal region also distinguished digits from other character categories, evidence of a numeral preference was inconclusive. There was, however, conclusive evidence of a lack of automatic categorical distinction in the left mirrored homologue of the candidate NFA. As in Chapter 2, the null finding of a left NFA could be due to the possibility that the ROI did not overlap with the left NFAs in most participants across three studies. Another possibility is that the left and right NFAs may be distinct in their representations.

Critically, the nature of the task (i.e., fixation-color change-detection) allows us to rule out the possibility that differences in response patterns were entirely accounted for by one’s attention to a character’s shape, identity, or category. Hence, Chapter 3 provides the first direct test of visual form representations in the NFA. Evidence for a lack of shape representations in both NFAs was replicated in Chapter 4 even when participants had to attend to the digit shapes in a visual search task. Based on these new insights, we propose that this region ought not to be referred to as a “Number Form Area”, but as “Inferior Temporal Numeral Area” (ITNA) to specify its gross anatomical location and preferred stimulus category. I will use this nomenclature going forward in the rest of this chapter.

One possibility is that instead of digit shapes, the ITNAs may instead represent abstract identities of digits, which is similar to the role of the letter case-invariant VWFA (L. Cohen & Dehaene, 2004; Dehaene & Cohen, 2011). Just like abstract letter identities are thought to be detected from activated stored sensory forms of allographs (e.g., “E” and “e”, or “4” and “4”)

(McCloskey & Schubert, 2014), abstract digit identities can also be detected from the activated stored sensory forms used to represent the same number (e.g., “5”, “V”,  and  share an abstract identity) (Abboud et al., 2015; Grotheer et al., 2018). Critically, these representations of abstract character identities do not contain information about shape or phonology (Friedmann & Coltheart, 2018; Rothlein & Rapp, 2014), which may account for the lack of evidence of shape representations in Chapters 3 and 4, and phonological representations in Chapter 4. Future studies may consider testing whether dice and finger patterns recruit the ITNAs even in non-numerical tasks (e.g., passive viewing), and whether they do so in individuals who have far less experience with Arabic numerals than with dice and fingers as visual representations of numbers (e.g., preschoolers).

In general, our findings in Chapters 3 and 4 also revealed a contrast between univariate and multivariate approaches. In Chapter 3, we found some degree of automatic preference for numerals in the right ITNA during passive viewing that was evident when we examined the exemplar-level multivoxel response patterns. Such a categorical distinction was elusive when regional responses averaged across voxels and exemplars were examined instead (Merkley et al., 2019; G. R. Price & Ansari, 2011). In Chapter 4, we found that the right ITNA exhibited greater category discriminability between digits and letters than its left counterpart during digit detection even though they showed equivalent digit sensitivities based on their regional mean responses. It is not surprising or novel that multivariate pattern analyses may be more sensitive than univariate analyses (Kriegeskorte et al., 2006; Kriegeskorte & Bandettini, 2007), which can even lead to qualitatively different and complementary conclusions (e.g., M. A. Cohen et al., 2019; Jimura & Poldrack, 2012; McGugin et al., 2015). It is, however, worth considering in future studies to include both approaches whenever possible to increase the richness of inferences we can make

from the same data.

In sum, both Chapters 3 and 4 add to the literature evidence that the ITNAs do not represent visual form of Arabic digits per se even though they do distinguish digits from other character sets. This is important because it explains why the right ITNA is recruited by Roman numerals (via auditory soundscapes) in congenitally blind adults (Abboud et al., 2015), and the bilateral ITNAs are recruited by dice and finger patterns in neurotypical adults (Grotheer et al., 2018). More broadly, it is consistent with findings beyond the numerical domain that indicate we may not have a separate region for different symbol sets used for similar representational purposes. For instance, English-Hebrew and Chinese-English bilinguals show overlap in the VWFAs involved in scripts of both languages (Baker et al., 2007; A. C. N. Wong et al., 2009).

Hemispheric Asymmetry of ITNAs' Functional and Representational Properties, and Relevance to Behavior

Hemispheric asymmetry of the bilateral ITNAs has been observed in the literature and in the first two studies of this thesis. Specifically, it is not uncommon to observe unilateral recruitment of the left or right ITNA (or vOT more generally) in numeral processing. Using a case study approach, in Chapter 4, we re-analyzed the data reported by Pollack and Price (2019) to probe whether the bilateral ITNAs are functional different. We found that some functional responses and representations are asymmetrically weighted between hemispheres, but there was no strong evidence that they are qualitatively different. First, categorical distinction between digit and letter multivoxel response patterns was found in both ITNAs, but the right ITNA exhibited higher category discriminability than the left ITNA. Interestingly, a right lateralization in the regional mean response amplitudes was also observed for both digit detection and letter detection regardless of whether a target was present or absent. This suggests an inherent right

lateralization for alphanumeric categorization, or possibly categorization between numerals and other character categories more generally. These findings and those of Chapter 3 suggest that the right ITNA may be more consistently involved in distinguishing numerals and other character categories than its left counterpart, and may likely underlie the right-hemispheric bias in numeral processing that is not uncommon in the literature.

Chapter 4 also adds to the sparse evidence base of the relation between individual differences in the functional and representational properties of the ITNAs and mathematical competencies. Although Chapter 4 is based on the same data as Pollack and Price (2019), we explored a richer set of brain-behavior relations including laterality indices, which no study of the ITNA has considered. First, in addition to the original finding that digit sensitivity of the right ITNA was positively correlated with calculation skills, we found a significant, but slightly weaker relation in the left ITNA. Critically, however, the degree of right lateralization in digit sensitivity was positively correlated with calculation skills. Second, digit discriminability was also positively correlated with calculation skills in the left ITNA, but the evidence was inconclusive in the right ITNA. Nonetheless, degree of laterality in digit discriminability was not associated with calculation skills, suggesting that laterality may not matter. Taken together, we provide evidence that the functional and representational properties of the ITNAs, as well as their laterality, are relevant for behavior such as calculation skills.

One explanation for the brain-behavior associations is that the ITNAs possess innate functional and structural properties (e.g., fovea processing and connectivity to magnitude processing circuits) that support the learning of numerals and performance in manipulating them (Abboud et al., 2015; Grotheer et al., 2019; Hannagan et al., 2015; Nemmi et al., 2018). For example, Nemmi and colleagues (2018) found that higher resting-state connectivity of the right

ITNA with left intraparietal sulcus and right dorsolateral prefrontal cortex was associated with higher math competencies. Another explanation is that the ITNAs develop ontogenetically as a result of our experience with numerals as a distinct set of symbols from other symbol sets (e.g., the Roman alphabet), and our competence in manipulating them in stereotypical and constrained contexts. Nonetheless, how these explanations relate to lateralization requires further research. An investigation of the similarities and differences in structural connectivity between the left and right ITNAs will be highly informative in constraining hypotheses about their functions and lateralization. As the investigation of brain-behavior relations is still limited and the findings have been mixed (Amalric & Dehaene, 2016; Nemmi et al., 2018; Pollack & Price, 2019; Wilkey et al., 2020), more research needs to be conducted to better characterize the specific functional and representational properties of the ITNA that are associated with individual differences in experience and competence with numerals.

In sum, our findings in Chapter 4 contribute to the field by highlighting the need to consider the individual and joint contributions of the left and right ITNAs to numeral processing and mathematical skills. Although we did not find evidence of a qualitative difference between the ITNAs, the quantitative asymmetries in some functional responses and representations cannot be ignored. Hence, future research should avoid generalizing findings about one ITNA to another. Our findings also speak to an ongoing debate about whether object representations are in fact distributed across both hemispheres but with connectivity-constrained asymmetric weightings that result in a graded (as opposed to categorical) lateralization for some categories (e.g., left-lateralized VWFA) (Behrmann & Plaut, 2013, 2015, 2020). Numeral representations do appear distributed across both hemispheres, but also show graded lateralization that are relevant for behavior (e.g., right lateralized digit sensitivity).

Taken together, the work in this thesis contributes to the literature by characterizing the so-called “Number Form Areas” in terms of their spatial reproducibility, their functional and representational properties, hemispheric asymmetry in those properties, as well as how those properties and their asymmetry relate to mathematical skills.

Future Directions

Although the ITNAs are involved in numeral processing, likely by virtue of their connections to parietal regions involved in magnitude processing, we are still far from understanding whether the ITNAs are truly specialized for the reading requirements specific to the Arabic numeral system, and what properties of numeral reading the ITNAs are tuned to as a result of our experience in reading and manipulating numerals: Are they mere abstract digit identity encoders? Or are they a visuospatial buffer of ordered digit identities as originally conceived by Cohen and Dehaene (1991)? Or do they support a collection of more specialized processes unique to numeral reading, especially of multi-digit numerals? Although the localization of the ITNAs has been demonstrated with both single- and multi-digit numerals (Amalric & Dehaene, 2016; Grotheer et al., 2018; Grotheer, Herrmann, et al., 2016; Shum et al., 2013), no study has directly compared whether the same region is involved regardless of string length, or whether there are distinct regions that subserve single- or multi-digit numeral processing. The latter is not impossible because single letters and letter strings recruit different regions of the left fusiform gyrus (James et al., 2005). However, unlike letters, single digits are not categorically distinct from digit strings (e.g., in their lexical status). The ITNAs are also responsive to single Roman numerals, dice, and finger patterns that do not conform to the base-10 place-value system (Abboud et al., 2015; Grotheer et al., 2018). Moreover, there is growing evidence that the neuronal populations surrounding or adjacent to the ITNAs are engaged by

meta-modal mathematical processing, such as mathematical statements presented aurally (Amalric & Dehaene, 2016; Baek et al., 2018; Bugden et al., 2019; Daitch et al., 2016; Grotheer et al., 2018; Hermes et al., 2017; Pinheiro-Chagas et al., 2018). To the extent that multi-digit numerals involve additive and multiplicative reasoning (e.g., $254 = 2 \times 100 + 5 \times 10 + 4$), multi-digit numeral reading likely engage the ITNAs and/or the neuronal populations adjacent to them. An immediate next step in this research program is to understand whether and how the ITNAs represent multi-digit numerals, and what specific stimulus characteristics they are tuned to.

There is an emerging consensus that processing of digits and letters diverge at the level of character strings, and that the processing of letter strings and digit strings are separate and qualitatively different, even between analogous sub-processes, possibly because they have distinct syntactic structures that require different parsing mechanisms (Dotan & Friedmann, 2019; Grainger & Hannagan, 2014). Although not necessary, these distinct mechanisms may require dedicated neural resources, and by extension, neural segregation for processing digit and letter strings (Dotan & Friedmann, 2019). **Figure 5-1** shows a cognitive model for reading aloud numerals proposed by Dotan and Friedmann (2018) alongside a similar model of reading aloud words (Friedmann & Coltheart, 2018) (for a review, see Dotan & Friedmann, 2019). There are other models of reading that can better account for behavioral and neuroimaging data, such as connectionist models, in which word recognition and naming are supported by a convergence of distributed representations of orthography, phonology and semantic codes¹⁰ (Harm & Seidenberg, 2004; Seidenberg, 2005; Seidenberg & McClelland, 1989). I highlight the pair of

¹⁰ It is plausible that a connectionist model can also better account for how we read numerals aloud or for meaning. For example, “2021” can be read as “two, zero, two, one”, “two thousand and twenty-one” or “twenty, twenty-one” to invoke a nominal label, a large numerical value, and a position on a timeline, respectively. How we parse the string and map the phonological code to the orthographic code depends on the context and semantic.

models by Friedmann and colleagues mainly because they are highly detailed in describing the parallel and non-parallel processes of the orthographic codes for letter and digit strings. Thus, they provide a starting point for developing focal hypotheses about the types of information the ITNAs may be tuned to relative to the VWFA. In contrast, a connectionist model in which information is distributed across codes lacks the specificity of processes involved within each code that can be empirically tested. As the goal here is to understand orthographic processing of numerals more specifically, I will focus on the components of the visual analyzer of digit strings (outlined in red in **Figure 5-1(a)**).

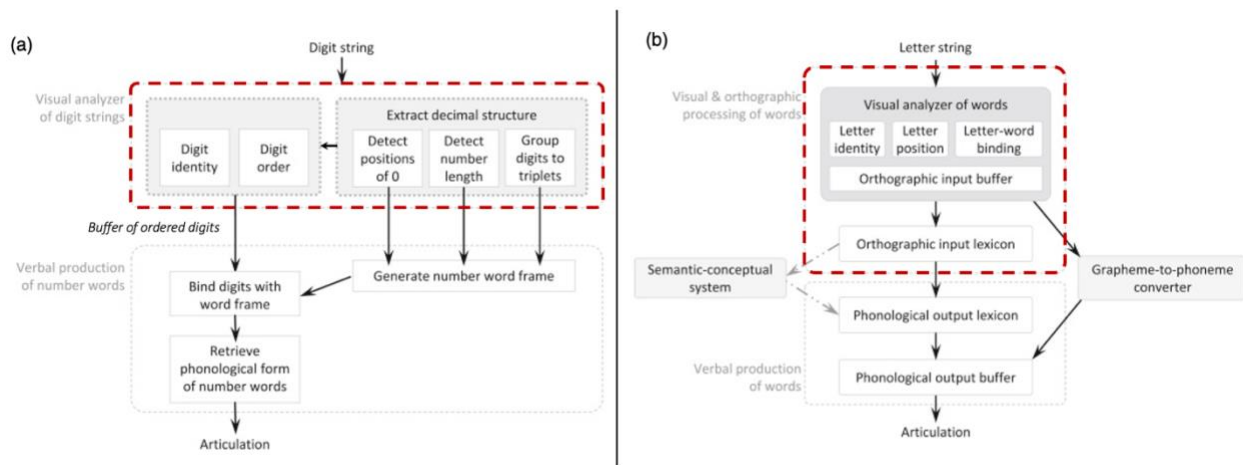


Figure 5-1. Cognitive models of reading aloud numerals and words
Cognitive model of reading (a) numerals (Dotan & Friedmann, 2018) and words (Friedmann & Coltheart, 2018). Reprinted and adapted from Dotan and Friedmann (2019) with permission from Elsevier.

In the visual analysis of digit strings in **Figure 5-1(a)**, the model postulates the existence of five distinct and dedicated processes, grouped into two sets. The first set includes three processes that serve to extract the decimal structure of a numeral. For instance, when presented

with the numeral 23706, the length of the numeral is detected (it has 5 digits), the numeral is parsed into triplets from right to left (23 and 706), and the positions of the digit 0 is detected (2nd from the right). Zero plays a special syntactic role as a placeholder in the base-10 system (e.g., zero is not semantically realized when there are no tens in 706). In most languages (with exceptions, e.g., Mandarin Chinese), zero is also not verbally realized in numeral-to-verbal transcoding (e.g., 23706 → “twenty-three thousand seven hundred and six”). The second set comprises the processes that encode the digit identities and the relative order of the digits (Friedmann et al., 2010). Finally, the output of the digit identity and digit order encoders are hypothesized to be represented in a buffer of order digit identities, which allows for the binding of the number word frame with the digits, at least for verbal number word production (Dotan & Friedmann, 2018). Dotan and Friedmann (2019) argue that these processes are distinct from those of the letter-string visual analyzer in **Figure 5-1(b)**, even between analogous processes such as those coding for identity and order. What seems truly unique about the digit-string visual analyzer that distinguishes it from the letter-string visual analyzer are the processes involved in extracting the numeral’s decimal structure. It would be worthwhile to investigate the sensitivity of the ITNAs to the decimal structure by systematically manipulating the different components.

Given that we now know the probable locations of the ITNAs, another line of investigation that is ripe for exploration concerns the structural and functional connections to and from the ITNAs. Behrmann & Plaut (2013) argue that functional specialization is an emergent property of the interactions between multiple regions within and between hemispheres via their structural and functional connections, rather than an intrinsic property of the isolated representations and computations of individual vOT regions. Hence, a better understanding of how we process numerals, especially in different contexts, will require us to go beyond analyses

of the ITNAs in isolation. A rich description of the regions that the ITNAs are connected to intrinsically or as dictated by task contexts will help constrain inferences that can be made about types of representations and computations in the extended neural circuits within and across hemispheres. Future work should therefore examine individual differences in structural and functional connectivity of the left and right ITNAs within and between hemispheres, and compare the functional responses and representations between heterotopic and homotopic regions in the circuits involved (Behrmann & Plaut, 2020). Developmental research will also be necessary to disentangle the contributions of genetically and experience-driven connectivity to lateralization effects in numeral processing.

A final line of possible investigation concerns the extent to which ITNAs that already exist to support Arabic numeral reading will support the learning of an alternative symbol set to represent numbers. The process-map account of category selectivity in the vOT suggests that category-selective regions arise because they support the processing goals required for specific categories (Gauthier, 2000). This account thus predicts that learning of multiple scripts for the same computational or representational purpose involves assimilation in the category-selective vOT nodes. For example, individual-subject analyses in bilinguals suggest that reading words in two different scripts (even between alphabetic and non-alphabetic) involve overlapping VWFAs (Baker et al., 2007; A. C. N. Wong et al., 2009). Training adults to read a novel alphabetic script comprised of images of houses also led to overlapping VWFAs of English and house-font orthographies, as well as learning-related increase in activity in pre-existing English-VWFA, which correlated with reading speed of the house-font (L. Martin et al., 2019). Finally, in Chinese-English bilinguals, greater proficiency in reading English was associated with higher multivoxel pattern similarity between English and Chinese words in the left mid-fusiform gyrus,

and a reverse relation was found in the right mid-fusiform gyrus that is less implicated in linguistic processing (Qu et al., 2019). Hence, assimilation within the key vOT nodes for reading different scripts for the same representational purpose can be evident in terms of spatial, functional, and representational properties. In the numerical domain, attempts have been made to study the neural changes in individuals who were trained to read multi-character Roman numerals (Masataka et al., 2007) and mapping abstract geometric patterns to approximate quantities (Lyons & Ansari, 2009). In general, these studies found that the new symbol set recruited regions that are commonly implicated in number processing, including vOT regions that span the ITG and fusiform gyri. However, because none of the studies included Arabic numeral processing for comparison, the degree of assimilation in the ITNAs and their extended networks, and how it relates to numeral reading competence remains unclear.

To sum up, some outstanding questions that can be pursued in future are to better understand the orthographic properties of multi-digit numerals the ITNAs are tuned to, the structural and functional connections with the ITNAs, the extent to which visually and/or syntactically distinct symbol sets for representing numbers tap into the same neural system, and how all of these relate to individual differences in numeral reading fluency and other symbolic mathematical skills.

Final Remarks

Arabic digits 0 – 9 are highly flexible tools that not only represent quantities (cardinal numerals), but also the positions of ordered sequences (ordinal numerals), and as identity labels (nominal numerals) across varied contexts within our environments. As such, reading Arabic numerals is a fundamental and necessary activity that we engage in most of our waking lives. The current thesis addresses how our brains facilitate fluent reading of Arabic numerals. This

work contributes to the numerical cognition and object recognition literatures with a programmatic characterization of the spatial, functional, and representational properties of the “Inferior Temporal Numeral Areas” that are hypothesized as the key vOT nodes of neural networks supporting numeral processing. More broadly, the findings provide insights on how cultural inventions shape our brains, and lay the groundwork for understanding the behavioral consequences when a brain struggles to incorporate the learning of arbitrary symbol sets within its existing neuronal architecture.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

Supplementary Methods

Comparison of Experimental and Neuroimaging Acquisition Parameters

Table A-1. Notable experimental and neuroimaging acquisition comparisons among datasets

Parameters	Dataset 1	Dataset 2	Dataset 3
<i>Experiment parameters</i>			
Run duration	16 min	9 min	11 min
Stimulus presentation duration	50 or 500 ms	500 ms	500 ms
Number of categories	8 (Ds, Ls, Scrambled Ds, Scrambled Ls for each presentation duration)	4 (Digits (Ds), Letters (Ls), Scrambled Ds, Scrambled Ls)	6 (Ds, Ls, Scrambled Ds, Scrambled Ls, Mirrored Ds, Mirrored Ls)
Number of exemplars per category	9	9	9
Number of trials per exemplar per run	2	2	2
Number of runs	2	1	4
Number of target trials/run	6	8	6
Inter-trial interval	4, 6 or 8 s ($M = 6$ s)	4, 5, 6, 7, or 8 s ($M = 6$ s)	1, 2 or 3 s ($M = 2$ s)
<i>fMRI acquisition parameters</i>			
MRI scanner model	3T Siemens Tim Trio	3T Phillips Achieva	3T Siemens Prisma Fit
Head coil type	32-channel	8-channel	32-channel
<i>Functional runs</i>			
Pulse sequence	T2-weighted echo-planar	Multislice 2D SENSE (factor 2) T2* gradient-echo, echo planar	T2*-weighted single-shot gradient-echo planar (Multiband acceleration factor: 4)
Echo time (TE)	30 ms	30 ms	30 ms
Time to repetition (TR)	2000 ms	2000 ms	1000 ms

Flip angle	90°	75°	40°
In-plane resolution	64 × 64 pixels	128 × 128 pixels	84 × 84 pixels
Field of view	192 mm	240 mm	208 mm
Slices per TR	38 axial	34 axial	48 axial
Voxel resolution	3 mm isotropic	1.875 × 1.875 × 3 mm ³	2.476 × 2.476 × 2.5 mm ³
Gap between slices	0 mm	1 mm	0 mm
Number of volumes/run	488	276	335

Behavioral Results

Table A-2. Number of errors of omission and commission across all runs in each dataset

Study	Omission errors ^a	Commission errors						
		Digits (500 ms)	Letters (500 ms)	Scrambled Digits (500 ms)	Scrambled Letters (500 ms)	Letters (50 ms)	Mirrored Digits (500 ms)	Mirrored Letters (500 ms)
Dataset 1 ($N = 19$)	5 (1 P with 3 errors; 2 Ps with 1 error each)	2 (2 Ps with 1 error each)	1	2 (2 Ps with 1 error each)	0	1	N.A.	N.A.
Dataset 2 ($N = 39$)	1	3 (2 Ps with 1 error each)	0	0	0	N.A.	N.A.	N.A.
Dataset 3 ($N = 40$)	8 (2 Ps with 2 errors each; 4 Ps with 1 error each)	7 (1 P with 1 error; 3 Ps with 2 errors each)	11 (1 P ^b with 6 errors; 1 P with 2 errors; 3 Ps with 1 error each)	4 (2 Ps with 2 errors each)	10 (1 P ^c with 4 errors; 1 P with 3 errors; 3 Ps with 1 error each)	N.A.	11 (1 P with 6 errors; 1 P with 3 errors; 2 Ps with 1 error each)	5 (1 P with 2 errors; 3 Ps with 1 error each)

Note. P = Participant.

These data included all runs, independent of the motion-related exclusion. Three participants from Dataset 3 were excluded from the final sample as no button responses were recorded from them.

^a Number of target trials per run: Dataset 1 – 6 trials; Dataset 2 – 8 trials; Dataset 3 – 6 trials.

^b This participant made 18 commission errors in one run (Digits: 2; Letters: 6; Scrambled Digits: 2; Scrambled Letters: 3; Mirrored Digits: 3; Mirrored Letters: 1) and 1 omission error. This run was excluded from analyses.

^c This participant made 18 commission errors in one run (Digits: 2; Letters: 2; Scrambled Digits: 2; Scrambled Letters: 4; Mirrored Digits: 6; Mirrored Letters: 2). This run was excluded from analyses.

Properties of Stimulus Sets

Table A-3. Visual properties of stimulus sets: Mean (SD)

	Visual angle (width × height)		Luminance		Perimetric complexity	
	Intact	Scrambled	Intact	Scrambled	Intact	Scrambled
Digits Set	4.63 (0.55) × 9.24 (0)	6.18 (1.00) × 8.25 (2.11)	8165.22 (1685.73)	7941.78 (1596.00)	82.39 (17.65)	86.53 (21.73)
Letters Set 1	5.67 (0.44) × 9.24 (0)	6.72 (1.77) × 10.83 (3.5)	10255.97 (1684.88)	10058.69 (1613.36)	88.70 (11.16)	90.61 (13.62)
Letters Set 2	5.67 (0.44) × 9.24 (0)	6.72 (1.77) × 10.83 (3.5)	10163.85 (1982.35)	9987.63 (1890.86)	88.80 (15.21)	91.49 (16.13)

Note:

Digits Set: Identical across all datasets

Letters Set 1: Datasets 1 and 2

Letters Set 2: Dataset 3

Table A-4 describes any substantial differences in luminance and perimetric complexity between the digits set and letters sets that may underlie any categorical differences. As we are not interested in making any inferences about the population parameters of the stimulus sets, frequentist hypothesis testing is inappropriate. Hence, we adopted Bayesian inferences (Bayesian Mann-Whitney in JASP 0.10.0 (JASP Team, 2019), 10,000 samples, with a Cauchy prior scale of 0.707) to quantify the degree of evidence for a difference between any two categories. By and large, albeit weak, there appears to be some evidence favoring a difference in luminance between the digits set and letters sets regardless of their intactness, and some evidence favoring a lack of difference in perimetric complexity between any two categories. These findings are consistent with those of Schubert (2017) for Arial font. Although the results suggested that luminance is not a huge issue ($BFs < 3$), it is still informative to assess the contribution of luminance as a confounding factor.

Table A-4. Pairwise comparisons of luminance (above diagonal) and perimetric complexity (below diagonal) of stimulus categories for Digits Set, Letters Set 1, and Letters Set 2 (bold)

	Intact digits	Intact letters	Scrambled digits	Scrambled letters
Intact digits	–	$W = 13,$ $BF_{10} = 2.44$ $W = 14,$ $BF_{10} = 1.88$	$W = 47,$ $BF_{10} = 0.48$	$W = 13,$ $BF_{10} = 2.55$ $W = 14,$ $BF_{10} = 2.03$
Intact letters	$W = 35,$ $BF_{10} = 0.44$ $W = 32,$ $BF_{10} = 0.50$	–	$W = 9,$ $BF_{10} = 3.21$ $W = 12,$ $BF_{10} = 2.37$	$W = 47,$ $BF_{10} = 0.48$ $W = 45,$ $BF_{10} = 0.44$
Scrambled digits	$W = 34,$ $BF_{10} = 0.47$	$W = 38,$ $BF_{10} = 0.43$ $W = 37,$ $BF_{10} = 0.43$	–	$W = 11,$ $BF_{10} = 3.01$ $W = 13,$ $BF_{10} = 2.08$
Scrambled letters	$W = 28,$ $BF_{10} = 0.68$ $W = 24,$ $BF_{10} = 0.86$	$W = 34,$ $BF_{10} = 0.46$ $W = 34,$ $BF_{10} = 0.46$	$W = 36,$ $BF_{10} = 0.43$ $W = 34,$ $BF_{10} = 0.46$	–

Note. W: Mann-Whitney U, two-sided.

Size and Temporal Signal-to-Noise Ratio of Participant-Specific Regions of Interest

The meta-analysis by Yeo and colleagues (2017) found convergent preferential activity to Arabic numerals than to other familiar symbols (e.g., Roman letters for English speakers or Chinese characters for Chinese speakers) across 20 studies in five regions (see **Figure 3-2(a)**): right inferior temporal gyrus (ITG, 55 3-mm isotropic voxels¹¹), left parietal lobule (PL; encompassing the inferior lobule and intraparietal sulcus, 114 3-mm isotropic voxels), right PL (encompassing both superior and inferior lobules, and the intraparietal sulcus, 362 3-mm isotropic voxels), right premotor cortex (PMC, encompassing both superior frontal gyrus and anterior cingulate cortex, 102 3-mm isotropic voxels), and right inferior frontal gyrus (IFG, 50 3-mm isotropic voxels). **Figure A-2** shows the group means and individual differences in the size of each region of interest (ROI) relative to the meta-analytic cluster in percentage, and the temporal signal-to-noise ratio (tSNR) per ROI. Apparent in **Figure A-2**, a participant in Dataset 3 had only 48.1% voxels retained in the left ITG, and the same participant had a tSNR of 44.7 in that region. We replicated the analyses pertaining to the left ITG cluster with the exclusion of this participant, and the findings were qualitatively similar to that of the full sample.

Of particular interest here is the tSNR in the ITG ROIs that are known to be susceptible to signal dropout. We thus provide a statistical comparison of the tSNR in these ROIs across the datasets as a potential factor that may explain any differences in dataset-specific findings. A one-way analysis of variance of the tSNRs across datasets revealed that there were statistically significant differences in the mean tSNRs in the left ITG ROI across the datasets [$F(2, 85) =$

¹¹ The volume of the original meta-analytic cluster was computed based on a map with 2-mm isotropic resolution. For the current study, the clusters were imported to BrainVoyager for analysis, which required first resampling them to 1-mm isotropic voxels and then to 3-mm isotropic voxels. The resampling procedure thus resulted in more voxels than one would expect based on the original cluster volumes.

5.27, $p = .007$, $\eta_p^2 = 0.11$, $BF_{10} = 5.50$]. Post-hoc comparisons indicated conclusive evidence that the mean tSNR was much higher in Dataset 3 ($M = 112.85$, $SD = 16.23$) than in Dataset 1 ($M = 97.55$, $SD = 18.02$) [$t = 3.22$, $p_{schefffe} = .008$, Cohen's $d = 0.91$, $BF_{10} = 16.22$], but there was inconclusive evidence that mean tSNRs differed between Dataset 1 and Dataset 2 ($M = 109.22$, $SD = 16.85$) [$t = 2.39$, $p_{schefffe} = .063$, $d = 0.68$, $BF_{10} = 2.49$], or Dataset 2 and Dataset 3 [$t = 0.89$, $p_{schefffe} = .673$, $d = 0.22$, $BF_{10} = 0.35$]. There were also statistically significant differences in the mean tSNRs in the right ITG ROI across the datasets [$F(2, 85) = 17.83$, $p < .001$, $\eta_p^2 = 0.30$, $BF_{10} = 41,766$]. Post-hoc comparisons indicated conclusive evidence that the mean tSNR was much higher in Dataset 3 ($M = 116.42$, $SD = 14.57$) than in Dataset 1 ($M = 89.88$, $SD = 18.59$) [$t = 5.72$, $p_{schefffe} < .001$, $d = 1.66$, $BF_{10} = 41,084$] and Dataset 2 ($M = 101.15$, $SD = 17.16$) [$t = 3.85$, $p_{schefffe} = .001$, $d = 0.97$, $BF_{10} = 146.66$], but inconclusive evidence that mean tSNRs differed between Datasets 1 and 2 [$t = 2.37$, $p_{schefffe} = .067$, $d = 0.64$, $BF_{10} = 1.97$].

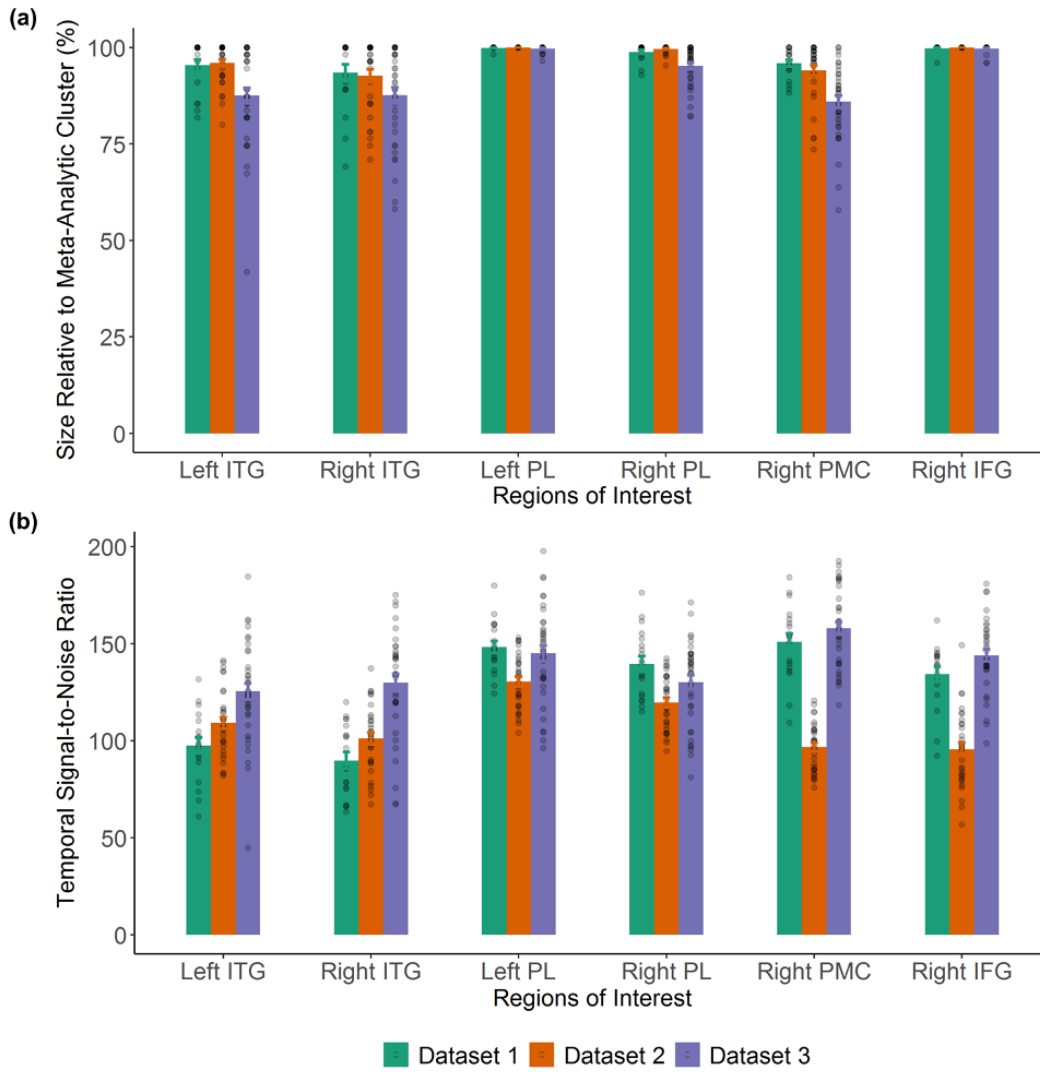


Figure A-2. Size (a) and temporal signal-to-noise ratio (b) in each region of interest across three datasets

ITG: inferior temporal gyrus; PL: parietal lobule; PMC: premotor cortex; IFG: inferior frontal gyrus.

Computational Details of Shape Distance Model

The shape distance measure for any pair of images I_j and I_k was computed using the MATLAB code from https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/shape/sc_digits.html (Belongie et al., 2002). For each image, the edges or contours of the character were first detected, and we sampled 100 approximately uniformly spaced points each along the inner and outer contours of the character. Then an iterative process finds the best one-to-one alignment between the 100 sampled points in character I_j and the 100 sampled points in character I_k . Specifically, for each sampled point in character I_j , the algorithm computed a “shape context” descriptor in the form of a histogram of that particular point to all other 99 points. This histogram is that point’s “shape context”. This was also computed for all the points in character I_k . By comparing the histograms (i.e., the shape contexts of the 100 points in character I_j and the 100 points in character I_k), a point-to-point alignment between the two characters was determined. This alignment was then used to spatially warp or deform one image to fit the other that minimizes the amount of “work” necessary to transform character I_j to align with character I_k . This was performed iteratively through six iterations in our case, which results in a rather stable solution (3 to 5 iterations were used in Belongie, Malik, & Puzicha, 2002). After the matching process, three separate distance measures were computed: (1) D_{jk}^{SC} is the summed shape context cost between the best-aligned “shape contexts” between characters I_j and I_k ; (2) D_{jk}^{BE} is a measure of the amount of spatial transformation (“bending energy”) needed to best align the characters; (3) D_{jk}^{AC} is the appearance cost (sum of squared differences in pixel-by-pixel intensities in Gaussian windows around corresponding points in an image) between the warped character I_j and the original character I_k . Finally, an overall dissimilarity index termed the “shape distance” is computed

using a weighted sum $D_{jk} = D_{jk}^{SC} + 1.6D_{jk}^{AC} + 0.3D_{jk}^{BE}$. The weights used here have previously been optimized using a large set of handwritten digits (Belongie et al., 2002), and have been applied for other object categories (e.g., Gotts, Milleville, Bellgowan, & Martin, 2011).

Representational Geometries Based on Pixel Overlap and Shape Distance

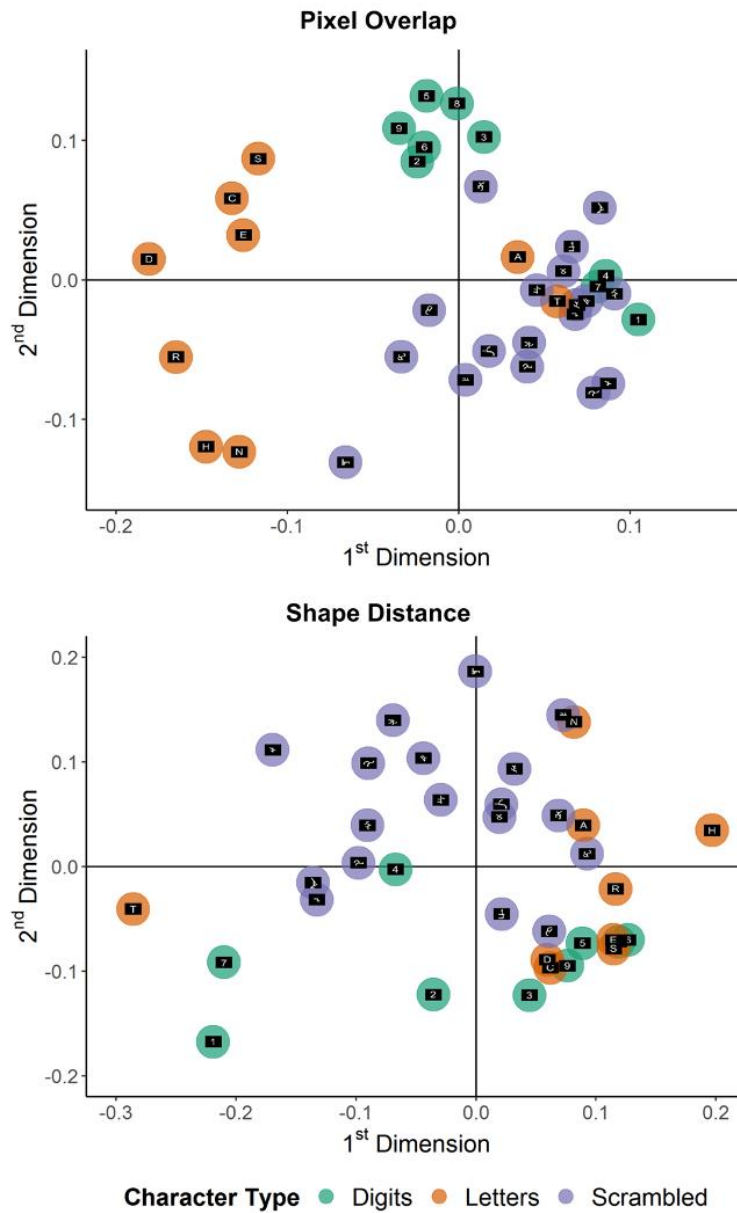


Figure A-3. Representational geometry of the 36 exemplars in two-dimensional space based on pixel overlap and shape distance for Letters Set 1

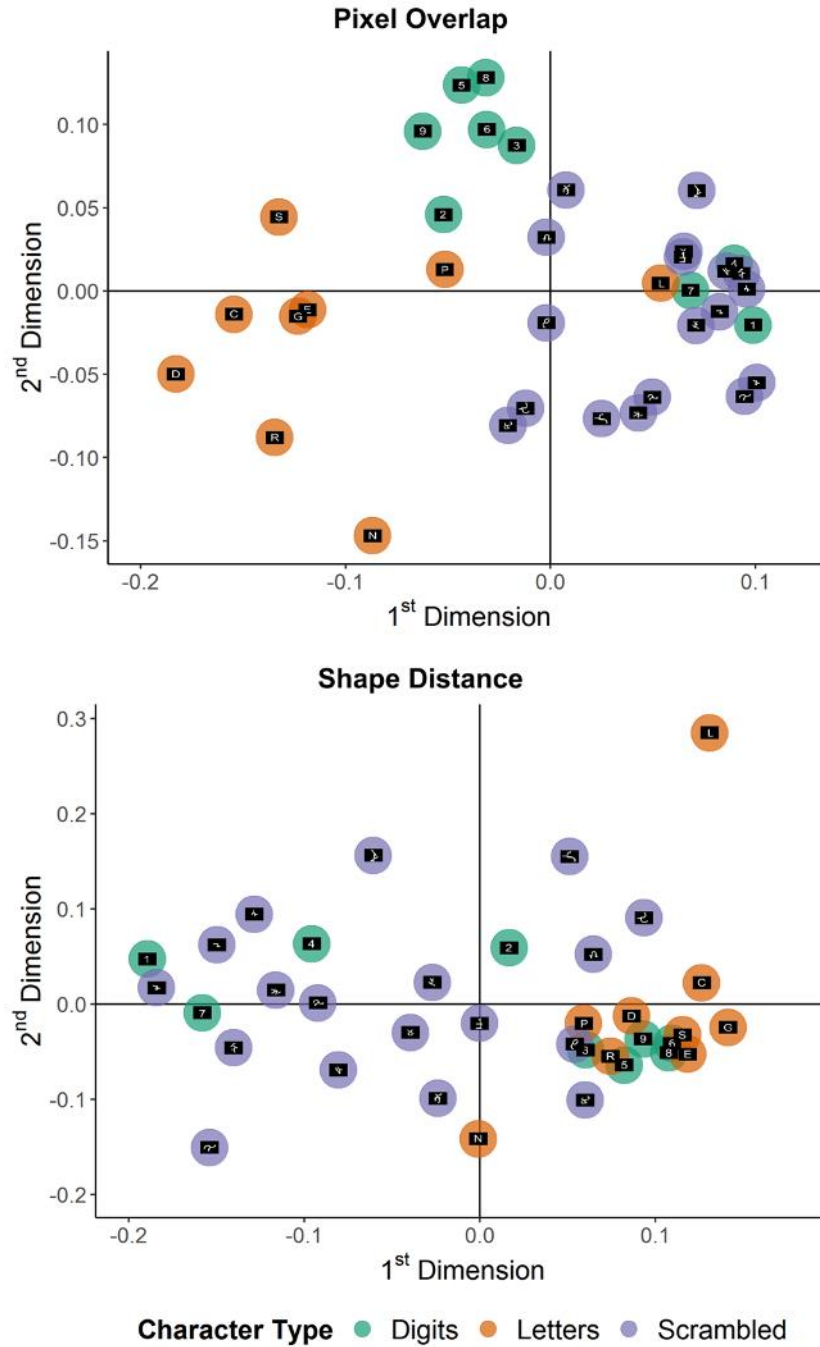


Figure A-4. Representational geometry of the 36 exemplars in two-dimensional space based on pixel overlap and shape distance for Letters Set 2

Comparison of Candidate Models between Letters Sets 1 and 2

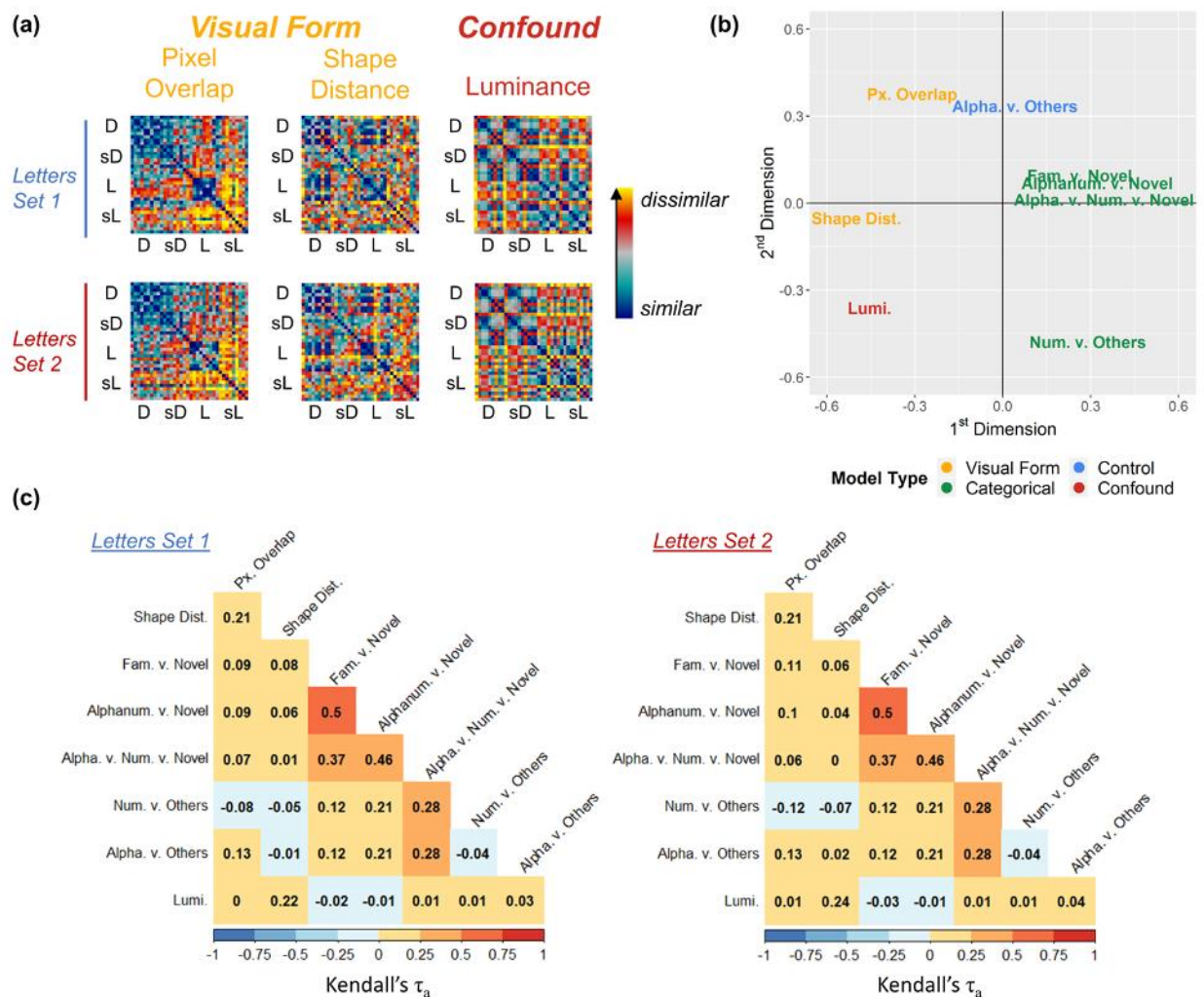


Figure A-5. Comparison of candidate representational dissimilarity models (RDMs) between Letters Set 1 (Datasets 1 and 2) and Letters Set 2 (Dataset 3) (a) Candidate RDMs constructed from 9 exemplars per category: Digits (D), Scrambled Digits (sD), Letters (L), and Scrambled Letters (sL). All models presented were rescaled to [0, 1] for comparative visualization. (b) Multidimensional scaling plot of the correlational distance among the candidate models using Letters Set 2 (see **Figure 3-2** for Letters Set 1). Both Letters Sets resulted in qualitatively similar representational geometry among the candidate representational models. (c) Pairwise rank correlations (Kendall's τ_a) between candidate models for Letters Sets 1 and 2.

Supplementary Results

Representational Geometry in Candidate Numeral-Preferring ITG Regions: Dataset-specific Summary

Table A-5 below provide the mean similarity values for each model within each dataset as shown in Figure 3-4.

Table A-5. Mean degree of dataset-specific similarity between model and neural RDMs in left and right ITG

Model	Left ITG				Right ITG			
	τ_a	SE	p	BF_{+0}	τ_a	SE	p	BF_{+0}
<u>Dataset 1</u>								
Pixel Overlap	.0030	.0060	.461	0.36	.0039	.0054	.354	0.45
Shape Distance	.0078	.0041	.044	1.97	.0003	.0056	.414	0.25
Familiar v. Novel	.0010	.0037	.430	0.29	.0075	.0046	.078	1.35
Alphanumeric v. Novel	.0012	.0039	.369	0.31	.0096	.0043	.016	3.36
Alphabet v. Numbers v. Novel	.0013	.0035	.369	0.32	.0100	.0036	.005	8.37
Numbers v. Others	.0027	.0037	.311	0.45	.0098	.0041	.033	4.34
Alphabet v. Others	-.0011	.0035	.952	0.19	.0028	.0046	.340	0.40
Luminance	.0046	.0038	.091	0.77	.0035	.0052	.297	0.42
<u>Dataset 2</u>								
Pixel Overlap	-.0007	.0036	.570	0.16	.0052	.0039	.131	0.77
Shape Distance	-.0035	.0035	.865	0.10	.00003	.0034	.674	0.19
Familiar v. Novel	.0020	.0045	.482	0.28	.0064	.0050	.402	0.70
Alphanumeric v. Novel	.0024	.0044	.387	0.30	.0072	.0054	.474	0.75
Alphabet v. Numbers v. Novel	.0023	.0036	.332	0.33	.0063	.0046	.216	0.78
Numbers v. Others	.0009	.0028	.482	0.25	.0017	.0027	.387	0.34
Alphabet v. Others	.0016	.0029	.372	0.30	.0045	.0042	.346	0.53
Luminance	-.0012	.0038	.781	0.15	-.0020	.0039	.839	0.13
<u>Dataset 3</u>								
Pixel Overlap	.0014	.0029	.383	0.27	.0029	.0022	.091	0.70
Shape Distance	.0039	.0039	.327	0.48	.0042	.0034	.185	0.63
Familiar v. Novel	.0026	.0034	.306	0.36	.0074	.0056	.285	0.69
Alphanumeric v. Novel	.0023	.0033	.394	0.33	.0095	.0057	.154	1.19
Alphabet v. Numbers v. Novel	.0013	.0029	.406	0.25	.0096	.0044	.093	2.76
Numbers v. Others	-.0032	.0023	.909	0.08	.0083	.0035	.026	4.07
Alphabet v. Others	.0030	.0031	.296	0.45	.0036	.0024	.165	0.91
Luminance	.0037	.0036	.197	0.49	.0071	.0037	.075	1.67

Note. τ_a : Kendall's Tau-a. SE: Standard error of the mean. BF_{+0} = Bayes Factor ($\tau_a > 0$ vs. $\tau_a = 0$). Bold: Result remained significant with FDR correction for multiple comparisons.

Dataset-specific Clustering Analyses of Neural Representations in Candidate Numeral-Preferring Right ITG Region

The multidimensional scaling plots in **Figure 3-5** could only provide a subjective appreciation of any clustering of the right ITG representations for each dataset. Hence, it is unclear whether a three-way distinction (numerals, letters, and novel characters) observed using a model-driven approach could also be observed using a data-driven approach. We addressed this question with a *k*-medoids clustering analysis, which partitions the data in the group-averaged RDM into a pre-specified *k* number of clusters using the “Partitioning Around Medoids” (PAM) algorithm (Kaufman & Rousseeuw, 1990) with the R package *cluster*. *k*-medoids clustering is similar to the more commonly known *k*-means clustering, but it uses medoids (actual data points) instead of centroids (possibly non-existent data points) as the most representative data point in a cluster. The PAM algorithm uses a dissimilarity matrix as input, and is less influenced by outlying data points compared to *k*-means. Thus, it is a more robust approach for clustering data points. Below, we focused on characterizing the 3-cluster structure observed in our primary analyses.

Dataset 1

As shown in **Table A-6** and **Figure A-6**, digits have slight dominance in Cluster #2 (45.5% digits, 36.4% novel, and 18.2% letters; medoid: digit), whereas novel characters have slight dominance in Cluster #1 (61.5% novel, 30.8% letters, 7.7% digits; medoid: novel) and Cluster #3 (50% novel, 25% letters, 25% digits; medoid: novel). 55.6% of digits are in Cluster #2, whereas no more than 45% of letters or of novel characters are found in any cluster. We then quantified the similarity between the cluster membership and the actual categorical membership, adjusted for chance, using the adjusted Rand Index [ranges from -1 (perfect disagreement) to 1

(perfect agreement), with 0 being chance-level agreement, using the R package *ClusterR*]. The adjusted Rand Index is 0.00659, suggesting that the observed clustering is as good as random.

Finally, we also used the average silhouette method to independently determine the optimal number of clusters k based on the quality of clustering. The silhouette method uses the average silhouette width, which is a measure of how close each data point is to other data points in its own cluster relative to how close it is to other data points in other clusters (Rousseeuw, 1987). It ranges from -1 (poorly clustered, and no different from random assignments) to 1 (well-clustered). By performing the PAM algorithm across a range of k values, the number of clusters that would maximize the average silhouette width is then selected as the optimal k . The silhouette method revealed that the number of clusters ranging from the least optimal to the most is 3, 2, and 4. The average silhouette widths range from 0.0154 to 0.0163, which are extremely low, and indicate that no substantial structure was found (i.e., the data points are very close to the decision boundaries). On the other hand, the adjusted Rand Indices for $k = 2$ and $k = 4$ are -0.00107 and -0.0169 , respectively, which suggest that the 2- and 4-cluster structures have slightly lower agreement with the actual category membership than a 3-cluster structure, and below that expected by chance. In a 4-cluster structure, digits have dominance in one cluster (50% digits, 20% letters, 30% novel among 10 exemplars; medoid: digit), and novel characters have dominance in the other three clusters (60%, 50%, and 62.5% novel among 10, 8, and 8 exemplars, respectively; medoids: all novel).

Table A-6. Relation between k -medoids cluster membership in the right ITG in Dataset 1 and the actual category membership

	Cluster			Total
	1	2	3	
Digits	1	5	3	9
Letters	4	2	3	9
Novel	8	4	6	18
Total	13	11	12	36

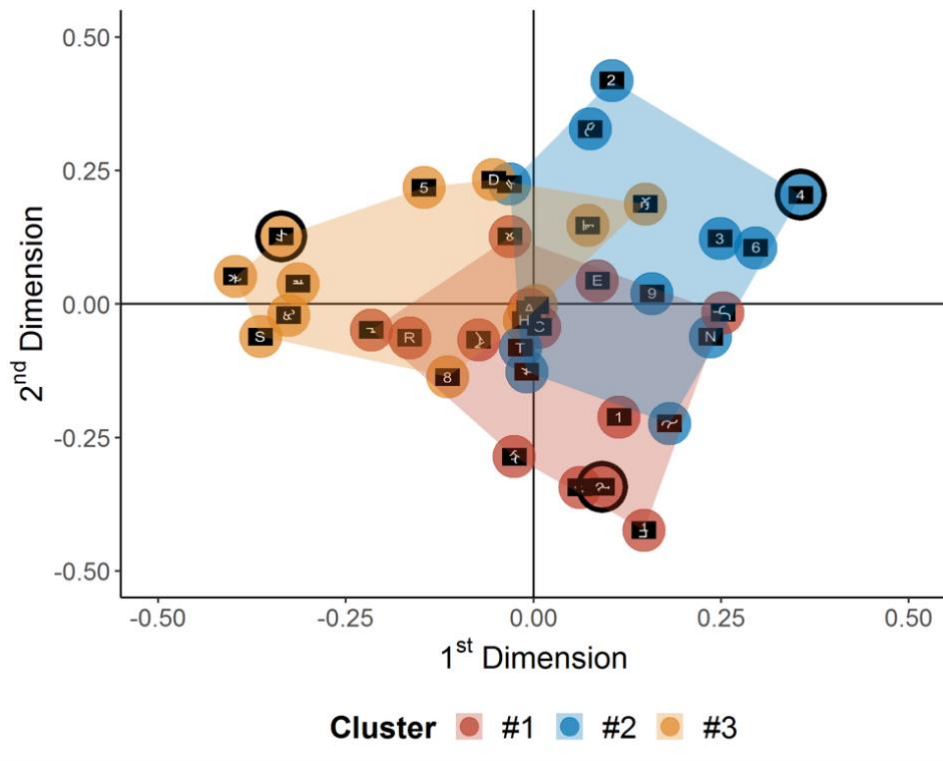


Figure A-6. k -medoids clustering of the neural representations of 36 exemplars in the numeral-preferring right ITG region in Dataset 1. Dimensions are identical to those in **Figure 3-5**. Circled exemplars are medoids of the clusters.

Dataset 2

As shown in **Table A-7** and **Figure A-7**, novel characters have slight dominance in Clusters #1 (60% novel and 40% digits; medoid: digit) and #3 (58.3% novel, and 41.7% letters, medoid: novel), and co-dominance with digits in Cluster #2 (35.7% novel, 35.7% digits, 28.6% letters; medoid: novel). However, the 3-cluster structure is noticeably weak. The adjusted Rand Index is 0.0414, suggesting that the observed clustering is as good as random.

The silhouette method revealed that the number of clusters ranging from the least optimal to the most is 3, 4, and 2. The average silhouette widths range from 0.0115 to 0.014, which are extremely low, and indicate that no substantial structure was found. The adjusted Rand Indices for $k = 2$ and $k = 4$ are 0.0273 and 0.0767, respectively. Taken together, these suggest that a 4-cluster structure may have a slightly higher agreement with the actual category membership than a 3-cluster structure. In a 4-cluster structure, one cluster has slight digit-dominance (55.6% digits and 44.4% novel among 9 exemplars; medoid: digit), one cluster has slight letter-dominance (50% letters, 25% digits, and 25% novel among 8 exemplars; medoid: digit), and two clusters have slight novel character-dominance (70% novel, 20% digits, and 10% letters among 10 exemplars in one cluster; medoid: novel; 55.6% novel and 44.4% letters among 9 exemplars in another cluster; medoid: novel).

Table A-7. Relation between k -medoids cluster membership in the right ITG in Dataset 2 and the actual category membership

	Cluster			Total
	1	2	3	
Digits	4	5	0	9
Letters	0	4	5	9
Novel	6	5	7	18
Total	10	14	12	36

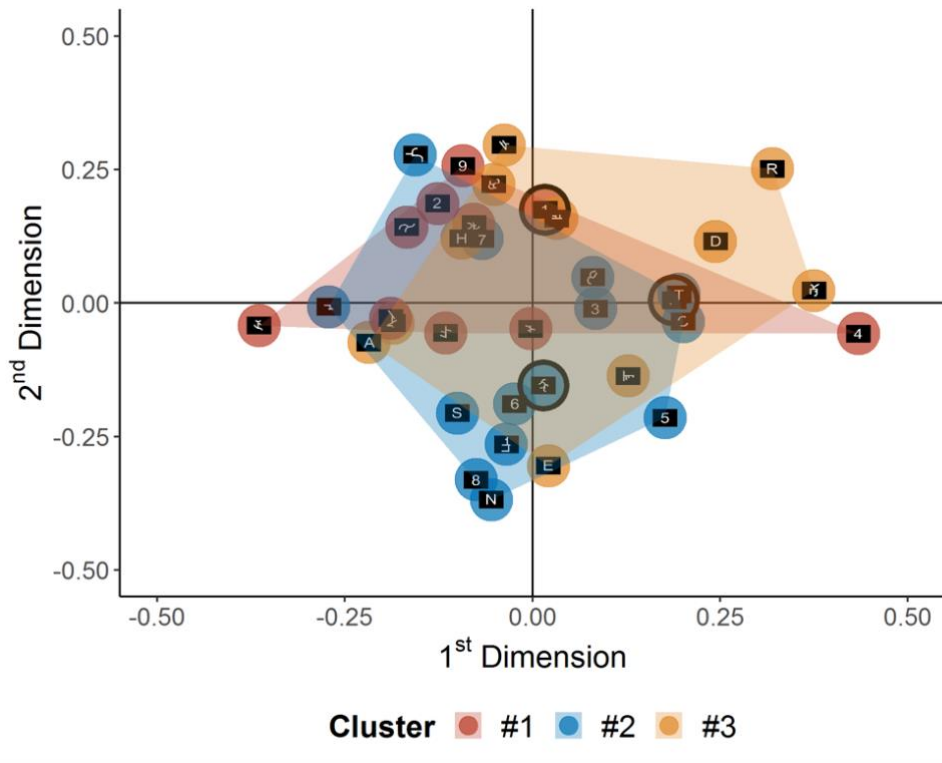


Figure A-7. *k*-medoids clustering of the neural representations of 36 exemplars in the numeral-preferring right ITG region in Dataset 2. Dimensions are identical to those in **Figure 3-5**. Circled exemplars are medoids of the clusters.

Dataset 3

As shown in **Table A-8** and **Figure A-8**, digits have slight dominance in Cluster #2 (50% digits, 33.3% novel, and 16.7% letters; medoid: digit), whereas letters have co-dominance with novel characters in Cluster #1 (41.7% letters, 41.7% novel, 16.7% digits; medoid: letter), and novel characters have dominance in Cluster #3 (75% novel, 16.7% letters, 8.3% digits; medoid: novel). 66.7% of digits are in Cluster #2, 55.6% of letters are in Cluster #1, and 50% of novel characters are in Cluster #3. The adjusted Rand Index is 0.0660, suggesting that the observed clustering is as good as random.

The silhouette method revealed that the number of clusters ranging from the least optimal to the most is 2, 4, and 3. The average silhouette widths range from 0.0124 to 0.0159, which are extremely low, and indicate that no substantial structure was found. The adjusted Rand Indices for $k = 2$ and $k = 4$ are -0.0297 and 0.0692 , respectively, which suggest that a 4-cluster structure may have a slightly higher agreement with the actual category membership than a 3-cluster structure. In a 4-cluster structure, digits have dominance in one cluster (66.7% digits, 22.2% novel, 11.1% letters among 9 exemplars; medoid: digit), novel characters have dominance in two clusters (66.7% and 75% among 9 and 8 exemplars, respectively; medoids: all novel), and letters have co-dominance with novel characters in one cluster (40% letters, 40% novel, 20% digits among 10 exemplars; medoid: letter). Taken together, a clear digit-dominated cluster could be observed in both 3- and 4-cluster structures.

Table A-8. Relation between k -medoids cluster membership in the right ITG in Dataset 3 and the actual category membership

	Cluster			Total
	1	2	3	
Digits	2	6	1	9
Letters	5	2	2	9
Novel	5	4	9	18
Total	12	12	12	36

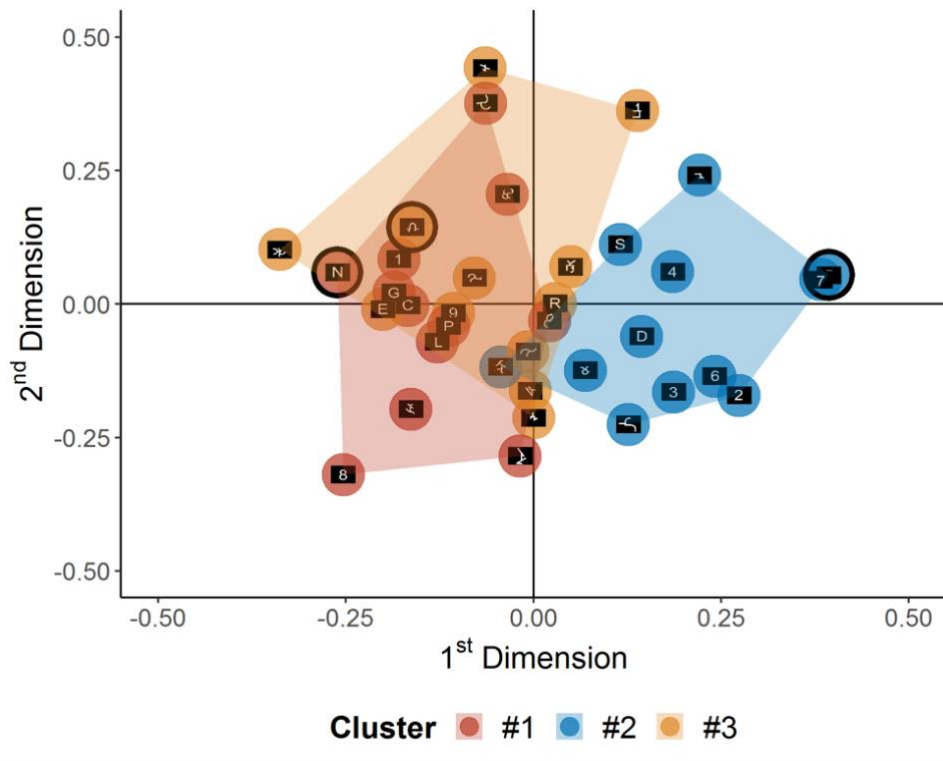


Figure A-8. k -medoids clustering of the neural representations of 36 exemplars in the numeral-preferring right ITG region in Dataset 3. Dimensions are identical to those in **Figure 3-5**. Circled exemplars are medoids of the clusters.

Representational Geometry in Candidate Numeral-Preferring Parietal Regions

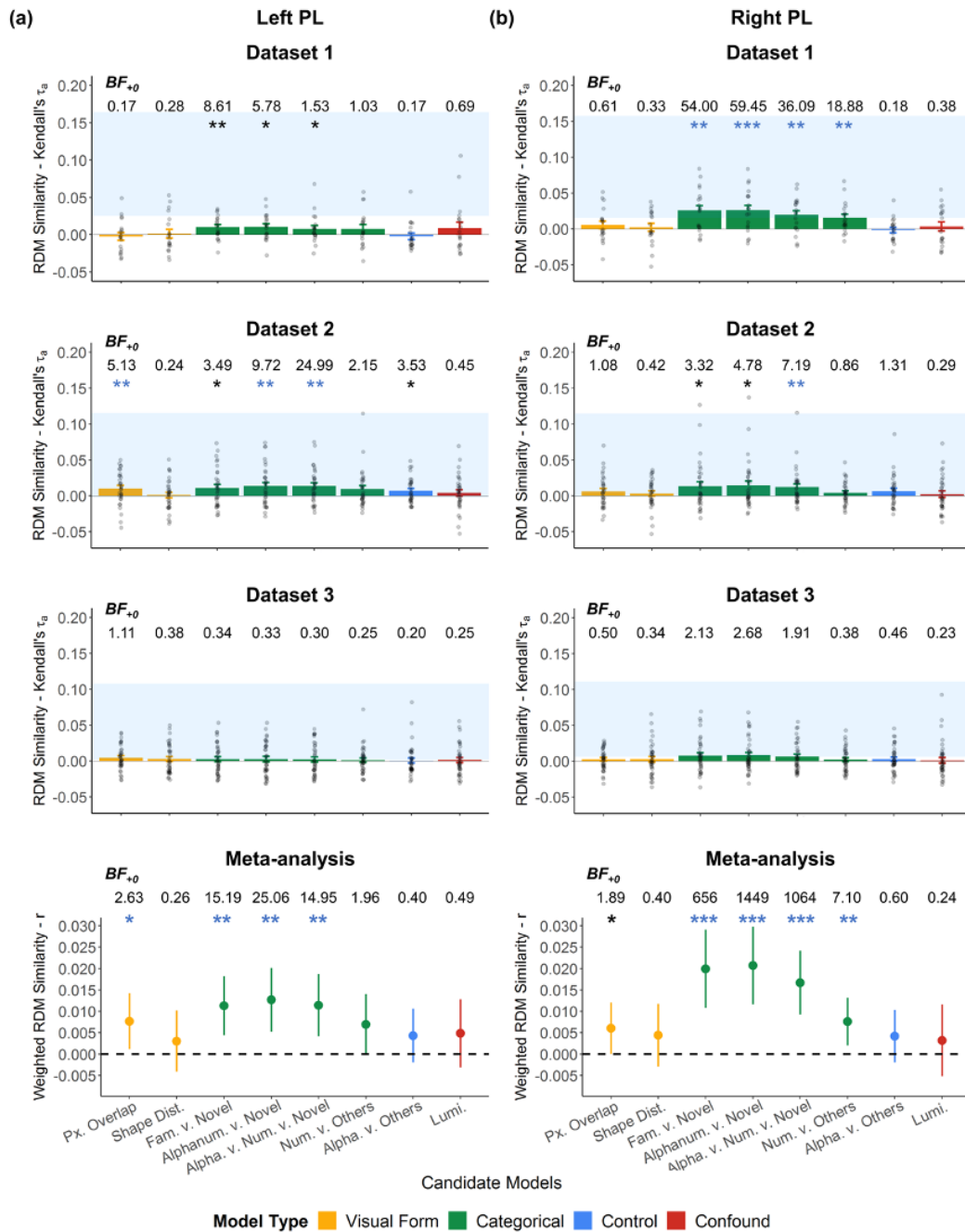


Figure A-9. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-preferring regions in (a) left and (b) right parietal lobule (PL). Blue bars indicate the estimated noise ceiling. Group means and standard errors of the similarity are indicated by the bar plots with error bars. Individual data points are shown as grey dots. Evidence of similarity is indicated by black asterisk: * $p < .05$, ** $p < .01$, *** $p < .001$, one-sided, uncorrected. Blue asterisks indicated results that remained significant with FDR correction. BF_{+0} = Bayes Factor [$r_z > 0$ vs. $r_z = 0$]. Lines in meta-analytic plots indicate the 95% confidence interval around the overall weighted r .

Table A-9. Meta-analyses for the degree of similarity between each model RDM and neural RDMs in left and right parietal lobules (PL)

Model	Left PL				Right PL			
	<i>r</i>	95% CI	<i>p</i>	BF_{+0}	<i>r</i>	95% CI	<i>p</i>	BF_{+0}
Pixel Overlap	.0077	[.0011, .0142]	.021	2.63	.0060	[.00003, .0120]	.049	1.89
Shape Distance	.0030	[-.0041, .0102]	.406	0.26	.0044	[-.0029, .0118]	.239	0.40
Familiar v. Novel	.0113	[.0044, .0182]	.001	15.19	.0199	[.0108, .0291]	<.001	655.53
Alphanumeric v. Novel	.0127	[.0052, .0202]	.001	25.06	.0207	[.0116, .0298]	<.001	1448.55
Alphabet v. Numbers v. Novel	.0114	[.0041, .0187]	.002	14.95	.0167	[.0092, .0242]	<.001	1064.43
Numbers v. Others	.0070	[-.0002, .0141]	.056	1.96	.0076	[.0020, .0132]	.008	7.10
Alphabet v. Others	.0043	[-.0020, .0106]	.180	0.40	.0042	[-.0020, .0103]	.182	0.60
Luminance	.0048	[-.0031, .0128]	.235	0.49	.0032	[-.0052, .0116]	.457	0.24

Note. 95% CI: 95% confidence interval. BF_{+0} = Bayes Factor ($r > 0$ vs. $r = 0$)

Left Parietal Lobule (PL)

Overall, there was strong meta-analytic evidence of similarity between the neural RDMs and categorical model RDMs that are merely category-sensitive (Familiar v. Novel, Alphanumeric v. Novel, and Alphabet v. Numbers v. Novel) and weak, inconclusive evidence of similarity between the neural RDMs and the Pixel Overlap model RDM (see **Figure A-9(a)** and **Table A-9**). Below we report whether the meta-analytic findings were also observed in each dataset.

Visual Form Models. There was moderate evidence of similarity between the neural and Pixel Overlap model RDMs only in Dataset 2 ($\tau_a = .0102$, $p = .008$, $BF_{+0} = 5.13$). There was no evidence of similarity between the neural and Shape Distance model RDMs in any dataset.

Categorical Models. For Dataset 1, there was moderate evidence of similarity between

the neural RDMs and the Familiar v. Novel model RDM ($\tau_a = .0101, p = .008, BF_{+0} = 8.61$), the Alphanumeric v. Novel model RDM ($\tau_a = .0104, p = .013, BF_{+0} = 5.78$), and weak evidence for the Alphabet v. Numbers v. Novel model RDM ($\tau_a = .0079, p = .048, BF_{+0} = 1.53$). There was still moderate to strong evidence of similarity between the neural RDMs and two of the three categorical model RDMs after controlling for the visual form and confound model RDMs (Familiar v. Novel: $\tau_a = .0149, p = .004, BF_{+0} = 15.08$; Alphanumeric v. Novel: $\tau_a = .0140, p = .007, BF_{+0} = 9.26$; Alphabet v. Numbers v. Novel: $\tau_a = .0118, p = .030, BF_{+0} = 1.66$). The evidence for the Alphabet v. Numbers v. Novel model was therefore inconclusive. Post-hoc pairwise comparisons of the three models revealed no evidence of within-participant differences between these model RDMs in their similarity to the neural RDMs (all $ps > .225, BF_{10}s < 0.45$ for both zero-order and semipartial correlations).

For Dataset 2, there was moderate to strong evidence of similarity between the neural RDMs and all three categorical model RDMs that are merely category-sensitive (Familiar v. Novel: $\tau_a = .0110, p = .042, BF_{+0} = 3.49$; Alphanumeric v. Novel: $\tau_a = .0137, p = .009, BF_{+0} = 9.72$; Alphabet v. Numbers v. Novel: $\tau_a = .0138, p = .002, BF_{+0} = 24.99$). There was still moderate to strong evidence of similarity between the neural RDMs and these categorical model RDMs after controlling for the visual form and confound model RDMs (Familiar v. Novel: $\tau_a = .0145, p = .044, BF_{+0} = 3.11$; Alphanumeric v. Novel: $\tau_a = .0168, p = .011, BF_{+0} = 8.00$; Alphabet v. Numbers v. Novel: $\tau_a = .0192, p = .003, BF_{+0} = 18.78$). However, it is important to note that the BFs for the Familiar v. Novel model in both the zero-order and semipartial correlations were not robust to varied priors as they decreased below 3 with wider Cauchy priors (≥ 1). Post-hoc pairwise comparisons of these three models revealed no evidence of within-participant differences between these three categorical model RDMs in their similarity to the

neural RDMs (all $ps > .080$, $BF_{10s} < 2.28$ for both zero-order and semipartial correlations).

For Dataset 3, there was no evidence of similarity between the neural RDMs and any of the categorical model RDMs.

Figure A-10(a) illustrates the group-averaged representational geometry of the 36 characters in this region for Datasets 1 and 2.

Control Model. There was moderate evidence of similarity between the neural and Alphabet v. Others model RDM only in Dataset 2 ($\tau_a = .0071$, $p = .032$, $BF_{+0} = 3.53$), but the BF was not robust to varied priors as it decreased below 3 with wider priors (≥ 1). Moreover, the evidence of their similarity was weakened after controlling for the visual form and confound model RDMs ($\tau_a = .0083$, $p = .092$, $BF_{+0} = 1.74$).

Confound Model. There was no evidence of similarity between the neural RDMs and Luminance model RDM in any dataset.

Right PL

Similar to the left PL, there was moderate to extreme meta-analytic evidence of similarity between the neural RDMs and all categorical model RDMs, including the numeral-sensitive model (Familiar v. Novel, Alphanumeric v. Novel, and Alphabet v. Numbers v. Novel, and Numbers v. Others) and weak, inconclusive evidence of similarity between the neural RDMs and the Pixel Overlap model RDM (see **Figure A-9(b)** and **Table A-9**). Below we report whether the meta-analytic findings were also observed in each dataset.

Visual Form Models. There was no evidence of similarity between the neural RDMs and Pixel Overlap or Shape Distance model RDMs in any dataset.

Categorical Models. For Dataset 1, there was strong to very strong evidence of similarity between the neural RDMs and all four categorical model RDMs (Familiar v. Novel: τ_a

= .0257, $p = .001$, $BF_{+0} = 54.00$; Alphanumeric v. Novel: $\tau_a = .0262$, $p < .001$, $BF_{+0} = 59.45$;
 Alphabet v. Numbers v. Novel: $\tau_a = .0199$, $p = .002$, $BF_{+0} = 36.09$; Numbers v. Others: $\tau_a =$
 .0156, $p = .002$, $BF_{+0} = 18.88$). There was still strong to very strong evidence of similarity
 between the neural RDMs and these categorical model RDMs after controlling for the visual
 form and confound model RDMs (Familiar v. Novel: $\tau_a = .0358$, $p = .002$, $BF_{+0} = 53.86$;
 Alphanumeric v. Novel: $\tau_a = .0334$, $p < .001$, $BF_{+0} = 63.29$; Alphabet v. Numbers v. Novel: τ_a
 = .0287, $p = .001$, $BF_{+0} = 39.22$; Numbers v. Others: $\tau_a = .0232$, $p = .001$, $BF_{+0} = 18.34$). Post-
 hoc pairwise comparisons revealed that within participants, the neural RDMs were more similar
 to the Alphanumeric v. Novel model RDM than to the Alphabet v. Numbers v. Novel model
 RDM ($p = .008$, $BF_{10} = 3.86$), although the BF decrease to below 3 with an ultrawide prior.
 Evidence for this pairwise difference appeared less conclusive after controlling for the visual
 form and confound models ($p = .029$, $BF_{10} = 0.71$). A somewhat similar trend was observed
 between the Alphabet v. Numbers v. Novel and Familiar v. Novel model RDMs, but the
 evidence appeared largely inconclusive (zero-order: $p = .018$, $BF_{10} = 0.85$; semipartial: $p = .040$,
 $BF_{10} = 0.58$). Other than the trends for these pairs of models, there was no evidence of other
 within-participant differences between these categorical model RDMs in their similarity to the
 neural RDMs (all $ps > .060$, $BF_{10}s < 0.90$ for both zero-order and semipartial correlations).

For Dataset 2, there was moderate evidence of similarity between the neural RDMs and
 three categorical model RDMs that are merely category-sensitive (Familiar v. Novel: $\tau_a = .0136$,
 $p = .044$, $BF_{+0} = 3.32$; Alphanumeric v. Novel: $\tau_a = .0146$, $p = .030$, $BF_{+0} = 4.78$; Alphabet v.
 Numbers v. Novel: $\tau_a = .0121$, $p = .006$, $BF_{+0} = 7.19$). There was still moderate evidence
 between the neural RDMs and two of these three these categorical model RDMs after controlling
 for the visual form and confound model RDMs (Familiar v. Novel: $\tau_a = .0183$, $p = .049$, $BF_{+0} =$

2.89; Alphanumeric v. Novel: $\tau_a = .0182$, $p = .030$, $BF_{+0} = 4.20$; Alphabet v. Numbers v. Novel: $\tau_a = .0172$, $p = .007$, $BF_{+0} = 6.49$). Note that the evidence for the Familiar v. Novel model was relatively weak. It is also important to note that the BF s for the Familiar v. Novel model in both the zero-order and semipartial correlations were not robust to varied priors as they decreased below 3 with wide priors. Post-hoc pairwise comparisons revealed no evidence of within-participant differences between these three categorical model RDMs in their similarity to the neural RDMs (all $ps > .253$, $BF_{10}s < 0.40$ for both zero-order and semipartial correlations).

For Dataset 3, there was no evidence of similarity between the neural RDMs and any of the categorical model RDMs.

Figure A-10(b) illustrates the group-averaged representational geometry of the 36 characters in this region for Datasets 1 and 2.

Control Model. There was no evidence of similarity between the neural RDMs and Alphabet v. Others model RDM in any dataset.

Confound Model. There was no evidence of similarity between the neural RDMs and Luminance model RDM in any dataset.

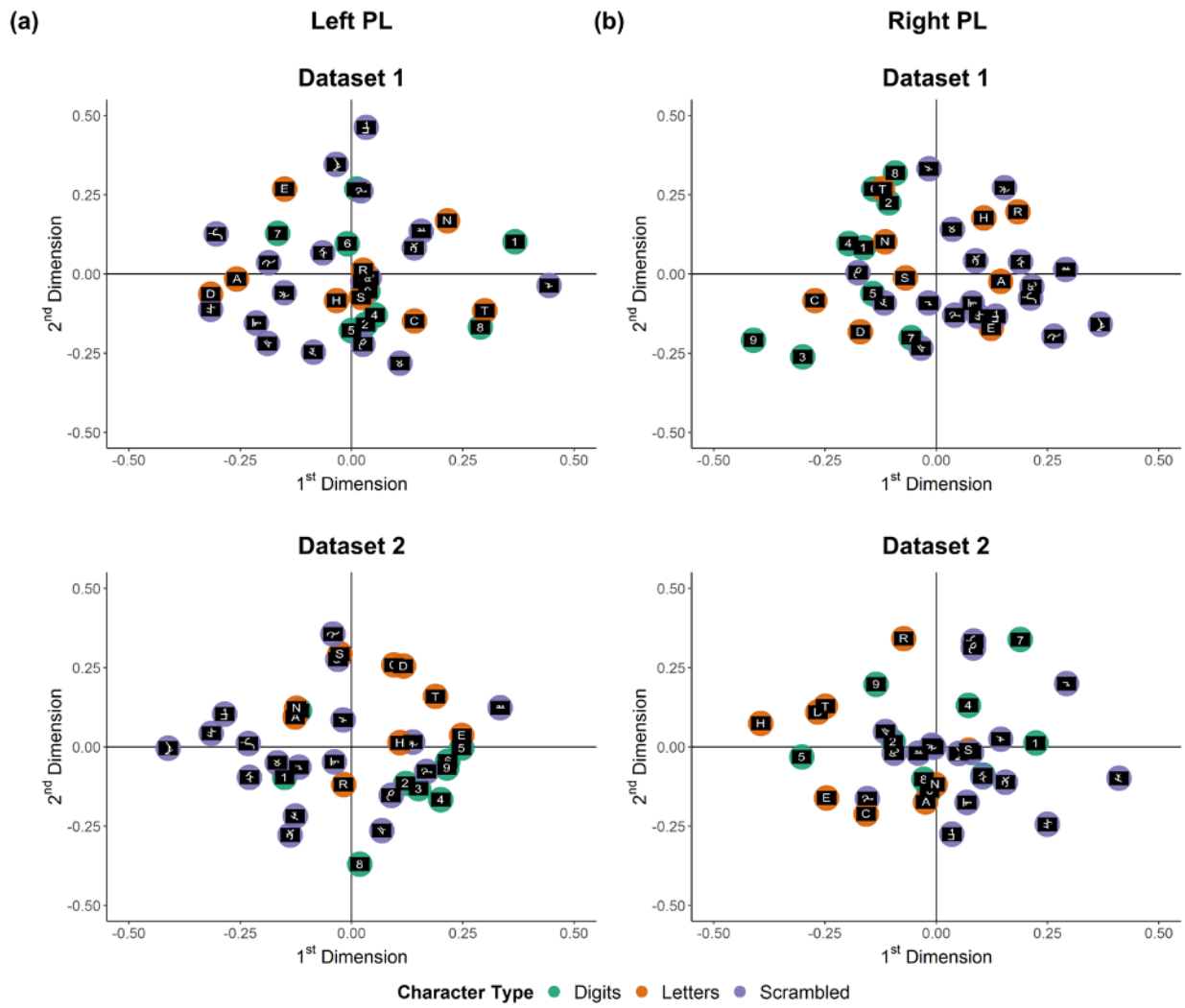


Figure A-10. Group-averaged representational geometry of the 36 exemplars in two-dimensional space in the numeral-preferring (a) left and (b) right parietal lobules (PL) in Datasets 1 and 2. Three-dimensional interactive plots are available at <https://osf.io/jwgk8/wiki/home>.

Representational Geometry in Candidate Numeral-Preferring Frontal Regions

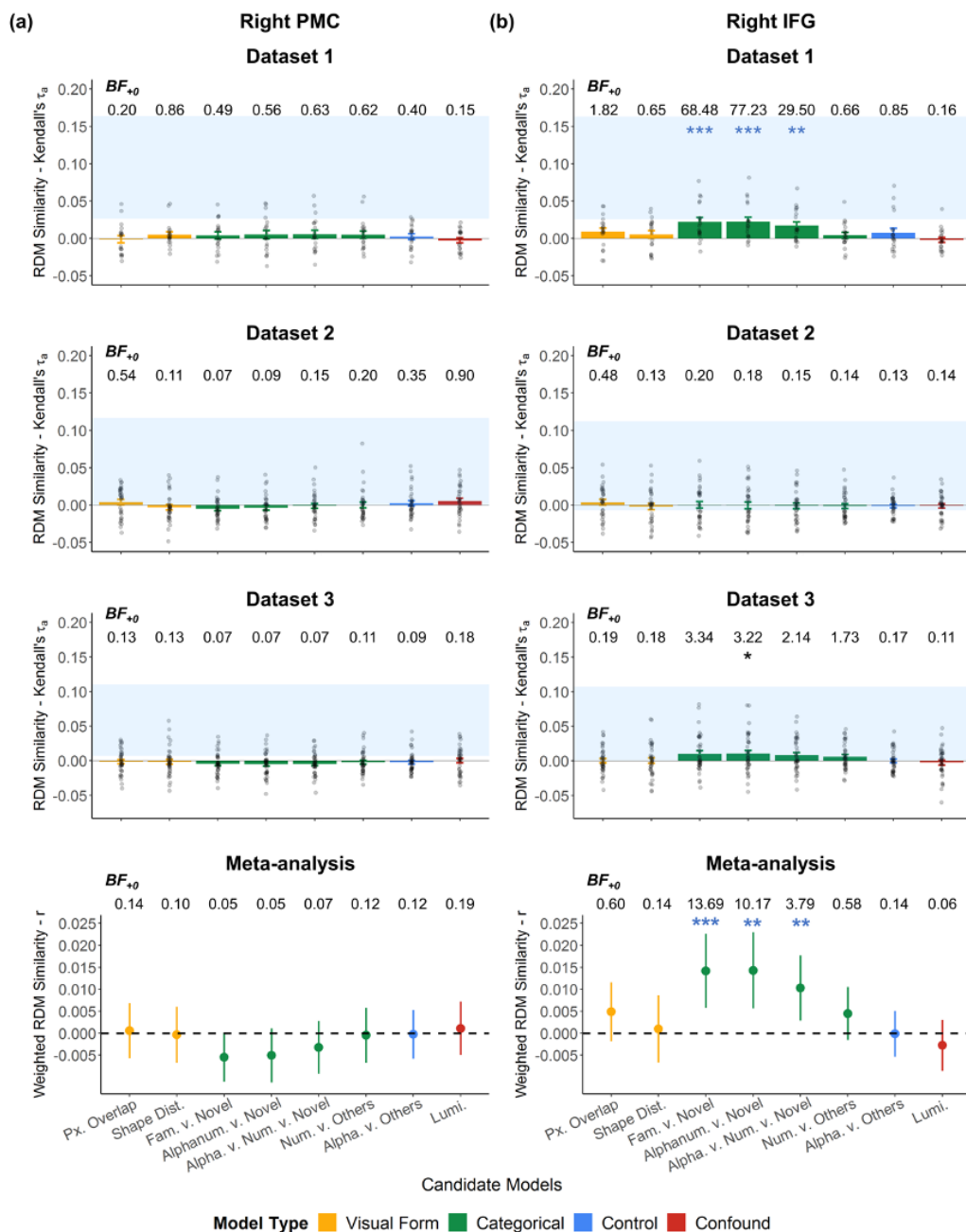


Figure A-11. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-preferring regions in the right (a) premotor cortex (PMC) and (b) inferior frontal gyrus (IFG)

Blue bars indicate the estimated noise ceiling. Group means and standard errors of the similarity are indicated by the bar plots with error bars. Individual data points are shown as grey dots. Evidence of similarity is indicated by black asterisk: * $p < .05$, ** $p < .01$, *** $p < .001$, one-sided, uncorrected. Blue asterisks indicated results that remained significant with FDR correction. BF_{+0} = Bayes Factor [$r_z > 0$ vs. $r_z = 0$]. Lines in meta-analytic plots indicate the 95% confidence interval around the overall weighted r .

Table A-10. Meta-analyses for the degree of similarity between each model RDM and neural RDMs in right premotor (PMC) and inferior frontal (IFG) regions

Model	Right PMC				Right IFG			
	<i>r</i>	95% CI	<i>p</i>	<i>BF</i> ₊₀	<i>r</i>	95% CI	<i>p</i>	<i>BF</i> ₊₀
Pixel Overlap	.0006	[-.0057, .0068]	.862	0.14	.0049	[-.0018, .0116]	.154	0.60
Shape Distance	-	[-.0067, .0060]	.909	0.10	.0010	[-.0067, .0086]	.806	0.14
Familiar v. Novel	-	[-.0110, .0001]	.056	0.05	.0142	[.0057, .0226]	<.001	13.69
Alphanumeric v. Novel	-	[-.0112, .0011]	.106	0.05	.0143	[.0056, .0229]	.001	10.17
Alphabet v. Numbers v. Novel	-	[-.0092, .0027]	.287	0.07	.0103	[.0028, .0177]	.007	3.79
Numbers v. Others	-	[-.0067, .0058]	.883	0.12	.0045	[-.0016, .0105]	.147	0.58
Alphabet v. Others	-	[-.0058, .0052]	.924	0.12	-	[-.0054, .0051]	.954	0.14
Luminance	.0011	[-.0050, .0072]	.726	0.19	-	[-.0086, .0030]	.348	0.06

Note. 95% CI: 95% confidence interval. *BF*₊₀ = Bayes Factor (*r* > 0 vs. *r* = 0)

Right Premotor Cortex (PMC)

Overall, there was moderate to strong meta-analytic evidence of a lack of similarity between the neural RDMs and any model RDM (see **Figure A-11(a)** and **Table A-10**). Below we report whether the meta-analytic findings were also observed in each dataset.

Visual Form Models. There was no evidence of similarity between the neural RDMs and Pixel Overlap or Shape Distance model RDM in any dataset.

Categorical Models. There was no evidence of similarity between the neural RDMs and any of the categorical model RDMs in any dataset.

Control Model. There was no evidence of similarity between the neural RDMs and Alphabet v. Others model RDM in any dataset.

Confound Model. There was no evidence of similarity between the neural RDMs and

Luminance model RDM in any dataset.

Right Inferior Frontal Gyrus (IFG)

Overall, there was moderate to strong meta-analytic evidence of similarity between the neural RDMs and categorical model RDMs that are merely category-sensitive (Familiar v. Novel, Alphanumeric v. Novel, and Alphabet v. Numbers v. Novel) (see **Figure A-11(b)** and **Table A-10**) However, the *BF* for the Alphabet v. Numbers v. Novel was not robust to varied priors as it decreased below 3 with wide priors.

As there was also considerable variance across datasets in the effect sizes for the Familiar v. Novel model (test of residual heterogeneity $p = .009$; τ^2 (between-study variance) = 0.00023, and I^2 (estimated proportion of variance across studies due to differences among true effect sizes) = 79.9%), Alphanumeric v. Novel model (test of residual heterogeneity $p = .006$; $\tau^2 = 0.00026$, and $I^2 = 81.2\%$), and Alphabet v. Numbers v. Novel model (test of residual heterogeneity $p = .013$; $\tau^2 = 0.00016$, and $I^2 = 78.0\%$), we recomputed the meta-analyses using a random-effects approach with the Restricted Maximum Likelihood method. The overage weighted correlations were .0165, 95% CI [-.0026, .0356], $p = .091$, for the Familiar v. Novel model; .0165, 95% CI [-.0037, .0367], $p = .109$, for the Alphanumeric v. Novel model; and .0119, 95% CI [-.0041, .0280], $p = .144$, for the Alphabet v. Numbers v. Novel model. Nonetheless, these residual heterogeneities were model-specific and not about the neural data per se because the tests for 5 out of 8 models did not show significant residual heterogeneity (all $ps > .236$). It should be noted that such assessments of residual heterogeneity have low power when implemented with a small number of studies (< 20) (Huedo-Medina et al., 2006), so these results should be interpreted with caution. Below we report whether the meta-analytic findings were also observed in each dataset.

Visual Form Models. There was no evidence of similarity between the neural RDMs and the Pixel Overlap or Shape Distance model RDM in any dataset.

Categorical Models. For Dataset 1, there was strong evidence of similarity between the neural RDMs and three categorical model RDMs that are merely category-sensitive (Familiar v. Novel: $\tau_a = .0222$, $p < .001$, $BF_{+0} = 68.48$; Alphanumeric v. Novel: $\tau_a = .0226$, $p < .001$, $BF_{+0} = 77.23$; Alphabet v. Numbers v. Novel: $\tau_a = .0170$, $p = .002$, $BF_{+0} = 29.50$). There was still strong evidence of similarity between the neural RDMs and these categorical model RDMs after controlling for the visual form and confound model RDMs (Familiar v. Novel: $\tau_a = .0298$, $p < .001$, $BF_{+0} = 49.00$; Alphanumeric v. Novel: $\tau_a = .0280$, $p < .001$, $BF_{+0} = 54.61$; Alphabet v. Numbers v. Novel: $\tau_a = .0243$, $p = .005$, $BF_{+0} = 22.75$). Post-hoc pairwise comparisons revealed that within participants, there was some weak evidence the neural RDMs were more similar to the Alphanumeric v. Novel model RDM than to the Alphabet v. Numbers v. Novel model RDM ($p = .045$, $BF_{10} = 2.19$). This pairwise difference was no longer statistically significant after controlling for the visual form and confound models ($p = .080$, $BF_{10} = 0.47$).

For Dataset 2, there was no evidence of similarity between the neural RDMs and any of the categorical model RDMs.

For Dataset 3, between the frequentist and Bayesian approaches, there was some inconsistent evidence of similarity between the neural RDMs and the Familiar v. Novel model RDM ($\tau_a = .0102$, $p = .064$, $BF_{+0} = 3.34$), but relatively more consistent evidence of similarity between the neural RDMs and the Alphanumeric v. Novel model RDM ($\tau_a = .0105$, $p = .040$, $BF_{+0} = 3.22$). There was somewhat more consistent evidence of similarity between the neural RDMs and these categorical model RDMs after controlling for the visual form and confound model RDMs (Familiar v. Novel: $\tau_a = .0143$, $p = .043$, $BF_{+0} = 3.68$; Alphanumeric v. Novel: τ_a

= .0135, $p = .025$, $BF_{+0} = 3.48$). However, the BF s for both of these models in both the zero-order and semipartial correlations were not robust to varied priors as they decreased below 3 with wide Cauchy priors (≥ 1). Post-hoc pairwise comparisons revealed no evidence of within-participant differences between these two categorical model RDMs in their similarity to the neural RDMs (all $ps > .213$, $BF_{10s} < 0.22$ for both zero-order and semipartial correlations).

Figure A-12 illustrates the group-averaged representational geometry of the 36 characters in this region for Datasets 1 and 3.

Control Model. There was no evidence of similarity between the neural RDMs and Alphabet v. Others model RDM in any dataset.

Confound Model. There was no evidence of similarity between the neural RDMs and Luminance model RDM in any dataset.

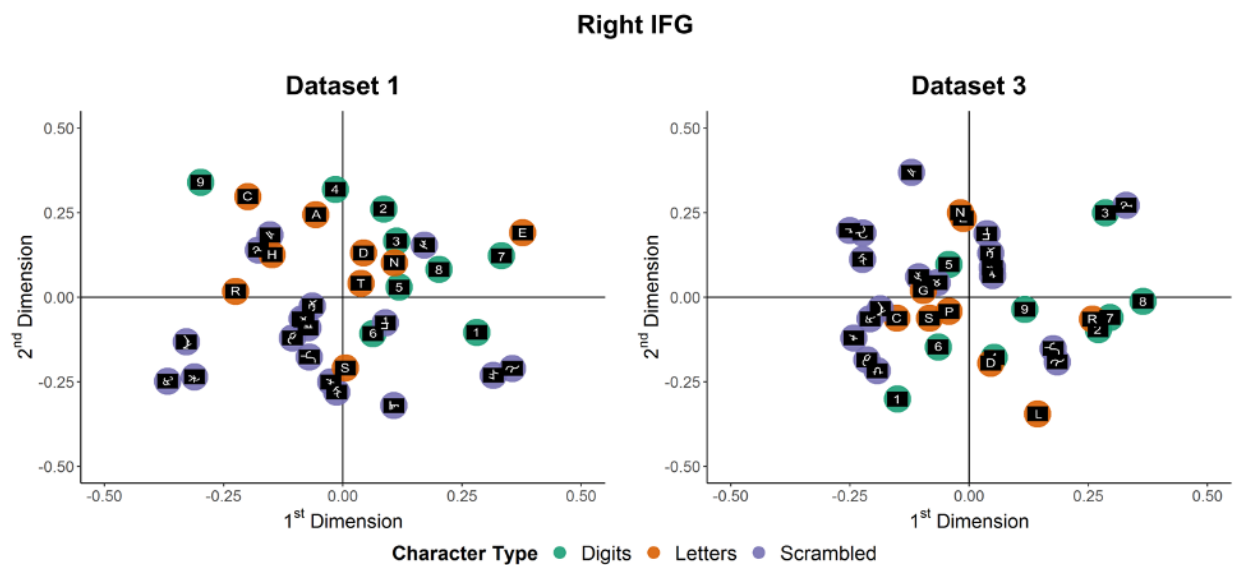


Figure A-12. Group-averaged representational geometry of the 36 exemplars in two-dimensional space in the numeral-preferring right inferior frontal gyrus (IFG) in Datasets 1 and 3. Three-dimensional interactive plots are available at <https://osf.io/jwgk8/wiki/home>.

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Supplementary Methods

Details of Stimuli

Table B-1. Stimulus list for a single run of digit detection and letter detection

Digit Detection		Letter Detection	
Digit Present	Digit Absent	Letter Present	Letter Absent
HN1DC	HNADC	782A6	78236
R1DTE	RNDTE	8A961	83961
TCS1D	TCSED	92A14	92714
CH2SD	CHTSD	5C473	59473
S2HNC	SRHNC	627C1	62713
TSN2R	TSNHR	73C69	73169
AH3NR	AHTNR	316D2	31692
EHT3C	EHTAC	5D296	53296
H3ERC	HTERC	89D23	89623
DST4A	DSTNA	2E738	29738
ER4CA	ERHCA	321E4	32174
R4DTA	RNDTA	73E69	73869
AET5D	AETHD	239H1	23971
D5CSA	DRCSA	8H537	86537
RC5HN	RCAHN	98H45	98345
C6SEH	CDSEH	153N2	15362
NRD6E	NRDAE	16N37	16437
TD6RH	TDERH	3N596	34596
CRN7E	CRNDE	639R7	63957
N7HRE	NSHRE	7R941	75941
RH7DA	RHCDA	93R78	93478
AE8DR	AETDR	51S93	51793
DHR8N	DHRCN	6S713	64713
H8CTA	HSCTA	759S4	75924
CR9AE	CRSAE	156T3	15623
E9HTR	EDHTR	4T768	41768
NCD9H	NCDEH	87T46	87946

Representational Geometries of Candidate Models

Figure B-1 – Figure B-4 illustrate the representational geometries based on phonological, ratio, joint frequency, and shape representations.

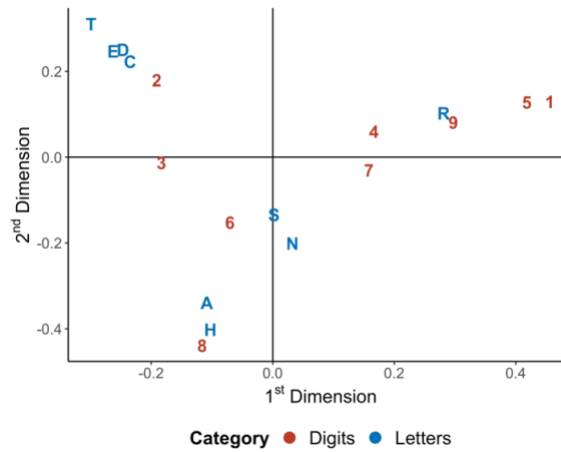


Figure B-1. Representational geometry of the digit and letter exemplars in two-dimensional space based on phonological similarity

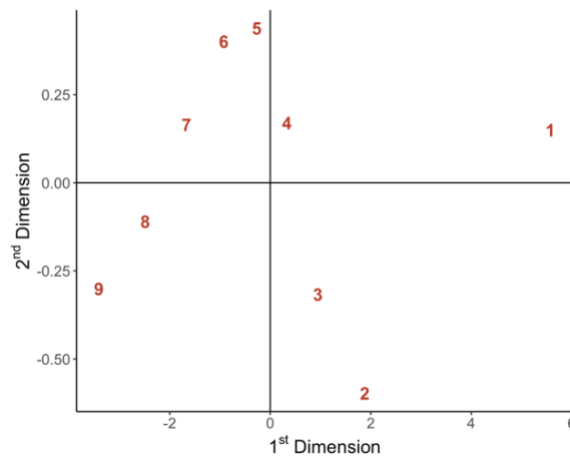


Figure B-2. Representational geometry of the digit exemplars in two-dimensional space based on ratio

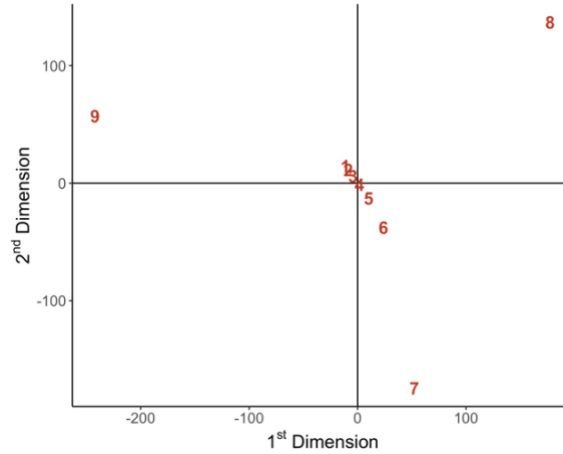


Figure B-3. Representational geometry of the digit exemplars in two-dimensional space based on joint frequency

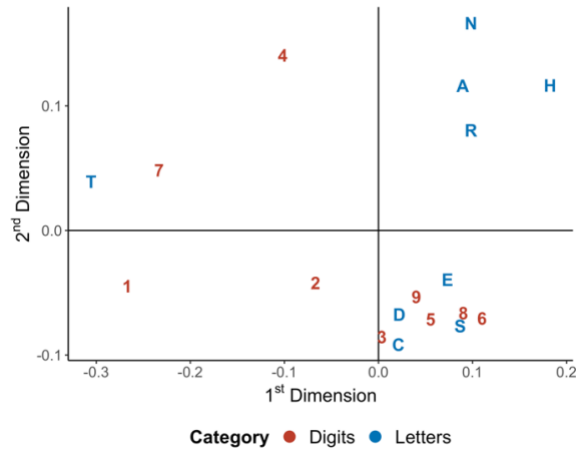


Figure B-4. Representational geometry of the digit and letter exemplars in two-dimensional space based on shape

Pairwise Similarity of Candidate Model RDMs

Table B-2. Spearman’s rank bivariate correlations of model RDMs (upper triangle: Digits only; lower triangle: Digits and letters combined)

Model	Phonological	Ratio	Frequency	Shape
Phonological	–	-.23	.16	.08
Ratio		–	-.65	.09
Frequency			–	-.12
Shape				–

Supplementary Results

Descriptive Statistics of Task Performance Measures

Table B-3. Mean, standard deviation, and range for in-scanner (included runs only) and behavioral measures

Measure	Mean	SD	Range
In-scanner detection tasks			
<u>Accuracy (%)</u>			
Digit Present	92.93	8.94	69.44 – 100
Digit Absent	92.42	11.16	53.70 – 100
Letter Present	92.40	11.53	49.38 – 100
Letter Absent	93.81	10.05	55.56 – 100
<u>Response time (ms)</u>			
Digit Present	746	82	630 – 937
Digit Absent	771	91	666 – 994
Letter Present	766	104	638 – 1079
Letter Absent	782	97	641 – 986
Standardized assessments			
Calculation	120	11.58	98 – 146
Math Fluency	114	14.36	87 – 149
Calculation Skills (Cluster)	121	12.56	93 – 145
Letter-Word ID	112	6.27	97 – 122

Note. Behavioral measures are expressed as standard scores. Accuracy rates reflect overall accuracy, including both errors of omission and commission ($N = 32$)

Regional Mean Letter Sensitivity

To assess the category-specific of our results pertaining to digit sensitivity, we replicated the analyses with the letter detection runs. On average, there was no hemispheric asymmetry in the regional mean letter sensitivity (i.e., [Letter Present – Letter Absent] contrast) (left: $M = 0.12$, $SD = 0.34$; right: $M = 0.10$, $SD = 0.36$, difference: $M = 0.02$, $SD = 0.35$), $t(31) = 0.35$, $d_z = 0.06$, $p = .728$, $BF_{01} = 5.00$ (**Figure B-5(a)**). The condition-wise response amplitudes was, however, strongly right-lateralized for both Letter Present (left: $M = 0.14$, $SD = 0.43$; right: $M = 0.79$, $SD = 0.67$, difference: $M = -0.65$, $SD = 0.64$) [$t(31) = -5.71$, $d_z = -1.01$, $p < .001$, $BF_{10} = 6689$] and Letter Absent (left: $M = 0.02$, $SD = 0.52$; right: $M = 0.69$, $SD = 0.76$, difference: $M = -0.67$, $SD = 0.66$) [$t(31) = -5.79$, $d_z = -1.02$, $p < .001$, $BF_{10} = 8277$] (**Figure B-5(c)**).

We found no conclusive or robust evidence that individuals with higher calculation skills had greater response amplitudes in both regions for Letter Present [left IT: $r(30) = .15$, $p = .203$, $BF_{0+} = 2.07$; $r_{skipped}(27) = -.34$, $p = .962$, $BF_{0+} = 11.22$; right IT: $r(30) = .24$, $p = .097$, $BF_{0+} = 1.13$; $r_{skipped}(29) = .30$, $p = .051$, $BF_{+0} = 1.50$] and Letter Absent [left IT: $r(30) = .11$, $p = .283$, $BF_{0+} = 2.73$; right IT: $r(30) = .22$, $p = .111$, $BF_{0+} = 1.26$; $r_{skipped}(26) = .58$, $p < .001$, $BF_{+0} = 73.22$].

Consistent with the whole-brain correlation analysis conducted originally by Pollack and Price (2019), greater letter sensitivity was not associated with higher calculation skills in both regions [left IT: $r(30) = .03$, $p = .426$, $BF_{0+} = 3.92$; right IT: $r(30) = -.03$, $p = .560$, $BF_{0+} = 5.09$]. However, the evidence was less conclusive in the right IT after exclusion of an outlier [$r_{skipped}(29) = .19$, $p = .150$, $BF_{0+} = 1.59$].

There was also evidence of no relation between calculation skills and degree of lateralization in their mean letter sensitivity [$r(30) = .06$, $p = .736$, $BF_{01} = 4.31$] (**Figure B-5(b)**)

and Letter Present response amplitudes [$r(30) = -.14, p = .434, BF_{01} = 3.39$]. Evidence of any relation for Letter Absent response amplitudes was inconclusive [$r(30) = -.17, p = .342, BF_{01} = 2.95; r_{skipped}(29) = -.24, p = .199, BF_{01} = 2.03$] (**Figure B-5(d)**).

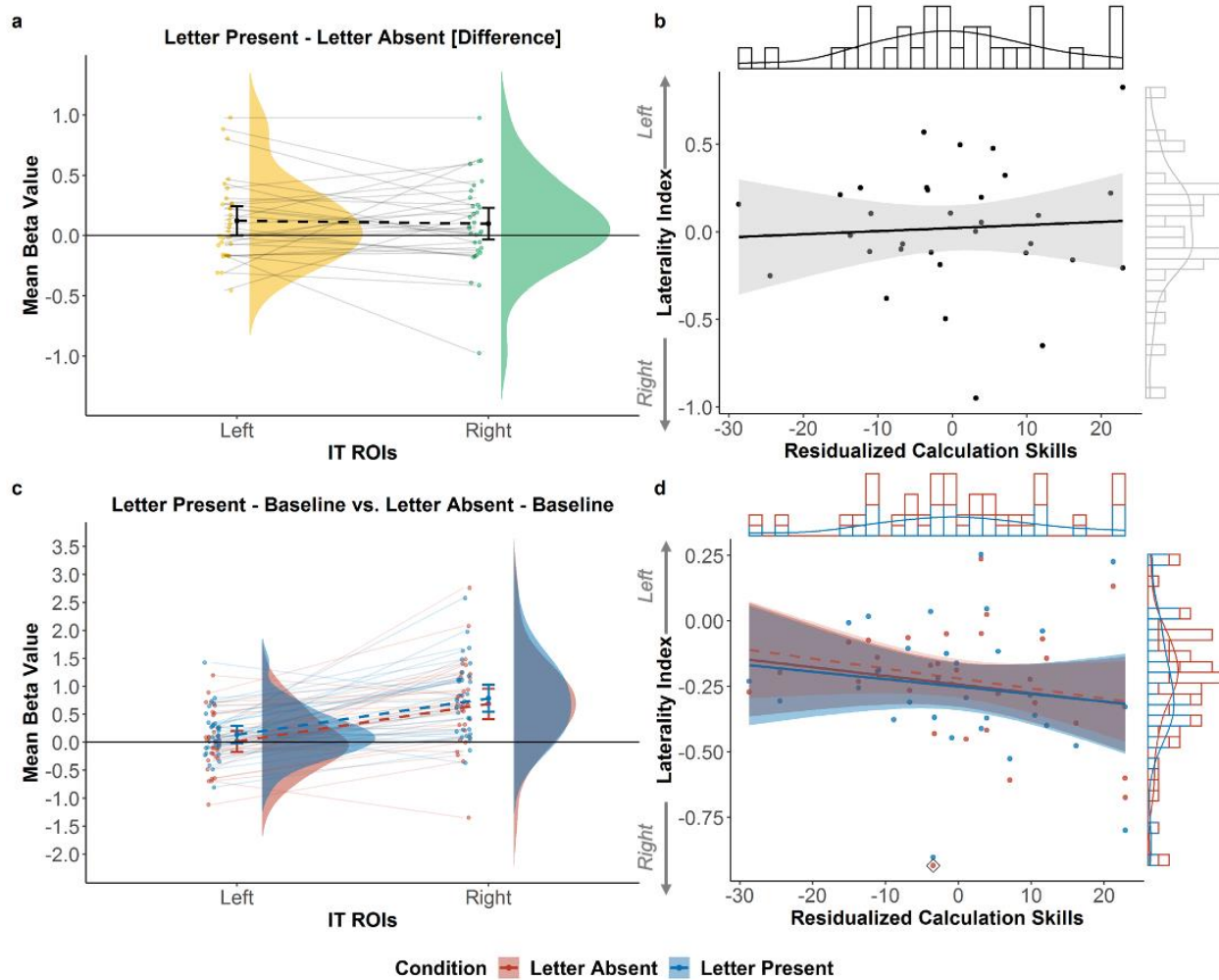


Figure B-5. Hemispheric asymmetries of regional mean response amplitudes and their relation to calculation skills for letter detection (a, b) letter sensitivity ([Letter Present – Letter Absent] contrast), and (c, d) condition-wise ([Letter Absent – Fixation] and [Letter Present – Fixation]). Error bars and bands are 95% confidence intervals. Dashed regression lines exclude bivariate outliers enclosed in \diamond .

Validation of Assumption That Co-Occurring Characters from the Non-Target Category Are Not Reliably Represented

As shown in **Table B-1**, some character(s) from the non-target category always co-occur with characters from the target category. For instance, the Digit Present trials with target ‘1’ always contain the letter ‘D’. Hence, the activation pattern that we assumed to be a reliable neural representation of the detected target ‘1’ likely also contain a reliable bottom-up representation of the non-target ‘D’ even if it was not the focus of detection. If that were true, the neural representation of detected target ‘1’ in the Digit Present Trials would be more similar to a detected target ‘D’ than to other detected letters in the Letter Present Trials due to the co-occurrence. To examine this hypothesis, for all affected pairs of co-occurring characters, we examined within each ROI whether the representational similarity (RS) between co-occurring characters (e.g., Digit Present trials with target ‘1’ that also contains non-target ‘D’, and Letter Present trials with target ‘D’) was greater than the mean RS between non-co-occurring characters (e.g., Digit Present trials with target ‘1’ and Letter Present trials with all other target letters but ‘D’). Specifically, we tested whether the difference in the Fisher’s z -transformed RS values (i.e., RS_z of co-occurring characters – Mean RS_z of non-co-occurring characters) was statistically greater than zero (one-sample t -test, right-tailed; Bayesian prior: Cauchy distribution, scale = 0.707).

Table B-4 below shows that in all cases, there was no evidence that a co-occurring character from the non-target category elicited a neural representation that was more similar to that when the said character was actually the target, relative to other target characters from the same category. Hence, our assumption that co-occurring characters from the non-target category are not as reliably represented as when they were from the target category is valid.

Table B-4. Fisher’s z -transformed representational similarity (RS_z) between co-occurring characters from target and non-target categories

Target character	Co-occurring non-target(s)	Left IT ROI	Right IT ROI
		RS_z of co-occurring characters – Mean RS_z of non-co-occurring characters	RS_z of co-occurring characters – Mean RS_z of non-co-occurring characters
1	D	$M = -0.04, t(31) = -1.26, d_z = -0.22, p = .892, BF_{0+} = 11.02$	$M = 0.03, t(31) = 0.88, d_z = 0.16, p = .192, BF_{0+} = 2.32$
2	S	$M = -0.07, t(31) = -1.81, d_z = -0.32, p = .960, BF_{0+} = 13.61$	$M = -0.02, t(31) = -0.35, d_z = -0.06, p = .637, BF_{0+} = 6.80$
3	H	$M = -0.001, t(31) = -0.04, d_z = -0.01, p = .515, BF_{0+} = 5.45$	$M = -0.05, t(31) = -1.19, d_z = -0.21, p = .879, BF_{0+} = 10.68$
4	A	$M = -0.03, t(31) = -0.78, d_z = -0.14, p = .780, BF_{0+} = 8.75$	$M = 0.009, t(31) = 0.19, d_z = 0.03, p = .424, BF_{0+} = 4.54$
7	R	$M = 0.01, t(31) = 0.37, d_z = 0.07, p = .355, BF_{0+} = 3.88$	$M = -0.009, t(31) = -0.23, d_z = -0.04, p = .592, BF_{0+} = 6.27$
C	7	$M = -0.04, t(31) = -0.99, d_z = -0.18, p = .834, BF_{0+} = 9.71$	$M = -0.01, t(31) = -0.33, d_z = -0.06, p = .628, BF_{0+} = 6.69$
D	2	$M = 0.02, t(31) = 0.58, d_z = 0.10, p = .283, BF_{0+} = 3.19$	$M = 0.05, t(31) = 1.50, d_z = 0.27, p = .072, BF_{0+} = 1.04$
E	3	$M = -0.02, t(31) = -0.55, d_z = -0.10, p = .708, BF_{0+} = 7.69$	$M = 0.07, t(31) = 1.40, d_z = 0.25, p = .085, BF_{0+} = 1.20$
N	3	$M = -0.03, t(31) = -0.97, d_z = -0.17, p = .829, BF_{0+} = 9.61$	$M = 0.002, t(31) = 0.06, d_z = 0.01, p = .478, BF_{0+} = 5.07$
R	7 & 9 ¹	$M = -0.05, t(31) = -1.66, d_z = -0.29, p = .946, BF_{0+} = 12.88$	$M = -0.04, t(31) = -1.35, d_z = -0.24, p = .906, BF_{0+} = 11.43$
T	6	$M = -0.007, t(31) = -0.18, d_z = -0.03, p = .570, BF_{0+} = 6.03$	$M = -0.08, t(31) = -1.86, d_z = -0.33, p = .964, BF_{0+} = 13.82$

Note. ¹ RS_z of R and 7, and RS_z of R and 9 were averaged as mean RS_z of co-occurring characters.

Letter Discriminability

There was no letter discriminability in the left IT ($M = -0.006$, $SD = 0.11$) [$t(29) = -0.30$, $d = -0.05$, $p = .615$, $BF_{0+} = 6.35$] and right IT ($M = -0.008$, $SD = 0.12$) [$t(29) = -0.38$, $d = -0.07$, $p = .645$, $BF_{0+} = 6.69$] (**Figure 4-5(b)**). There was no hemispheric asymmetry in letter discriminability (difference: $M = 0.002$, $SD = 0.18$), $t(29) = 0.08$, $d_z = 0.01$, $p = .940$, $BF_{01} = 5.13$.

Greater letter discriminability was not associated with higher calculation skills in the left IT [$r(28) = -.22$, $p = .881$, $BF_{0+} = 8.96$] and right IT [$r(28) = -.003$, $p = .507$, $BF_{0+} = 4.46$] (**Figure B-6**). There was also no relation between the degree of lateralization of letter discriminability and calculation skills, $r(28) = -.13$, $p = .479$, $BF_{01} = 3.47$.

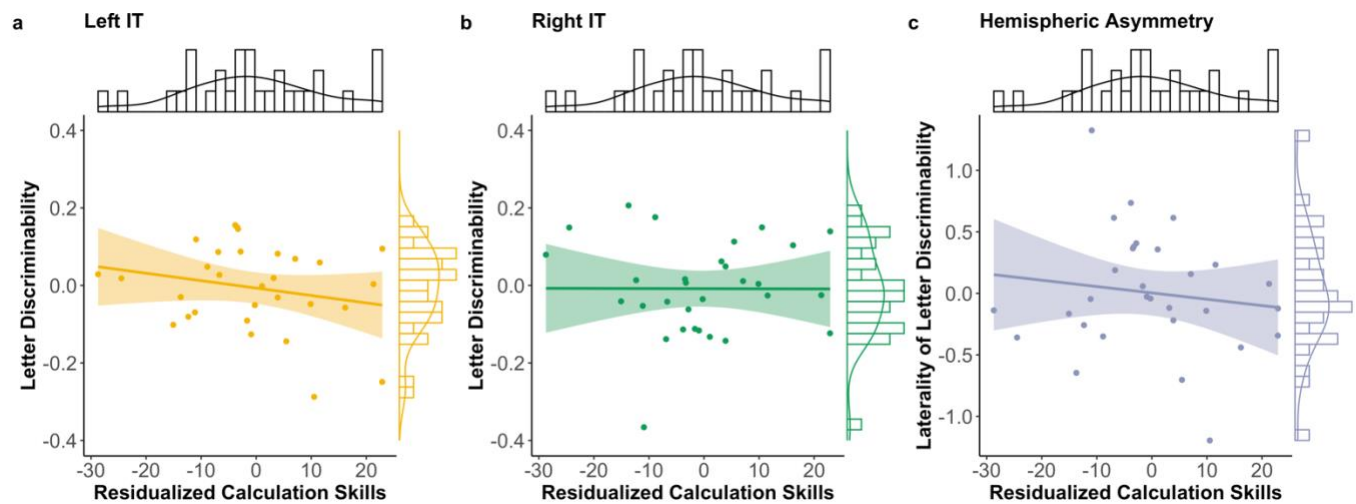


Figure B-6. Relation between calculation skills and letter discriminability (a) in the left IT, (b) in the right IT, and (c) its lateralization (negative: left lateralization, positive: right lateralization) ($N = 30$). Error bands are 95% confidence intervals.

Hemispheric Asymmetry of Representational Geometries of Letters

On average, there was a small, but significant positive correlation between the representational geometries of letters (i.e., Letters-RDMs) in the left and right IT (mean $\rho_z = .12$, $SD = .25$), $t(30) = 2.64$, $d_z = 0.47$, $p = .006$, $BF_{+0} = 7.08$. There was inconclusive evidence that lower between-hemisphere similarity (i.e., greater asymmetry) in the representational geometries of letters was associated with higher calculation skills, $r(29) = -.12$, $p = .269$, $BF_{0-} = 2.57$.

Representational Content in Digits-RDMs

Table B-5. Pairwise difference in the similarity between the left IT Digits-RDMs and model RDMs ($N = 32$)

Model 1	vs.	Model 2	t	d_z	p	BF_{10}
Phonological	-	Ratio	-0.02	-0.004	.983	0.19
Phonological	-	Frequency	-1.80	-0.32	.082	0.79
Phonological	-	Shape	0.69	0.12	.498	0.24
Ratio	-	Frequency	-1.17	-0.21	.253	0.35
Ratio	-	Shape	0.77	0.14	.447	0.25
Frequency	-	Shape	2.02	0.36	.052	1.13

Note. p -values are uncorrected for multiple comparisons.

Table B-6. Pairwise difference in the similarity between the right IT Digits-RDMs and model RDMs ($N = 32$)

Model 1	vs.	Model 2	t	d_z	p	BF_{10}
Phonological	-	Ratio	-1.84	-0.32	.076	0.84
Phonological	-	Frequency	0.67	0.12	.511	0.23
Phonological	-	Shape	0.31	0.06	.756	0.20
Ratio	-	Frequency	1.58	0.28	.124	0.58
Ratio	-	Shape	2.61	0.46	.014	3.36
Frequency	-	Shape	-0.35	-0.06	.726	0.20

Note. p -values are uncorrected for multiple comparisons.

Representational Content in Full-RDMs (Alphanumeric Set)

Left IT

The left IT Full-RDMs were not similar to the RDMs of the Phonological model (Mean $\rho_z = .0008$, $SD = .08$) [$t(30) = 0.50$, $d = 0.01$, $p = .479$, $BF_{0+} = 5.01$] and Shape model (Mean $\rho_z = -.03$, $SD = .12$) [$t(30) = -1.12$, $d = -0.20$, $p = .865$, $BF_{0+} = 10.20$] (**Figure B-7**).

In terms of pairwise model comparisons, there was no within-participant difference between the pair of neural RDM-model RDM similarities, $t(30) = 1.01$, $d_z = 0.18$, $p = .323$, $BF_{01} = 3.29$.

Right IT

The right IT Full-RDMs were not similar to the RDMs of the Phonological model (Mean $\rho_z = .002$, $SD = .06$) [$t(30) = 0.23$, $d = 0.04$, $p = .410$, $BF_{0+} = 4.34$] and Shape model (Mean $\rho_z = -.01$, $SD = .06$) [$t(30) = -0.96$, $d = -0.17$, $p = .827$, $BF_{0+} = 9.43$] (**Figure B-7**).

In terms of pairwise model comparisons, there was no within-participant difference between the pair of neural RDM-model RDM similarities, $t(30) = 0.91$, $d_z = 0.16$, $p = .369$, $BF_{01} = 3.56$.

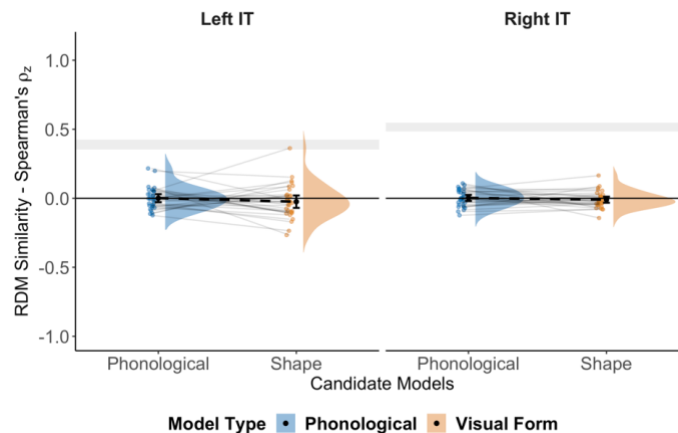


Figure B-7. Similarity between the alphanumeric model RDMs and Full-RDMs in the left and right IT ROIs ($N = 31$)
Error bars are 95% confidence intervals. Grey bars indicate the estimated upper and lower bounds of the expected similarity achievable by the unknown true model given the degree of between-participant variability.

Left IT vs. Right IT

Neural RDM-model RDM similarities were not different between the left and right IT for both Phonological model [difference: $M = -0.002$, $SD = 0.10$, $t(30) = -0.10$, $d_z = -0.02$, $p = .924$, $BF_{01} = 5.20$] and Shape model [difference: $M = -0.01$, $SD = 0.14$, $t(30) = -0.59$, $d_z = -0.11$, $p = .560$, $BF_{01} = 4.45$] (**Figure B-8**).

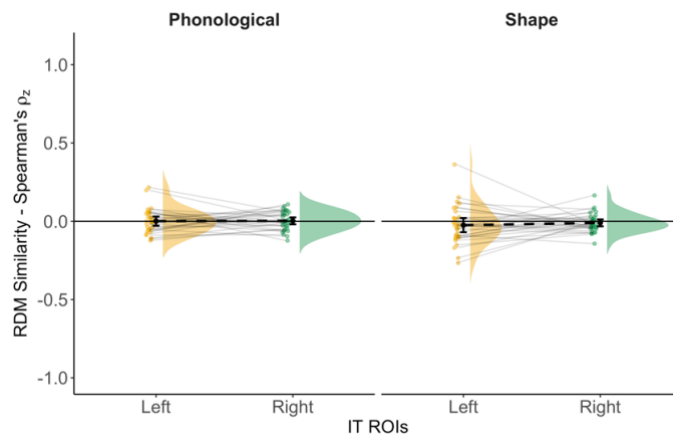


Figure B-8. Hemispheric asymmetry in similarity between the alphanumeric model RDMs and neural Full-RDMs ($N = 31$)
Error bars are 95% confidence intervals.

REFERENCES

- Abboud, S., Maidenbaum, S., Dehaene, S., & Amedi, A. (2015). A number-form area in the blind. *Nature Communications*, 6(1), 6026. <https://doi.org/10.1038/ncomms7026>
- Abdulrahman, H., & Henson, R. N. (2016). Effect of trial-to-trial variability on optimal event-related fMRI design: Implications for Beta-series correlation and multi-voxel pattern analysis. *NeuroImage*, 125, 756–766. <https://doi.org/10.1016/j.neuroimage.2015.11.009>
- Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human Extrastriate Visual Cortex and the Perception of Faces, Words, Numbers, and Colors. *Cerebral Cortex*, 4(5), 544–554. <https://doi.org/10.1093/cercor/4.5.544>
- Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, 113(18), 4909–4917. <https://doi.org/10.1073/pnas.1603205113>
- Amalric, M., & Dehaene, S. (2018). Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain’s semantic networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1740), 20160515. <https://doi.org/10.1098/rstb.2016.0515>
- Amalric, M., & Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage*, 189, 19–31. <https://doi.org/10.1016/j.neuroimage.2019.01.001>
- Anderson, B., Soliman, S., O’Malley, S., Danckert, J., & Besner, D. (2015). Control over the strength of connections between modules: a double dissociation between stimulus format and task revealed by Granger causality mapping in fMRI. *Frontiers in Psychology*, 6, 1–10. <https://doi.org/10.3389/fpsyg.2015.00321>
- Anderson, E. J., Mannan, S. K., Husain, M., Rees, G., Sumner, P., Mort, D. J., McRobbie, D., & Kennard, C. (2007). Involvement of prefrontal cortex in visual search. *Experimental Brain Research*, 180(2), 289–302. <https://doi.org/10.1007/s00221-007-0860-0>
- Andres, M., Michaux, N., & Pesenti, M. (2012). Common substrate for mental arithmetic and finger representation in the parietal cortex. *NeuroImage*, 62(3), 1520–1528. <https://doi.org/10.1016/j.neuroimage.2012.05.047>
- Andres, M., Pelgrims, B., Michaux, N., Olivier, E., & Pesenti, M. (2011). Role of distinct parietal areas in arithmetic: An fMRI-guided TMS study. *NeuroImage*, 54(4), 3048–3056. <https://doi.org/10.1016/j.neuroimage.2010.11.009>
- Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, 9(4), 278–291. <https://doi.org/10.1038/nrn2334>
- Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *NeuroReport*, 16(16), 1769–1773. <https://doi.org/10.1097/01.wnr.0000183905.23396.f1>
- Arsalidou, M., & Taylor, M. J. (2011). Is $2+2=4$? Meta-analyses of brain areas needed for

- numbers and calculations. *NeuroImage*, 54(3), 2382–2393.
<https://doi.org/10.1016/j.neuroimage.2010.10.009>
- Attout, L., Fias, W., Salmon, E., & Majerus, S. (2014). Common Neural Substrates for Ordinal Representation in Short-Term Memory, Numerical and Alphabetical Cognition. *PLoS ONE*, 9(3), e92049. <https://doi.org/10.1371/journal.pone.0092049>
- Aurtenetxe, S., Molinaro, N., Davidson, D., & Carreiras, M. (2020). Early dissociation of numbers and letters in the human brain. *Cortex*, 130(May), 192–202.
<https://doi.org/10.1016/j.cortex.2020.03.030>
- Baek, S., Daitch, A. L., Pinheiro-Chagas, P., & Parvizi, J. (2018). Neuronal Population Responses in the Human Ventral Temporal and Lateral Parietal Cortex during Arithmetic Processing with Digits and Number Words. *Journal of Cognitive Neuroscience*, 30(9), 1315–1322. https://doi.org/10.1162/jocn_a_01296
- Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, 104(21), 9087–9092.
<https://doi.org/10.1073/pnas.0703300104>
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 449–454. <https://doi.org/10.1073/pnas.0507062103>
- Basso, G., Nichelli, P., Wharton, C. M., Peterson, M., & Grafman, J. (2003). Distributed neural systems for temporal production: A functional MRI study. *Brain Research Bulletin*, 59(5), 405–411. [https://doi.org/10.1016/S0361-9230\(02\)00941-3](https://doi.org/10.1016/S0361-9230(02)00941-3)
- Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences*, 17(5), 210–219.
<https://doi.org/10.1016/j.tics.2013.03.007>
- Behrmann, M., & Plaut, D. C. (2015). A vision of graded hemispheric specialization. *Annals of the New York Academy of Sciences*, 1359(1), 30–46. <https://doi.org/10.1111/nyas.12833>
- Behrmann, M., & Plaut, D. C. (2020). Hemispheric Organization for Visual Object Recognition: A Theoretical Account and Empirical Evidence. *Perception*, 49(4), 373–404.
<https://doi.org/10.1177/0301006619899049>
- Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 509–522.
<https://doi.org/10.1109/34.993558>
- Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289–300.
- Bolger, D. J., Perfetti, C. A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*, 25(1), 92–104.

<https://doi.org/10.1002/hbm.20124>

- Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*, *143*, 128–140. <https://doi.org/10.1016/j.neuroimage.2016.08.068>
- Bouhali, F., Mongelli, V., & Cohen, L. (2017). Musical literacy shifts asymmetries in the ventral visual cortex. *NeuroImage*, *156*, 445–455. <https://doi.org/10.1016/j.neuroimage.2017.04.027>
- Bouhali, F., Mongelli, V., Thiebaut de Schotten, M., & Cohen, L. (2020). Reading music and words: The anatomical connectivity of musicians' visual cortex. *NeuroImage*, *212*(November 2019), 116666. <https://doi.org/10.1016/j.neuroimage.2020.116666>
- Bracci, S., Caramazza, A., & Peelen, M. V. (2015). Representational Similarity of Body Parts in Human Occipitotemporal Cortex. *Journal of Neuroscience*, *35*(38), 12977–12985. <https://doi.org/10.1523/JNEUROSCI.4698-14.2015>
- Bracci, S., Daniels, N., & Op de Beeck, H. (2017). Task Context Overrides Object- and Category-Related Representational Content in the Human Parietal Cortex. *Cerebral Cortex*, *27*(1), 310–321. <https://doi.org/10.1093/cercor/bhw419>
- Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *Journal of Neuroscience*, *36*(2), 432–444. <https://doi.org/10.1523/JNEUROSCI.2314-15.2016>
- Bracci, S., Ritchie, J. B., & Op de Beeck, H. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia*, *105*, 153–164. <https://doi.org/10.1016/j.neuropsychologia.2017.06.010>
- Bradshaw, A. R., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Methodological considerations in assessment of language lateralisation with fMRI: a systematic review. *PeerJ*, *5*(7), e3557. <https://doi.org/10.7717/peerj.3557>
- Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. <https://doi.org/10.1038/nrn756>
- Brown, W. (1910). Some Experimental Results in the Correlation of Mental Abilities. *British Journal of Psychology*, *1904-1920*, *3*(3), 296–322. <https://doi.org/10.1111/j.2044-8295.1910.tb00207.x>
- Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1), 1–38. <https://doi.org/10.5334/joc.72>
- Bugden, S., Woldorff, M. G., & Brannon, E. M. (2019). Shared and distinct neural circuitry for nonsymbolic and symbolic double-digit addition. *Human Brain Mapping*, *40*(4), 1328–1343. <https://doi.org/10.1002/hbm.24452>
- Bynner, J., & Parsons, S. (1997). *Does Numeracy Matter? Evidence from the National Child Development Study on the Impact of Poor Numeracy on Adult Life*. The Basic Skills Agency.
- Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of

- numerical processing in adults and 4-y-old children. *PLoS Biology*, 4(5), 844–854.
<https://doi.org/10.1371/journal.pbio.0040125>
- Cantlon, J. F., Libertus, M. E., Pinel, P., Dehaene, S., Brannon, E. M., & Pelphrey, K. A. (2009). The Neural Development of an Abstract Concept of Number. *Journal of Cognitive Neuroscience*, 21(11), 2217–2229. <https://doi.org/10.1162/jocn.2008.21159>
- Cantlon, J. F., Pinel, P., Dehaene, S., & Pelphrey, K. A. (2011). Cortical Representations of Symbols, Objects, and Faces Are Pruned Back during Early Childhood. *Cerebral Cortex*, 21(1), 191–199. <https://doi.org/10.1093/cercor/bhq078>
- Cappelletti, M., Lee, H. L., Freeman, E. D., & Price, C. J. (2010). The Role of Right and Left Parietal Lobes in the Conceptual Processing of Numbers. *Journal of Cognitive Neuroscience*, 22(2), 331–346. <https://doi.org/10.1162/jocn.2009.21246>
- Carlos, B. J., Hirshorn, E. A., Durisko, C., Fiez, J. A., & Coutanche, M. N. (2019). Word inversion sensitivity as a marker of visual word form area lateralization: An application of a novel multivariate measure of laterality. *NeuroImage*, 191(August 2018), 493–502. <https://doi.org/10.1016/j.neuroimage.2019.02.044>
- Carreiras, M., Monahan, P. J., Lizarazu, M., Duñabeitia, J. A., & Molinaro, N. (2015). Numbers are not like words: Different pathways for literacy and numeracy. *NeuroImage*, 118, 79–89. <https://doi.org/10.1016/j.neuroimage.2015.06.021>
- Carreiras, M., Quiñones, I., Hernández-Cabrera, J. A., & Duñabeitia, J. A. (2015). Orthographic Coding: Brain Activation for Letters, Symbols, and Digits. *Cerebral Cortex*, 25(12), 4748–4760. <https://doi.org/10.1093/cercor/bhu163>
- Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., & Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *eLife*, 8, 1–26. <https://doi.org/10.7554/eLife.45160>
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, 129(3), 564–583. <https://doi.org/10.1093/brain/awl004>
- Chassy, P., & Grodd, W. (2012). Comparison of quantities: Core and format-dependent regions as revealed by fMRI. *Cerebral Cortex*, 22(6), 1420–1430. <https://doi.org/http://dx.doi.org/10.1093/cercor/bhr219>
- Chochon, F., Cohen, L., Moortele, P. F. van de, & Dehaene, S. (1999). Differential Contributions of the Left and Right Inferior Parietal Lobules to Number Processing. *Journal of Cognitive Neuroscience*, 11(6), 617–630. <https://doi.org/10.1162/089892999563689>
- Chouinard, P. A., Morrissey, B. F., Köhler, S., & Goodale, M. A. (2008). Repetition suppression in occipital-temporal visual areas is modulated by physical rather than semantic features of objects. *NeuroImage*, 41(1), 130–144. <https://doi.org/10.1016/j.neuroimage.2008.02.011>
- Chrisomalis, S. (2010). *Numerical Notation: A Comparative History* (1st ed.). Cambridge University Press.
- Church, J. A., Balota, D. A., Petersen, S. S. E., & Schlaggar, B. L. (2011). Manipulation of Length and Lexicality Localizes the Functional Neuroanatomy of Phonological Processing in Adult Readers. *Journal of Cognitive Neuroscience*, 23(6), 1475–1493. <https://doi.org/10.1162/jocn.2010.21515>

- Coderre, E. L., Filippi, C. G., Newhouse, P. A., & Dumas, J. A. (2009). Ichi, ni, 3, 4: Neural representation of kana, kanji, and Arabic numbers in native Japanese speakers. *Brain and Cognition*, *70*(3), 289–296. <https://doi.org/10.1016/j.bandc.2009.03.002>
- Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review*, *16*(2), 332–336. <https://doi.org/10.3758/PBR.16.2.332>
- Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-Dependent and -Independent Representations of Numbers in the Parietal Lobes. *Neuron*, *53*(2), 307–314. <https://doi.org/10.1016/j.neuron.2006.12.025>
- Cohen Kadosh, R., Cohen Kadosh, K., Schuhmann, T., Kaas, A., Goebel, R., Henik, A., & Sack, A. T. (2007). Virtual Dyscalculia Induced by Parietal-Lobe TMS Impairs Automatic Magnitude Processing. *Current Biology*, *17*(8), 689–693. <https://doi.org/10.1016/j.cub.2007.02.056>
- Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132–147. <https://doi.org/10.1016/j.pneurobio.2007.11.001>
- Cohen, L., & Dehaene, S. (1991). Neglect Dyslexia for Numbers? A Case Report. *Cognitive Neuropsychology*, *8*(1), 39–58. <https://doi.org/10.1080/02643299108253366>
- Cohen, L., & Dehaene, S. (1995). Number processing in pure alexia: The effect of hemispheric asymmetries and task demands. *Neurocase*, *1*(2), 121–137. <https://doi.org/10.1080/13554799508402356>
- Cohen, L., & Dehaene, S. (1996). Cerebral networks for number processing: Evidence from a case of posterior callosal lesion. *Neurocase*, *2*(November), 155–174. <https://doi.org/10.1080/13554799608402394>
- Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology*, *17*(6), 563–583. <https://doi.org/10.1080/02643290050110656>
- Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage*, *22*(1), 466–476. <https://doi.org/10.1016/j.neuroimage.2003.12.049>
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*(2), 291–307. <https://doi.org/10.1093/brain/123.2.291>
- Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054–1069. <https://doi.org/10.1093/brain/awf094>
- Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic Jargon Sparring Numbers: A Category-specific Phonological Impairment. *Cognitive Neuropsychology*, *14*(7), 1029–1061. <https://doi.org/10.1080/026432997381349>
- Cohen, M. A., Dilks, D. D., Koldewyn, K., Weigelt, S., Feather, J., Kell, A. J., Keil, B., Fischl,

- B., Zöllei, L., Wald, L., Saxe, R., & Kanwisher, N. (2019). Representational similarity precedes category selectivity in the developing ventral visual pathway. *NeuroImage*, *197*, 565–574. <https://doi.org/10.1016/j.neuroimage.2019.05.010>
- Colvin, M. K., Funnell, M. G., & Gazzaniga, M. S. (2005). Numerical processing in the two hemispheres: Studies of a split-brain patient. *Brain and Cognition*, *57*(1), 43–52. <https://doi.org/http://dx.doi.org/10.1016/j.bandc.2004.08.019>
- Conrad, B. N., Wilkey, E. D., Yeo, D. J., & Price, G. R. (2020). Network topology of symbolic and nonsymbolic number comparison. *Network Neuroscience*, *4*(3), 714–745. https://doi.org/10.1162/netn_a_00144
- Corballis, M. C. (1994). Can commissurotomed subjects compare digits between the visual fields? *Neuropsychologia*, *32*(12), 1475–1486. [https://doi.org/10.1016/0028-3932\(94\)90119-8](https://doi.org/10.1016/0028-3932(94)90119-8)
- Cui, J., Yu, X., Yang, H., Chen, C., Liang, P., & Zhou, X. (2013). Neural correlates of quantity processing of numeral classifiers. *Neuropsychology*, *27*(5), 583–594. <https://doi.org/http://dx.doi.org/10.1037/a0033630>
- Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. <https://doi.org/10.1038/nn.3381>
- Cummine, J., Chouinard, B., Szepesvari, E., & Georgiou, G. K. K. G. K. (2015). An examination of the rapid automatized naming–reading relationship using functional magnetic resonance imaging. *Neuroscience*, *305*, 49–66. <https://doi.org/10.1016/j.neuroscience.2015.07.071>
- Cummine, J., Szepesvari, E., Chouinard, B., Hanif, W., & Georgiou, G. K. (2014). A functional investigation of RAN letters, digits, and objects: How similar are they? *Behavioural Brain Research*, *275*, 157–165. <https://doi.org/http://dx.doi.org/10.1016/j.bbr.2014.08.038>
- Daitch, A. L., Foster, B. L., Schrouff, J., Rangarajan, V., Kaşikçi, I., Gattas, S., & Parvizi, J. (2016). Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. *Proceedings of the National Academy of Sciences*, *113*(46), E7277–E7286. <https://doi.org/10.1073/pnas.1608434113>
- Damarla, S. R., Cherkassky, V. L., & Just, M. A. (2016). Modality-independent representations of small quantities based on brain activation patterns. *Human Brain Mapping*, *37*(4), 1296–1307. <https://doi.org/10.1002/hbm.23102>
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biology*, *16*(3), e2004103. <https://doi.org/10.1371/journal.pbio.2004103>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)
- Dehaene, S. (1996). The Organization of Brain Activations in Number Comparison: Event-Related Potentials and the Additive-Factors Method. *Journal of Cognitive Neuroscience*, *8*(1), 47–68. <https://doi.org/10.1162/jocn.1996.8.1.47>
- Dehaene, S., & Akhavein, R. (1995). Attention, Automaticity, and Levels of Representation in Number Processing. In *Journal of Experimental Psychology: Learning, Memory, and*

- Cognition* (Vol. 21, Issue 2, pp. 314–326). <https://doi.org/10.1037/0278-7393.21.2.314>
- Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: a neuronal model. *Journal of Cognitive Neuroscience*, 5(4), 390–407. <https://doi.org/10.1162/jocn.1993.5.4.390>
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1(1), 83–120.
- Dehaene, S., & Cohen, L. (1997). Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal and Quantitative Knowledge of Arithmetic. *Cortex*, 33(2), 219–250. [https://doi.org/10.1016/S0010-9452\(08\)70002-9](https://doi.org/10.1016/S0010-9452(08)70002-9)
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398. <https://doi.org/10.1016/j.neuron.2007.10.004>
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, 16(4), 234–244. <https://doi.org/10.1038/nrn3924>
- Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J.-B., Le Bihan, D., & Cohen, L. (2004). Letter Binding and Invariant Recognition of Masked Words. Behavioral and Neuroimaging Evidence. *Psychological Science*, 15(5), 307–313. <https://doi.org/10.1111/j.0956-7976.2004.00674.x>
- Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43(1), 1–29. [https://doi.org/10.1016/0010-0277\(92\)90030-L](https://doi.org/10.1016/0010-0277(92)90030-L)
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. Le, Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4(7), 752–758. <https://doi.org/10.1038/89551>
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359–1364. <https://doi.org/10.1126/science.1194140>
- Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20(3–6), 487–506. <https://doi.org/http://dx.doi.org/10.1080/02643290244000239>
- Dickson, D. S., & Federmeier, K. D. (2017). The language of arithmetic across the hemispheres: An event-related potential investigation. *Brain Research*, 1662, 46–56. <https://doi.org/10.1016/j.brainres.2017.02.019>
- Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE*, 10(4), e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. In

- PLoS Computational Biology* (Vol. 13, Issue 4).
<https://doi.org/10.1371/journal.pcbi.1005508>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5(July), 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78–89. <https://doi.org/10.1016/j.jmp.2015.10.003>
- Dienes, Z., & Mclatchie, N. (2017). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin and Review*, 1–12. <https://doi.org/10.3758/s13423-017-1266-z>
- Diester, I., & Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biology*, 5(11), 2684–2695.
<https://doi.org/10.1371/journal.pbio.0050294>
- Dotan, D., & Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*, 63, 317–351. <https://doi.org/10.1016/j.cortex.2014.08.014>
- Dotan, D., & Friedmann, N. (2018). A cognitive model for multidigit number reading: Inferences from individuals with selective impairments. *Cortex*, 101, 249–281.
<https://doi.org/10.1016/j.cortex.2017.10.025>
- Dotan, D., & Friedmann, N. (2019). Separate mechanisms for number reading and word reading: Evidence from selective impairments. *Cortex*, 114, 176–192.
<https://doi.org/10.1016/j.cortex.2018.05.010>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>; [10.1037/0012-1649.43.6.1428.supp](https://doi.org/10.1037/0012-1649.43.6.1428.supp) (Supplemental)
- Edwards, L. A., Wagner, J. B., Simon, C. E., & Hyde, D. C. (2016). Functional brain organization for number processing in pre-verbal infants. *Developmental Science*, 19(5), 757–769. <https://doi.org/10.1111/desc.12333>
- Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., & Kleinschmidt, A. (2003). A Supramodal Number Representation in Human Intraparietal Cortex. *Neuron*, 37(4), 719–726.
[https://doi.org/10.1016/S0896-6273\(03\)00036-9](https://doi.org/10.1016/S0896-6273(03)00036-9)
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, 59(3), 2349–2361.
<https://doi.org/10.1016/j.neuroimage.2011.09.017>
- Eickhoff, S. B., Laird, A. R., Fox, P. M., Lancaster, J. L., & Fox, P. T. (2017). Implementation errors in the GingerALE Software: Description and recommendations. *Human Brain Mapping*, 38(1), 7–11. <https://doi.org/10.1002/hbm.23342>
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain*

- Mapping*, 30(9), 2907–2926. <https://doi.org/10.1002/hbm.20718>
- Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., Bzdok, D., & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, 137, 70–85. <https://doi.org/10.1016/j.neuroimage.2016.04.072>
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325–1335. <https://doi.org/10.1016/j.neuroimage.2004.12.034>
- Emerson, R. W., & Cantlon, J. F. (2015). Continuity and change in children’s longitudinal neural responses to numbers. *Developmental Science*, 18(2), 314–326. <https://doi.org/10.1111/desc.12215>
- Fairhall, S. L., Anzellotti, S., Pajtas, P. E., & Caramazza, A. (2011). Concordance between perceptual and categorical repetition effects in the ventral visual stream. *Journal of Neurophysiology*, 106(1), 398–408. <https://doi.org/10.1152/jn.01138.2010>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Feng, X., Altarelli, I., Monzalvo, K., Ding, G., Ramus, F., Shu, H., Dehaene, S., Meng, X., & Dehaene-Lambertz, G. (2020). A universal reading network and its modulation by writing system and reading ability in French and Chinese children. *ELife*, 9, 1–51. <https://doi.org/10.7554/eLife.54591>
- Fernandes, M. A., Moscovitch, M., Ziegler, M., & Grady, C. (2005). Brain regions associated with successful and unsuccessful retrieval of verbal episodic memory as revealed by divided attention. *Neuropsychologia*, 43(8), 1115–1127. <https://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2004.11.026>
- Fias, W., Lammertyn, J., Caessens, B., & Orban, G. A. (2007). Processing of abstract ordinal knowledge in the horizontal segment of the intraparietal sulcus. *The Journal of Neuroscience*, 27(33), 8952–8956. <https://doi.org/http://dx.doi.org/10.1523/JNEUROSCI.2076-07.2007>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 201711978. <https://doi.org/10.1073/pnas.1711978115>
- Flowers, D. L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T. A., Wood, F. B., & Eden, G. F. (2004). Attention to single letters activates left extrastriate cortex. *NeuroImage*, 21(3), 829–839. <https://doi.org/10.1016/j.neuroimage.2003.10.002>
- Fox, P. T., Laird, A. R., Eickhoff, S. B., Lancaster, J. L., Fox, M., Uecker, A. M., Robertson, M., & Ray, K. L. (2013). *User Manual for GingerALE 2.3*. Research Imaging Institute, UT Health Science Center.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural

- connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7), 2458–2463. <https://doi.org/10.1073/pnas.0509389103>
- Friedmann, N., & Coltheart, M. (2018). Types of Developmental Dyslexia. In A. Bar-On & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives* (pp. 119–151). De Gruyter Mouton. https://doi.org/10.1007/978-94-017-8545-7_6
- Friedmann, N., Dotan, D., & Rahamim, E. (2010). Is the visual analyzer orthographic-specific? Reading words and numbers in letter position dyslexia. *Cortex*, 46(8), 982–1004. <https://doi.org/http://dx.doi.org/10.1016/j.cortex.2009.08.007>
- Fulbright, R. K., Manson, S. C., Skudlarski, P., Lacadie, C. M., & Gore, J. C. (2003). Quantity determination and the distance effect with letters, numbers, and shapes: A functional MR imaging study of number processing. *American Journal of Neuroradiology*, 24(2), 193–200.
- Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex? *Trends in Cognitive Sciences*, 4(1), 1–2. [https://doi.org/10.1016/S1364-6613\(99\)01416-3](https://doi.org/10.1016/S1364-6613(99)01416-3)
- Gauthier, I. (2018). Domain-Specific and Domain-General Individual Differences in Visual Object Recognition. *Current Directions in Psychological Science*, 27(2), 97–102. <https://doi.org/10.1177/0963721417737151>
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191–197. <https://doi.org/10.1038/72140>
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6), 568–573. <https://doi.org/10.1038/9224>
- Gazzaniga, M. S., & Smylie, C. S. (1984). Dissociation of Language and Cognition. *Brain*, 107(1), 145–153. <https://doi.org/10.1093/brain/107.1.145>
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *American Statistician*, 60(4), 328–331. <https://doi.org/10.1198/000313006X152649>
- Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences*, 110(28), 11267–11271. <https://doi.org/10.1073/pnas.1220568110>
- Gilpin, A. R. (1993). Table for Conversion of Kendall’S Tau to Spearman’S Rho Within the Context of Measures of Magnitude of Effect for Meta-Analysis. *Educational and Psychological Measurement*, 53(1), 87–92. <https://doi.org/10.1177/0013164493053001007>
- Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, 76(2), 104–122. <https://doi.org/10.1006/jecp.2000.2564>
- Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for Highly Selective Neuronal Tuning to Whole Words in the “Visual Word Form Area.” *Neuron*, 62(2), 199–204. <https://doi.org/10.1016/j.neuron.2009.03.017>

- Glezer, L. S., & Riesenhuber, M. (2013). Individual Variability in Location Impacts Orthographic Selectivity in the “Visual Word Form Area.” *Journal of Neuroscience*, *33*(27), 11221–11226. <https://doi.org/10.1523/JNEUROSCI.5002-12.2013>
- Goffin, C., Sokolowski, H. M., Slipenkyj, M., & Ansari, D. (2019). Does writing handedness affect neural representation of symbolic number? An fMRI adaptation study. *Cortex*, *121*, 27–43. <https://doi.org/10.1016/j.cortex.2019.07.017>
- Goffin, C., Vogel, S. E., Slipenkyj, M., & Ansari, D. (2020). A comes before B, like 1 comes before 2. Is the parietal cortex sensitive to ordinal relationships in both numbers and letters? An fMRI-adaptation study. *Human Brain Mapping*, *41*(6), 1591–1610. <https://doi.org/10.1002/hbm.24897>
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social and Personality Psychology Compass*, *10*(10), 535–549. <https://doi.org/10.1111/spc3.12267>
- Gotts, S. J., Milleville, S. C., Bellgowan, P. S. F., & Martin, A. (2011). Broad and narrow conceptual tuning in the human frontal lobes. *Cerebral Cortex*, *21*(2), 477–491. <https://doi.org/10.1093/cercor/bhq113>
- Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., & Ebner, F. (2013). The function of the left angular gyrus in mental arithmetic: evidence from the associative confusion effect. *Human Brain Mapping*, *34*(5), 1013–1024. <https://doi.org/10.1002/hbm.21489>
- Grainger, J., & Hannagan, T. (2014). What is special about orthographic processing? *Written Language & Literacy*, *17*(2), 225–252. <https://doi.org/10.1075/wll.17.2.03gra>
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203. [https://doi.org/10.1016/S0896-6273\(00\)80832-6](https://doi.org/10.1016/S0896-6273(00)80832-6)
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*(1), 649–677. <https://doi.org/10.1146/annurev.neuro.27.070203.144220>
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. <https://doi.org/10.1038/nrn3747>
- Gross, J., Hudson, C., & Price, D. (2009). *The long term costs of numeracy difficulties*.
- Grotheer, M., Ambrus, G. G., & Kovács, G. (2016). Causal evidence of the involvement of the number form area in the visual detection of numbers and letters. *NeuroImage*, *132*, 314–319. <https://doi.org/10.1016/j.neuroimage.2016.02.069>
- Grotheer, M., Herrmann, K.-H., & Kovacs, G. (2016). Neuroimaging Evidence of a Bilateral Representation for Visually Presented Numbers. *Journal of Neuroscience*, *36*(1), 88–97. <https://doi.org/10.1523/JNEUROSCI.2129-15.2016>
- Grotheer, M., Jeska, B., & Grill-Spector, K. (2018). A preference for mathematical processing outweighs the selectivity for Arabic numbers in the inferior temporal gyrus. *NeuroImage*, *175*(November 2017), 188–200. <https://doi.org/10.1016/j.neuroimage.2018.03.064>

- Grotheer, M., Zhen, Z., Lerma-Usabiaga, G., & Grill-Spector, K. (2019). Separate lanes for adding and reading in the white matter highways of the human brain. *Nature Communications*, *10*(1), 3675. <https://doi.org/10.1038/s41467-019-11424-1>
- Gullick, M. M., & Temple, E. (2011). Are historic years understood as numbers or events? An fMRI study of numbers with semantic associations. *Brain and Cognition*, *77*(3), 356–364. <https://doi.org/10.1016/j.bandc.2011.09.004>
- Gwilliams, L., & King, J.-R. (2020). Recurrent processes support a cascade of hierarchical decisions. *ELife*, *9*, 1–20. <https://doi.org/10.7554/eLife.56603>
- Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., & Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends in Cognitive Sciences*, *19*(7), 374–382. <https://doi.org/10.1016/j.tics.2015.05.006>
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. <https://doi.org/10.1037/0033-295X.111.3.662>
- Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. <https://doi.org/10.3758/BF03202431>
- Hermes, D., Rangarajan, V., Foster, B. L., King, J.-R., Kasikci, I., Miller, K. J., & Parvizi, J. (2017). Electrophysiological Responses in the Ventral Temporal Cortex During Reading of Numerals and Calculation. *Cerebral Cortex*, *27*(1), 567–575. <https://doi.org/10.1093/cercor/bhv250>
- Hibbard, J. H., Peters, E., Dixon, A., & Tusler, M. (2007). Consumer competencies and the use of comparative quality information: It isn't just about literacy. *Medical Care Research and Review*, *64*(4), 379–394. <https://doi.org/10.1177/1077558707301630>
- Hochman Cohen, H., Berger, A., Rubinsten, O., & Henik, A. (2014). Does the learning of two symbolic sets of numbers affect the automaticity of number processing in children? *Journal of Experimental Child Psychology*, *121*(1), 96–110. <https://doi.org/10.1016/j.jecp.2013.11.005>
- Holloway, I. D., Battista, C., Vogel, S. E., & Ansari, D. (2013). Semantic and Perceptual Processing of Number Symbols: Evidence from a Cross-linguistic fMRI Adaptation Study. *Journal of Cognitive Neuroscience*, *25*(3), 388–400. https://doi.org/10.1162/jocn_a_00323
- Holloway, I. D., van Atteveldt, N., Blomert, L., & Ansari, D. (2015). Orthographic dependency in the neural correlates of reading: Evidence from audiovisual integration in English readers. *Cerebral Cortex*, *25*(6), 1544–1553. <https://doi.org/10.1093/cercor/bht347>
- Houdé, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*(6), 876–885. <https://doi.org/10.1111/j.1467-7687.2009.00938.x>
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435–448. <https://doi.org/10.1038/nrn1684>

- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. <https://doi.org/10.1037/1082-989X.11.2.193>
- Hull, A. J. (1973). A letter-digit metric of auditory confusions. *British Journal of Psychology*, *64*(4), 579–585.
- Hurst, M., Anderson, U., & Cordes, S. (2017). Mapping Among Number Words, Numerals, and Nonsymbolic Quantities in Preschoolers. *Journal of Cognition and Development*, *18*(1), 41–62. <https://doi.org/10.1080/15248372.2016.1228653>
- Hyde, D. C., Boas, D. A., Blair, C., & Carey, S. (2010). Near-infrared spectroscopy shows right parietal specialization for number in pre-verbal infants. *NeuroImage*, *53*(2), 647–652. <https://doi.org/10.1016/j.neuroimage.2010.06.030>
- Ischebeck, A., Heim, S., Siedentopf, C., Zamarian, L., Schocke, M., Kremser, C., Egger, K., Streng, H., Scheperjans, F., & Delazer, M. (2008). Are numbers special? Comparing the generation of verbal materials from ordered categories (months) to numbers and other categories (animals) in an fMRI study. *Human Brain Mapping*, *29*(8), 894–909. <https://doi.org/http://dx.doi.org/10.1002/hbm.20433>
- Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology*, *6*(2), 0275–0285. <https://doi.org/10.1371/journal.pbio.0060011>
- James, K. H., James, T. W., Jobard, G., Wong, A. C.-N., & Gauthier, I. (2005). Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(4), 452–466. <https://doi.org/10.3758/CABN.5.4.452>
- Jang, S., & Hyde, D. C. (2020). Hemispheric asymmetries in processing numerical meaning in arithmetic. *Neuropsychologia*, *146*(June), 107524. <https://doi.org/10.1016/j.neuropsychologia.2020.107524>
- Jansen, A., Menke, R., Sommer, J., Förster, A. F., Bruchmann, S., Hämle, J., Weber, B., & Knecht, S. (2006). The assessment of hemispheric lateralization in functional MRI—Robustness and reproducibility. *NeuroImage*, *33*(1), 204–217. <https://doi.org/10.1016/j.neuroimage.2006.06.019>
- JASP Team. (2019). *JASP (Version 0.10.0) [Computer software]*. <https://jasp-stats.org/>
- JASP Team. (2020). *JASP (Version 0.14.1) [Computer software]*. <https://jasp-stats.org/>
- Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, *50*(4), 544–552. <https://doi.org/10.1016/j.neuropsychologia.2011.11.007>
- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170. <https://doi.org/10.1073/pnas.1005062107>
- Karipidis, I. I., Pleisch, G., Röthlisberger, M., Hofstetter, C., Dornbierer, D., Stämpfli, P., & Brem, S. (2017). Neural initialization of audiovisual integration in prereaders at varying risk for developmental dyslexia. *Human Brain Mapping*, *38*(2), 1038–1055.

<https://doi.org/10.1002/hbm.23437>

- Karolis, V. R., Corbetta, M., & Thiebaut de Schotten, M. (2019). The architecture of functional lateralisation and its relationship to callosal connectivity in the human brain. *Nature Communications*, *10*(1), 1417. <https://doi.org/10.1038/s41467-019-09344-1>
- Kaufman, L., & Rousseuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc.
- Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L.-B., Felber, S., Ischebeck, A., & Ischebeck, A. (2006). Neural correlates of the number-size interference task in children. *Neuroreport*, *17*(6), 587–591. <https://doi.org/http://dx.doi.org/10.1097/00001756-200604240-00007>
- Kaufmann, L., Vogel, S. E., Starke, M., Kremser, C., & Schocke, M. (2009). Numerical and non-numerical ordinality processing in children with and without developmental dyscalculia: Evidence from fMRI. *Cognitive Development*, *24*(4), 486–494. <https://doi.org/10.1016/j.cogdev.2009.09.001>
- Kaufmann, L., Wood, G., Rubinsten, O., & Henik, A. (2011). Meta-Analyses of Developmental fMRI Studies Investigating Typical and Atypical Trajectories of Number Processing and Calculation. *Developmental Neuropsychology*, *36*(6), 763–787. <https://doi.org/10.1080/87565641.2010.549884>
- Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. *ELife*, *6*, 1–29. <https://doi.org/10.7554/elife.22341>
- Kim, K. K., Karunanayaka, P., Privitera, M. D., Holland, S. K., & Szaflarski, J. P. (2011). Semantic association investigated with functional MRI and independent component analysis. *Epilepsy and Behavior*, *20*(4), 613–622. <https://doi.org/10.1016/j.yebeh.2010.11.010>
- Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. <https://doi.org/10.5351/CSAM.2015.22.6.665>
- Kimura, D. (1966). Dual functional asymmetry of the brain in visual perception. *Neuropsychologia*, *4*(3), 275–285. [https://doi.org/10.1016/0028-3932\(66\)90033-9](https://doi.org/10.1016/0028-3932(66)90033-9)
- Knops, A., Nuerk, H.-C., Fimm, B., Vohn, R., & Willmes, K. (2006). A special role for numbers in working memory? An fMRI study. *NeuroImage*, *29*(1), 1–14. <https://doi.org/10.1016/j.neuroimage.2005.07.009>
- Kojouharova, P., & Krajcsi, A. (2019). Two components of the Indo-Arabic numerical size effect. *Acta Psychologica*, *192*(June 2018), 163–171. <https://doi.org/10.1016/j.actpsy.2018.11.009>
- Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., & et al. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 723–735. <https://doi.org/10.1037//0096-1523.15.4.723>
- Koul, A., Tyagi, V., & Singh, N. C. (2014). Notational usage modulates attention networks in binumerates. *Frontiers in Human Neuroscience*, *8*, 326.

<https://doi.org/10.3389/fnhum.2014.00326>

- Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology*, 7(NOV), 1–16. <https://doi.org/10.3389/fpsyg.2016.01795>
- Kriegeskorte, N., & Bandettini, P. A. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, 38(4), 649–662. <https://doi.org/10.1016/j.neuroimage.2007.02.022>
- Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3863–3868. <https://doi.org/10.1073/pnas.0600244103>
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412. <https://doi.org/10.1016/j.tics.2013.06.007>
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2(November), 4. <https://doi.org/10.3389/neuro.06.004.2008>
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, 60(6), 1126–1141. <https://doi.org/10.1016/j.neuron.2008.10.043>
- Laird, A. R., Lancaster, J. L., & Fox, P. T. (2005). BrainMap: The Social Evolution of a Human Brain Mapping Database. *Neuroinformatics*, 3(1), 065–078. <https://doi.org/10.1385/NI:3:1:065>
- Laird, A. R., Robinson, J. L., McMillan, K. M., Tordesillas-Gutiérrez, D., Moran, S. T., Gonzales, S. M., Ray, K. L., Franklin, C., Glahn, D. C., Fox, P. T., & Lancaster, J. L. (2010). Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: Validation of the Lancaster transform. *NeuroImage*, 51(2), 677–683. <https://doi.org/10.1016/j.neuroimage.2010.02.048>
- Lakens, D., & Etz, A. J. (2017). Too True to be Bad: When Sets of Studies With Significant and Nonsignificant Findings Are Probably True. *Social Psychological and Personality Science*, 8(8), 875–881. <https://doi.org/10.1177/1948550617693058>
- Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C., & Fox, P. T. (2007). Bias between MNI and talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, 28(11), 1194–1205. <https://doi.org/10.1002/hbm.20345>
- Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, 32(1), 1–10. <https://doi.org/10.5334/irsp.289>
- Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*, 74(March 2017), 150–156. <https://doi.org/10.1016/j.jesp.2017.09.011>

- Libertus, M. E., Brannon, E. M., & Pelphrey, K. A. (2009). Developmental changes in category-specific brain responses to numbers and letters in a working memory task. *NeuroImage*, *44*(4), 1404–1414. <https://doi.org/10.1016/j.neuroimage.2008.10.027>
- Libertus, M. E., Pruitt, L. B., Woldorff, M. G., & Brannon, E. M. (2009). Induced alpha-band oscillations reflect ratio-dependent number discrimination in the infant brain. *Journal of Cognitive Neuroscience*, *21*(12), 2398–2406. <https://doi.org/10.1162/jocn.2008.21162>
- Liu, Y., Dunlap, S., Fiez, J., & Perfetti, C. (2007). Evidence for neural accommodation to a writing system following learning. *Human Brain Mapping*, *28*(11), 1223–1234. <https://doi.org/10.1002/hbm.20356>
- Lochy, A., & Schiltz, C. (2019). Lateralized Neural Responses to Letters and Digits in First Graders. *Child Development*, *90*(6), 1866–1874. <https://doi.org/10.1111/cdev.13337>
- Lyons, I. M., & Ansari, D. (2009). The cerebral basis of mapping nonsymbolic numerical quantities onto abstract symbols: An fMRI training study. *Journal of Cognitive Neuroscience*, *21*(9), 1720–1735. <https://doi.org/http://dx.doi.org/10.1162/jocn.2009.21124>
- Lyons, I. M., & Beilock, S. L. (2018). Characterizing the neural coding of symbolic quantities. *NeuroImage*, *178*(September), 503–518. <https://doi.org/10.1016/j.neuroimage.2018.05.062>
- Lyons, I. M., Vogel, S. E., & Ansari, D. (2016). On the ordinality of numbers: A review of neural and behavioral studies. In *The Mathematical Brain Across the Lifespan* (1st ed., pp. 187–221). Elsevier B.V. <https://doi.org/10.1016/bs.pbr.2016.04.010>
- Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-Related Properties Shape Object Representations in the Ventral Stream. *Neuron*, *55*(3), 507–520. <https://doi.org/10.1016/j.neuron.2007.07.011>
- Marangolo, P., Nasti, M., & Zorzi, M. (2004). Selective impairment for reading numbers and number words: A single case study. *Neuropsychologia*, *42*(8), 997–1006. <https://doi.org/10.1016/j.neuropsychologia.2004.01.004>
- Marangolo, P., Piras, F., & Fias, W. (2005). “I can write seven but I can’t say it”: A case of domain-specific phonological output deficit for numbers. *Neuropsychologia*, *43*(8), 1177–1188. <https://doi.org/10.1016/j.neuropsychologia.2004.11.001>
- Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife*, *7*, e31873. <https://doi.org/10.7554/eLife.31873>
- Martin, L., Durisko, C., Moore, M. W., Coutanche, M. N., Chen, D., & Fiez, J. A. (2019). The VWFA Is the Home of Orthographic Learning When Houses Are Used as Letters. *ENEURO*, *6*(1), ENEURO.0425-17.2019. <https://doi.org/10.1523/ENEURO.0425-17.2019>
- Martin, R., & Hodgson, H. (2014). *Pro Bono Numeracy Economics Report for National Numeracy: Cost of outcomes associated with low levels of adult numeracy in the UK*.
- Masataka, N., Ohnishi, T., Imabayashi, E., Hirakata, M., & Matsuda, H. (2007). Neural correlates for learning to read Roman numerals. *Brain and Language*, *100*(3), 276–282. <https://doi.org/10.1016/j.bandl.2006.11.011>
- Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining

- mathematics learning disability in the primary school-age years. *Annals of Dyslexia*, 53(1), 218–253. <https://doi.org/10.1007/s11881-003-0011-7>
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7), 293–299. [https://doi.org/10.1016/S1364-6613\(03\)00134-7](https://doi.org/10.1016/S1364-6613(03)00134-7)
- McCloskey, M., & Schubert, T. (2014). Shared versus separate processes for letter and digit identification. *Cognitive Neuropsychology*, 31(5–6), 437–460. <https://doi.org/10.1080/02643294.2013.869202>
- McGugin, R. W., Newton, A. T., Gore, J. C., & Gauthier, I. (2014). Robust expertise effects in right FFA. *Neuropsychologia*, 63, 135–144. <https://doi.org/10.1016/j.neuropsychologia.2014.08.029>
- McGugin, R. W., Van Gulick, A. E., Tamber-Rosenau, B. J., Ross, D. A., & Gauthier, I. (2015). Expertise effects in face-selective areas are robust to clutter and diverted attention, but not to competition. *Cerebral Cortex*, 25(9), 2610–2622. <https://doi.org/10.1093/cercor/bhu060>
- Merkley, R., Conrad, B., Price, G. R., & Ansari, D. (2019). Investigating the visual number form area: a replication study. *Royal Society Open Science*, 6(10), 182067. <https://doi.org/10.1098/rsos.182067>
- Messina, G., Denes, G., & Basso, A. (2009). Words and number words transcoding: A retrospective study on 57 aphasic subjects. *Journal of Neurolinguistics*, 22(5), 486–494. <https://doi.org/10.1016/j.jneuroling.2009.04.001>
- Meyer, L., Obleser, J., Anwander, A., & Friederici, A. D. (2012). Linking ordering in Broca's area to storage in left temporo-parietal regions: the case of sentence processing. *NeuroImage*, 62(3), 1987–1998. <https://doi.org/10.1016/j.neuroimage.2012.05.052>
- Miozzo, M., & Caramazza, A. (1998). Varieties of pure alexia: The case of failure to access graphemic representations. *Cognitive Neuropsychology*, 15(1–2), 203–238. <https://doi.org/10.1080/026432998381267>
- Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, 53(1), 103–118. <https://doi.org/10.1016/j.neuroimage.2010.05.051>
- Moeller, K., Willmes, K., & Klein, E. (2015). A review on functional and structural brain connectivity in numerical cognition. *Frontiers in Human Neuroscience*, 9(May), 1–14. <https://doi.org/10.3389/fnhum.2015.00227>
- Mongelli, V., Dehaene, S., Vinckier, F., Peretz, I., Bartolomeo, P., & Cohen, L. (2017). Music and words in the visual cortex: The impact of musical expertise. *Cortex*, 86, 260–274. <https://doi.org/10.1016/j.cortex.2016.05.016>
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.12-4.2. <https://cran.r-project.org/package=BayesFactor>
- Naparstek, S., & Henik, A. (2010). Count me in! On the automaticity of numerosity processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(4), 1053–1059. <https://doi.org/10.1037/a0019766>

- Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Di Oleggio Castello, M. V., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex*, 27(8), 4277–4291. <https://doi.org/10.1093/cercor/bhx138>
- Nelson, J. R., Liu, Y., Fiez, J., & Perfetti, C. A. (2009). Assimilation and accommodation patterns in ventral occipitotemporal cortex in learning a second writing system. *Human Brain Mapping*, 30(3), 810–820. <https://doi.org/10.1002/hbm.20551>
- Nemmi, F., Schel, M. A., & Klingberg, T. (2018). Connectivity of the Human Number Form Area Reveals Development of a Cortical Network for Mathematics. *Frontiers in Human Neuroscience*, 12(November), 1–15. <https://doi.org/10.3389/fnhum.2018.00465>
- Nieder, A. (2009). Prefrontal cortex and the evolution of symbolic reference. *Current Opinion in Neurobiology*, 19(1), 99–108. <https://doi.org/10.1016/j.conb.2009.04.008>
- Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, 63(3), 1646–1669. <https://doi.org/10.1016/j.neuroimage.2012.06.065>
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14(9), 1105–1107. <https://doi.org/10.1038/nn.2886>
- Nili, H., Walther, A., Alink, A., & Kriegeskorte, N. (2020). Inferring exemplar discriminability in brain representations. *PLOS ONE*, 15(6), e0232551. <https://doi.org/10.1371/journal.pone.0232551>
- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, 10(4), e1003553. <https://doi.org/10.1371/journal.pcbi.1003553>
- Olulade, O. A., Flowers, D. L., Napoliello, E. M., & Eden, G. F. (2015). Dyslexic children lack word selectivity gradients in occipito-temporal and inferior frontal cortex. *NeuroImage: Clinical*, 7, 742–754. <https://doi.org/10.1016/j.nicl.2015.02.013>
- Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage*, 49(3), 1943–1948. <https://doi.org/10.1016/j.neuroimage.2009.02.047>
- Op de Beeck, H. P., Pillet, I., & Ritchie, J. B. (2019). Factors Determining Where Category-Selective Areas Emerge in Visual Cortex. *Trends in Cognitive Sciences*, 23(9), 784–797. <https://doi.org/10.1016/j.tics.2019.06.006>
- Op de Beeck, H. P., Torfs, K., & Wagemans, J. (2008). Perceived Shape Similarity among Unfamiliar Objects and the Organization of the Human Object Vision Pathway. *Journal of Neuroscience*, 28(40), 10111–10123. <https://doi.org/10.1523/JNEUROSCI.2511-08.2008>
- Ortuño, F., Ojeda, N., Arbizu, J., López, P., Martí-Clement, J. M., Peñuelas, I., & Cervera, S. (2002). Sustained attention in a counting task: normal performance and functional neuroanatomy. *NeuroImage*, 17(1), 411–420. <https://doi.org/10.1006/nimg.2002.1168>
- Park, J., Chiang, C., Brannon, E. M., & Woldorff, M. G. (2014). Experience-dependent Hemispheric Specialization of Letters and Numbers Is Revealed in Early Visual Processing.

- Journal of Cognitive Neuroscience*, 26(10), 2239–2249.
https://doi.org/10.1162/jocn_a_00621
- Park, J., Hebrank, A., Polk, T. A., & Park, D. C. (2012). Neural Dissociation of Number from Letter Recognition and Its Relationship to Parietal Numerical Processing. *Journal of Cognitive Neuroscience*, 24(1), 39–50. https://doi.org/10.1162/jocn_a_00085
- Park, J., van den Berg, B., Chiang, C., Woldorff, M. G., & Brannon, E. M. (2018). Developmental trajectory of neural specialization for letter and number visual processing. *Developmental Science*, 21(3), 1–14. <https://doi.org/10.1111/desc.12578>
- Parsons, S., & Bynner, J. (1997). Numeracy and employment. *Education & Training*, 39(2), 43–51. <https://doi.org/10.1108/00400919710164125>
- Parsons, S., & Bynner, J. (2005). *Does Numeracy Matter More?*
- Parviainen, T. (2006). Cortical Sequence of Word Perception in Beginning Readers. *Journal of Neuroscience*, 26(22), 6052–6061. <https://doi.org/10.1523/jneurosci.0673-06.2006>
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362(6418), 342–345.
<https://doi.org/10.1038/362342a0>
- Peer, M., Abboud, S., Hertz, U., Amedi, A., & Arzy, S. (2016). Intensity-based masking: A tool to improve functional connectivity results of resting-state fMRI. *Human Brain Mapping*, 37(7), 2407–2418. <https://doi.org/10.1002/hbm.23182>
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46(28), 4646–4674.
<https://doi.org/10.1016/j.visres.2006.04.023>
- Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L. H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain’s reading network. *Bilingualism*, 10(2), 131–146. <https://doi.org/10.1017/S1366728907002891>
- Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, 21(1), 31–35.
<https://doi.org/10.1177/0963721411429960>
- Peters, L., De Smedt, B., & Op de Beeck, H. P. (2015). The neural representation of Arabic digits in visual cortex. *Frontiers in Human Neuroscience*, 9(September), 517.
<https://doi.org/10.3389/fnhum.2015.00517>
- Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, 1106(1), 177–188. <https://doi.org/10.1016/j.brainres.2006.05.104>
- Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, 53(2), 293–305. <https://doi.org/10.1016/j.neuron.2006.11.022>
- Pinel, P., & Dehaene, S. (2010). Beyond Hemispheric Dominance: Brain Regions Underlying the Joint Lateralization of Language and Arithmetic to the Left Hemisphere. *Journal of Cognitive Neuroscience*, 22(1), 48–66. <https://doi.org/10.1162/jocn.2009.21184>

- Pinel, P., & Dehaene, S. (2013). Genetic and environmental contributions to brain activation during calculation. *NeuroImage*, *81*, 306–316. <https://doi.org/10.1016/j.neuroimage.2013.04.118>
- Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage*, *14*(5), 1013–1026. <https://doi.org/10.1006/nimg.2001.0913>
- Pinel, P., Le Clec'H, G., van de Moortele, P.-F. F., Naccache, L., Le Bihan, D., Dehaene, S., Le Clec'H, G., van de Moortele, P.-F. F., Naccache, L., Le Bihan, D., & Dehaene, S. (1999). Event-related fMRI analysis of the cerebral circuit for number comparison. *NeuroReport*, *10*(7), 1473–1479. <https://doi.org/10.1097/00001756-199905140-00015>
- Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*(6), 983–993. [https://doi.org/10.1016/S0896-6273\(04\)00107-2](https://doi.org/10.1016/S0896-6273(04)00107-2)
- Pinheiro-Chagas, P., Daich, A., Parvizi, J., & Dehaene, S. (2018). Brain Mechanisms of Arithmetic: A Crucial Role for Ventral Temporal Cortex. *Journal of Cognitive Neuroscience*, *30*(12), 1757–1772. https://doi.org/10.1162/jocn_a_01319
- Piras, F., & Marangolo, P. (2009). Word and number reading in the brain: Evidence from a Voxel-based Lesion-symptom Mapping study. *Neuropsychologia*, *47*(8–9), 1944–1953. <https://doi.org/10.1016/j.neuropsychologia.2009.03.006>
- Pleisch, G., Karipidis, I. I., Brauchli, C., Röthlisberger, M., Hofstetter, C., Stämpfli, P., Walitza, S., & Brem, S. (2019). Emerging neural specialization of the ventral occipitotemporal cortex to characters through phonological association learning in preschool children. *NeuroImage*, *189*, 813–831. <https://doi.org/10.1016/j.neuroimage.2019.01.046>
- Polk, T. A., & Farah, M. J. (1995). Brain localization for arbitrary stimulus categories: a simple account based on Hebbian learning. *Proceedings of the National Academy of Sciences*, *92*(26), 12370–12373. <https://doi.org/10.1073/pnas.92.26.12370>
- Polk, T. A., & Farah, M. J. (1998). The neural development and organization of letter recognition: Evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proceedings of the National Academy of Sciences*, *95*(3), 847–852. <https://doi.org/10.1073/pnas.95.3.847>
- Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D'Esposito, M., Detre, J. A., & Farah, M. J. (2002). Neural Specialization for Letter Recognition. *Journal of Cognitive Neuroscience*, *14*(2), 145–159. <https://doi.org/10.1162/089892902317236803>
- Pollack, C., & Price, G. R. (2019). Neurocognitive mechanisms of digit processing and their relationship with mathematics competence. *NeuroImage*, *185*, 245–254. <https://doi.org/10.1016/j.neuroimage.2018.10.047>
- Price, C. J. (1998). The functional anatomy of word comprehension and production. *Trends in Cognitive Sciences*, *2*(8), 281–288. [https://doi.org/10.1016/S1364-6613\(98\)01201-7](https://doi.org/10.1016/S1364-6613(98)01201-7)
- Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*(3), 473–481. [https://doi.org/10.1016/S1053-8119\(03\)00084-3](https://doi.org/10.1016/S1053-8119(03)00084-3)
- Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal

- contributions to reading. *Trends in Cognitive Sciences*, *15*(6), 246–253.
<https://doi.org/10.1016/j.tics.2011.04.001>
- Price, C. J., Devlin, J. T., Moore, C. J., Morton, C., & Laird, A. R. (2005). Meta-analyses of object naming: Effect of baseline. *Human Brain Mapping*, *25*(1), 70–82.
<https://doi.org/10.1002/hbm.20132>
- Price, G. R., & Ansari, D. (2011). Symbol processing in the left angular gyrus: Evidence from passive perception of digits. *NeuroImage*, *57*(3), 1205–1211.
<https://doi.org/10.1016/j.neuroimage.2011.05.035>
- Price, G. R., Mazzocco, M. M. M., & Ansari, D. (2013). Why Mental Arithmetic Counts: Brain Activation during Single Digit Arithmetic Predicts High School Math Scores. *Journal of Neuroscience*, *33*(1), 156–163. <https://doi.org/10.1523/JNEUROSCI.2936-12.2013>
- Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology*, *105*(2), 453–464. <https://doi.org/10.1037/a0031753>
- Qu, J., Zhang, L., Chen, C., Xie, P., Li, H., Liu, X., & Mei, L. (2019). Cross-Language Pattern Similarity in the Bilateral Fusiform Cortex Is Associated with Reading Proficiency in Second Language. *Neuroscience*, *410*, 254–263.
<https://doi.org/10.1016/j.neuroscience.2019.05.019>
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing [Computer software]*. <https://cran.r-project.org/>
- Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology*, *21*(5), 363–368.
<https://doi.org/10.1016/j.cub.2011.01.040>
- Reinke, K., Fernandes, M., Schwindt, G., O’Craven, K., Grady, C. L., O’Craven, K., & Grady, C. L. (2008). Functional specificity of the visual word form area: General activation for words and symbols but specific network activation for words. *Brain and Language*, *104*(2), 180–189. <https://doi.org/10.1016/j.bandl.2007.04.006>
- Richler, J. J., Tomarken, A. J., Sunday, M. A., Vickery, T. J., Ryan, K. F., Floyd, R. J., Sheinberg, D., Wong, A. C. N., & Gauthier, I. (2019). Individual differences in object recognition. *Psychological Review*, *126*(2), 226–251. <https://doi.org/10.1037/rev0000129>
- Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition*, *166*, 42–55.
<https://doi.org/10.1016/j.cognition.2017.05.019>
- Ringo, J. L., Doty, R. W., Demeter, S., & Simard, P. Y. (1994). Time Is of the Essence: A Conjecture that Hemispheric Specialization Arises from Interhemispheric Conduction Delay. *Cerebral Cortex*, *4*(4), 331–343. <https://doi.org/10.1093/cercor/4.4.331>
- Rivera-Batiz, F. L. (1992). Quantitative Literacy and the Likelihood of Employment among Young Adults in the United States. *The Journal of Human Resources*, *27*(2), 313.
<https://doi.org/10.2307/145737>
- Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental Changes in Mental Arithmetic: Evidence for Increased Functional Specialization in the Left Inferior

- Parietal Cortex. *Cerebral Cortex*, 15(11), 1779–1790. <https://doi.org/10.1093/cercor/bhi055>
- Rothlein, D., & Rapp, B. (2014). The similarity structure of distributed neural responses reveals the multiple representations of letters. *NeuroImage*, 89, 331–344. <https://doi.org/10.1016/j.neuroimage.2013.11.054>
- Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem’s ESP claim. *Psychonomic Bulletin & Review*, 18(4), 682–689. <https://doi.org/10.3758/s13423-011-0088-7>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rousselet, G. A., & Pernet, C. R. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience*, 6(May), 119. <https://doi.org/10.3389/fnhum.2012.00119>
- Roux, F. E., Lubrano, V., Lauwers-Cances, V., Giussani, C., & Demonet, J.-F. (2008). Cortical areas involved in Arabic number reading. *Neurology*, 70(3), 210–217. <https://doi.org/10.1212/01.wnl.0000297194.14452.a0>
- Rubinsten, O., & Henik, A. (2005). Automatic Activation of Internal Magnitudes: A Study of Developmental Dyscalculia. *Neuropsychology*, 19(5), 641–648. <https://doi.org/10.1037/0894-4105.19.5.641>
- Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology*, 81(1), 74–92. <https://doi.org/10.1006/jecp.2001.2645>
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, 20(3), e12372. <https://doi.org/10.1111/desc.12372>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis. *Child Development*, 89(5), 1467–1484. <https://doi.org/10.1111/cdev.13068>
- Schubert, T. M. (2017). Why are digits easier to identify than letters? *Neuropsychologia*, 95(December 2016), 136–155. <https://doi.org/10.1016/j.neuropsychologia.2016.12.016>
- Seghier, M. L. (2008). Laterality index in functional MRI: methodological issues. *Magnetic Resonance Imaging*, 26(5), 594–601. <https://doi.org/10.1016/j.mri.2007.10.010>
- Seghier, M. L. (2013). The Angular Gyrus. *The Neuroscientist*, 19(1), 43–61. <https://doi.org/10.1177/1073858412440596>
- Seghier, M. L. (2019). Categorical laterality indices in fMRI: a parallel with classic similarity indices. *Brain Structure and Function*, 224(3), 1377–1383. <https://doi.org/10.1007/s00429-019-01833-9>
- Seidenberg, M. S. (2005). Connectionist Models of Word Reading. *Current Directions in Psychological Science*, 14(5), 238–242. <https://doi.org/10.1111/j.0963-7214.2005.00372.x>

- Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, *96*(4), 523–568. <https://doi.org/10.1037/0033-295X.96.4.523>
- Sergent, J. (1990). Furtive incursions into bicameral minds. *Brain*, *113*(2), 537–568. <https://doi.org/10.1093/brain/113.2.537>
- Seymour, S. E., Reuter-lorenz, P. A., & Gazzaniga, M. S. (1994). The disconnection syndrome: Basic findings reaffirmed. *Brain*, *117*(1), 105–115. <https://doi.org/10.1093/brain/117.1.105>
- Shovman, M. M., & Ahissar, M. (2006). Isolating the impact of visual perception on dyslexics' reading ability. *Vision Research*, *46*(20), 3514–3525. <https://doi.org/10.1016/j.visres.2006.05.011>
- Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., Miller, K. J., & Parvizi, J. (2013). A brain area for visual numerals. *The Journal of Neuroscience*, *33*(16), 6709–6715. <https://doi.org/10.1523/JNEUROSCI.4558-12.2013>
- Smith, D. E., & Karpinski, L. C. (1911). *The Hindu-Arabic Numerals*. Ginn and Company.
- Sokolowski, H. M., Fias, W., Mousa, A., & Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: A functional neuroimaging meta-analysis. *NeuroImage*, *146*(October 2016), 376–394. <https://doi.org/10.1016/j.neuroimage.2016.10.028>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *1904-1920*, *3*(3), 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Stanescu-Cosson, R., Pinel, P., van de Moortele, P. F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*, *123*(11), 2240–2255. <https://doi.org/10.1093/brain/123.11.2240>
- Starrfelt, R., & Behrmann, M. (2011). Number reading in pure alexia—A review. *Neuropsychologia*, *49*(9), 2283–2298. <https://doi.org/10.1016/j.neuropsychologia.2011.04.028>
- Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron*, *76*(3), 640–652. <https://doi.org/10.1016/j.neuron.2012.08.026>
- Szwed, M., Qiao, E., Jobert, A., Dehaene, S., & Cohen, L. (2014). Effects of Literacy in Early Visual and Occipitotemporal Areas of Chinese and French Readers. *Journal of Cognitive Neuroscience*, *26*(3), 459–475. https://doi.org/10.1162/jocn_a_00499
- Talairach, J., & Tournoux, P. (1988). Co-Planar Stereotaxis Atlas of the Human Brain. In *Direct* (Vol. 270).
- Tang, J., Critchley, H. D., Glaser, D. E., Dolan, R. J., & Butterworth, B. (2006). Imaging informational conflict: a functional magnetic resonance imaging study of numerical stroop. *Journal of Cognitive Neuroscience*, *18*(12), 2049–2062. <https://doi.org/10.1162/jocn.2006.18.12.2049>
- Taylor, D. A. (1978). Identification and categorization of letters and digits. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 4(3), 423–439.
<https://doi.org/10.1037/0096-1523.4.3.423>
- Teng, E. L., & Sperry, R. W. (1973). Interhemispheric interaction during simultaneous bilateral presentation of letters or digits in commissurotomed patients. *Neuropsychologia*, 11(2), 131–140. [https://doi.org/10.1016/0028-3932\(73\)90001-8](https://doi.org/10.1016/0028-3932(73)90001-8)
- Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, 33(1), 1–13. <https://doi.org/10.1002/hbm.21186>
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How Many Studies Do You Need? A Primer on Statistical Power for Meta-Analysis. *Journal of Educational and Behavioral Statistics*, 35(2), 215–247. <https://doi.org/10.3102/1076998609346961>
- van der Mark, S., Bucher, K., Maurer, U., Schulz, E., Brem, S., Buckelmüller, J., Kronbichler, M., Loenneker, T., Klaver, P., Martin, E., & Brandeis, D. (2009). Children with dyslexia lack multiple specializations along the visual word-form (VWF) system. *NeuroImage*, 47(4), 1940–1949. <https://doi.org/10.1016/j.neuroimage.2009.05.021>
- van der Mark, S., Klaver, P., Bucher, K., Maurer, U., Schulz, E., Brem, S., Martin, E., & Brandeis, D. (2011). The left occipitotemporal system in reading: Disruption of focal fMRI connectivity to left inferior frontal and inferior parietal language areas in children with dyslexia. *NeuroImage*, 54(3), 2426–2436. <https://doi.org/10.1016/j.neuroimage.2010.10.002>
- van der Ven, F., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2016). Non-symbolic and symbolic notations in simple arithmetic differentially involve intraparietal sulcus and angular gyrus activity. *Brain Research*, 1643, 91–102.
<https://doi.org/10.1016/j.brainres.2016.04.050>
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, 55(1), 143–156.
<https://doi.org/10.1016/j.neuron.2007.05.031>
- Visser, R. M., de Haan, M. I. C., Beemsterboer, T., Haver, P., Kindt, M., & Scholte, H. S. (2016). Quantifying learning-dependent changes in the brain: Single-trial multivoxel pattern analysis requires slow event-related fMRI. *Psychophysiology*, 53(8), 1117–1127.
<https://doi.org/10.1111/psyp.12665>
- Vogel, S. E., Goffin, C., & Ansari, D. (2015). Developmental specialization of the left parietal cortex for the semantic representation of Arabic numerals: An fMR-adaptation study. *Developmental Cognitive Neuroscience*, 12(1), 61–73.
<https://doi.org/10.1016/j.dcn.2014.12.001>
- Vogel, S. E., Goffin, C., Bohnenberger, J., Koschutnig, K., Reishofer, G., Grabner, R. H., & Ansari, D. (2017). The left intraparietal sulcus adapts to symbolic number in both the visual and auditory modalities: Evidence from fMRI. *NeuroImage*, 153(June), 16–27.
<https://doi.org/10.1016/j.neuroimage.2017.03.048>
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspectives on Psychological Science*, 4(3), 274–290. <https://doi.org/10.1111/j.1745-6924.2009.01125.x>

- Walker, D. A. (2003). JMASM9: Converting Kendall's Tau For Correlational Or Meta-Analytic Analyses. *Journal of Modern Applied Statistical Methods*, 2(2), 525–530. <https://doi.org/10.22237/jmasm/1067646360>
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, 137, 188–200. <https://doi.org/10.1016/j.neuroimage.2015.12.012>
- Wiener, M., Hamilton, R., Turkeltaub, P., Matell, M. S., & Coslett, H. B. (2010). Fast Forward: Supramarginal Gyrus Stimulation Alters Time Measurement. *Journal of Cognitive Neuroscience*, 22(1), 23–31. <https://doi.org/10.1162/jocn.2009.21191>
- Wiener, M., Turkeltaub, P. E., & Coslett, H. B. (2010). Implicit timing activates the left inferior parietal cortex. *Neuropsychologia*, 48(13), 3967–3971. <https://doi.org/10.1016/j.neuropsychologia.2010.09.014>
- Wilcox, R. R. (2004). Inferences based on a skipped correlation coefficient. *Journal of Applied Statistics*, 31(2), 131–143. <https://doi.org/10.1080/0266476032000148821>
- Wilkey, E. D., Barone, J. C., Mazzocco, M. M. M., Vogel, S. E., & Price, G. R. (2017). The effect of visual parameters on neural activation during nonsymbolic number comparison and its relation to math competency. *NeuroImage*, 159(August), 430–442. <https://doi.org/10.1016/j.neuroimage.2017.08.023>
- Wilkey, E. D., Conrad, B. N., Yeo, D. J., & Price, G. R. (2020). Shared Numerosity Representations Across Formats and Tasks Revealed with 7 Tesla fMRI: Decoding, Generalization, and Individual Differences in Behavior. *Cerebral Cortex Communications*, 1(1), 1–19. <https://doi.org/10.1093/texcom/tgaa038>
- Wong, A. C. N., Jobard, G., James, K. H., James, T. W., & Gauthier, I. (2009). Expertise with characters in alphabetic and nonalphabetic writing systems engage overlapping occipito-temporal areas. *Cognitive Neuropsychology*, 26(1), 111–127. <https://doi.org/10.1080/02643290802340972>
- Wong, Y. K., & Gauthier, I. (2010). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, 22(4), 695–713. <https://doi.org/10.1162/jocn.2009.21229>
- Wong, Y. K., Peng, C., Fratus, K. N., Woodman, G. F., & Gauthier, I. (2014). Perceptual Expertise and Top-Down Expectation of Musical Notation Engages the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, 26(8), 1629–1643. https://doi.org/10.1162/jocn_a_00616
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside.
- Wu, S. S., Chang, T. T., Majid, A., Caspers, S., Eickhoff, S. B., & Menon, V. (2009). Functional Heterogeneity of Inferior Parietal Cortex during Mathematical Cognition Assessed with Cytoarchitectonic Probability Maps. *Cerebral Cortex*, 19(12), 2930–2945. <https://doi.org/10.1093/cercor/bhp063>
- Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*, 8(7). <https://doi.org/10.1371/journal.pone.0068910>

- Yeo, D. J., Pollack, C., Merkle, R., Ansari, D., & Price, G. R. (2020). The “Inferior Temporal Numeral Area” distinguishes numerals from other character categories during passive viewing: A representational similarity analysis. *NeuroImage*, *214*, 116716. <https://doi.org/10.1016/j.neuroimage.2020.116716>
- Yeo, D. J., Wilkey, E. D., & Price, G. R. (2017). The search for the number form area: A functional neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *78*, 145–160. <https://doi.org/10.1016/j.neubiorev.2017.04.027>
- Yin, L.-J., Lou, Y.-T., Fan, M.-X., Wang, Z.-X., & Hu, Y. (2015). Neural evidence for the use of digit-image mnemonic in a superior memorist: An fMRI study. *Frontiers in Human Neuroscience*, *9*, 109. <https://doi.org/10.3389/fnhum.2015.00109>
- Zago, L., Petit, L., Turbelin, M.-R., Andersson, F., Vigneau, M., & Tzourio-Mazoyer, N. (2008). How verbal and spatial manipulation networks contribute to calculation: An fMRI study. *Neuropsychologia*, *46*(9), 2403–2414. <https://doi.org/10.1016/j.neuropsychologia.2008.03.001>
- Zamarian, L., Ischebeck, A., & Delazer, M. (2009). Neuroscience of learning arithmetic—Evidence from brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *33*(6), 909–925. <https://doi.org/10.1016/j.neubiorev.2009.03.005>
- Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Reishofer, G., & Ischebeck, A. (2012). The influence of verbalization on the pattern of cortical activation during mental arithmetic. *Behavioral and Brain Functions*, *8*(1), 13. <https://doi.org/http://dx.doi.org/10.1186/1744-9081-8-13>
- Zeithamova, D., de Araujo Sanchez, M. A., & Adke, A. (2017). Trial timing and pattern-information analyses of fMRI data. *NeuroImage*, *153*(November 2016), 221–231. <https://doi.org/10.1016/j.neuroimage.2017.04.025>
- Zhang, B., He, S., & Weng, X. (2018). Localization and functional characterization of an occipital visual word form sensitive area. *Scientific Reports*, *8*(1), 1–11. <https://doi.org/10.1038/s41598-018-25029-z>
- Zhang, H., Chen, C., & Zhou, X. (2012). Neural correlates of numbers and mathematical terms. *NeuroImage*, *60*(1), 230–240. <https://doi.org/10.1016/j.neuroimage.2011.12.006>
- Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, *57*(3), 271–295. [https://doi.org/10.1016/0010-0277\(95\)00674-3](https://doi.org/10.1016/0010-0277(95)00674-3)
- Ziegler, J. C., Pech-Georgel, C., Dufau, S., & Grainger, J. (2010). Rapid processing of letters, digits and symbols: What purely visual-attentional deficit in developmental dyslexia? *Developmental Science*, *13*(4), 8–14. <https://doi.org/10.1111/j.1467-7687.2010.00983.x>