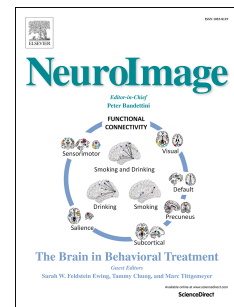


Journal Pre-proof

The “Inferior Temporal Numeral Area” distinguishes numerals from other character categories during passive viewing: A representational similarity analysis

Darren J. Yeo, Courtney Pollack, Rebecca Merkley, Daniel Ansari, Gavin R. Price



PII: S1053-8119(20)30203-2

DOI: <https://doi.org/10.1016/j.neuroimage.2020.116716>

Reference: YNIMG 116716

To appear in: *NeuroImage*

Received Date: 10 October 2019

Revised Date: 26 February 2020

Accepted Date: 3 March 2020

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

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.

Yeo, Pollack, Merkley, Ansari, and Price (2020)

CRedit authorship contribution statement

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

Title

The “Inferior Temporal Numeral Area” distinguishes numerals from other character categories during passive viewing: A representational similarity analysis

Authors

Darren J. Yeo ^{a, b}, Courtney Pollack ^c, Rebecca Merkley ^{d, e}, Daniel Ansari ^d, Gavin R. Price ^a

Author Affiliations

^a Department of Psychology & Human Development, Peabody College, Vanderbilt University, 230 Appleton Place, Nashville, TN, 37203.

^b Division of Psychology, School of Social Sciences, Nanyang Technological University, 48 Nanyang Avenue, Singapore, 639818.

^c Lynch School of Education, Boston College, Campion Hall, 140 Commonwealth Ave, Chestnut Hill, MA 02467.

^d Brain and Mind Institute, and Department of Psychology, University of Western Ontario, Westminister Hall, London, ON, N6A 3K7, Canada.

^e Institute of Cognitive Science, 2201 Dunton Tower, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada.

Corresponding Author:

Gavin R. Price

Email: gavin.price@vanderbilt.edu

Abstract

A region in the posterior inferior temporal gyrus (pITG) is thought to be specialized for processing Arabic numerals, but fMRI studies that compared passive viewing of numerals to other character types (e.g., letters and novel characters) have not found evidence of numeral preference in the pITG. However, recent studies showed that the engagement of the pITG is modulated by attention and task contexts, suggesting that passive viewing paradigms may be ill-suited for examining numeral specialization in the pITG. It is possible, however, that even if the strengths of responses to different category types are similar, the distributed response patterns (i.e., neural representations) in a candidate numeral-preferring pITG region (“pITG-numerals”) may reveal categorical distinctions, even during passive viewing. Using representational similarity analyses with three datasets that share the same task paradigm and stimulus sets (total $N = 88$), we tested whether the neural representations of digits, letters, and novel characters in pITG-numerals were organized according to visual form and/or conceptual categories (e.g., familiar versus novel, numbers versus others). Small-scale frequentist and Bayesian meta-analyses of our dataset-specific findings revealed that the organization of neural representations in pITG-numerals is unlikely to be described by differences in abstract shape, but can be described by a categorical “digits versus letters” distinction, or even a “digits versus others” distinction (suggesting greater numeral sensitivity). Evidence of greater numeral sensitivity during passive viewing suggest that pITG-numerals is likely part of a neural pathway that has been developed for automatic processing objects with potential numerical relevance. Given that numerals and letters do not differ categorically in terms of shape, categorical distinction in pITG-numerals during passive viewing must reflect ontogenetic differentiation of symbol set representations based on repeated usage of numbers and letters in differing task contexts.

Keywords

numerical cognition; symbol processing; character categorization; inferior temporal gyrus; representational similarity; number form area

1. Introduction

Modern societies around the world use written characters as tools for visually representing spoken communication. Besides having to master the writing script(s) used in one's native language (e.g., Latin alphabet, Chinese, Devanagari, and Arabic), any member of a technology-driven society needs to also master the Arabic numeral system comprising the digits 0 – 9. Although any written script can represent spoken number words (e.g., /siks/ can be represented using “six” or “VI” in Latin script), mastery of the Arabic numeral system is indispensable because its positional decimal notation allows numbers to be represented very efficiently (J. Zhang & Norman, 1995). Despite the ubiquity and importance of Arabic numerals, little is known about how the brain supports the processing of numerals as visual objects.

More than two decades ago, neuropsychological findings of patients with brain lesions led to the hypothesis of a “Number Form Area” (NFA) in the ventral occipito-temporal cortex (vOT) that supports processing of numerals (Cohen & Dehaene, 1991, 1995, 2000; Dehaene & Cohen, 1995). One technique to demonstrate the existence of this specialized region is to use functional magnetic resonance imaging (fMRI) to localize neuronal populations that respond preferentially (i.e., show greater activation) to numerals than to other similarly learned symbols such as letters. Evidence of an NFA has been elusive using fMRI despite the relatively robust localization of an analogous letter-preferring region in the left fusiform gyrus (Baker et al., 2007; Cohen & Dehaene, 2004; James, James, Jobard, Wong, & Gauthier, 2005; Park, Hebrank, Polk, & Park, 2012; Polk & Farah, 1998; Polk et al., 2002). Nonetheless, with intracranial electrophysiological recordings (Daitch et al., 2016; Shum et al., 2013) and intracranial cortical stimulation (Roux, Lubrano, Lauwers-Cances, Giussani, & Demonet, 2008), a numeral-preferring region has been successfully localized in the lateral posterior inferior temporal gyrus (pITG). Moreover, Shum and colleagues (2013) found that this numeral-preferring pITG region does not show the same degree of activation for stimuli with similar curvilinear features (e.g., letters, scrambled digits), semantics (number words, e.g., “one”), and phonology (similar-sounding non-number words, e.g., “won”). This led the authors to conclude that the region is tuned specifically to the overall visual form of the numerals,

1 rather than driven by the mere presence of constituent visual features (lines, curves, and angles), or by
2 top-down influences from regions involved in phonological or semantic processing.

3 The prior lack of evidence of a reproducible localization of the putative NFA using fMRI has
4 been attributed to possible dropout of fMRI signal because the candidate region in the pITG is more
5 lateral compared to the letter-preferring fusiform region and close to the air-tissue boundary (Abboud,
6 Maidenbaum, Dehaene, & Amedi, 2015; Grotheer, Herrmann, & Kovacs, 2016; Shum et al., 2013).
7 However, a growing body of evidence increasingly supports the existence of a numeral-preferring region
8 in the pITG, and suggests that the previous lack of reliable localization was likely due to modulation by
9 task demands rather than signal dropout (Amalric & Dehaene, 2016; Bugden, Woldorff, & Brannon,
10 2018; Grotheer, Jeska, & Grill-Spector, 2018; Pollack & Price, 2019; see Yeo, Wilkey, & Price, 2017, for
11 review). For instance, Grotheer and colleagues (2018) showed that, during a repetition detection (i.e., one-
12 back) task, regions in the bilateral pITG demonstrated numeral preference. They probed the function of
13 this region and showed that, during an addition task using numerals (e.g., $2 + 3 = 5?$), dice patterns, or
14 finger representations, these very same regions were not more engaged for addition with numerals than
15 addition with dice patterns and finger representations (Grotheer et al., 2018). The authors concluded that
16 the “numeral-preferring” pITG regions initially identified are not involved in processing the visual form
17 of the numerals because they should otherwise show consistent numeral preference regardless of the task.
18 Instead, Grotheer and colleagues (2018) hypothesize that the neuronal populations in the pITG
19 predominantly “ascribe numerical content to the visual input” (p. 188). More recently, Pollack and Price
20 (2019) found that a region in the left pITG was preferentially engaged for numerals when participants had
21 to detect a digit amongst a string of letters, but the same region showed no numeral preference when
22 participants had to detect a letter amongst a string of digits (i.e., when the digits were task-irrelevant).
23 Taken together, whether the candidate region for an NFA shows greater engagement for numerals than
24 other visual object categories may be highly dependent on attention to the stimulus identity and/or
25 category, as well as task contexts. However, there is also evidence that this region is involved in non-
26 quantitative contexts, such as whether a character is familiar or novel (Grotheer, Ambrus, & Kovács,

2016), whether a character is read aloud (Shum et al., 2013), or whether a character is repeated (Grotheer, Herrmann, et al., 2016; Grotheer et al., 2018). Thus, the exact computational mechanisms subserved by this numeral-preferring pITG region remain opaque.

In light of these new insights on the sensitivity of the numeral-preferring region in the pITG to task demands, we questioned whether 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 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 dataset), 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 (Cantlon, Pinel, Dehaene, & Pelphrey, 2011; Dehaene-Lambertz, Monzalvo, & Dehaene, 2018; Glezer, Jiang, & Riesenhuber, 2009; Karipidis et al., 2017; Kay & Yeatman, 2017; Parviainen, 2006; Pleisch et al., 2019; Polk et al., 2002; Vinckier et al., 2007; B. Zhang, He, & Weng, 2018), and is recommended for understanding models of experience-driven neural coding in the vOT (Dehaene & Cohen, 2011). Moreover, having a task that 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) 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. 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 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.

2. 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 region (hereafter, “pITG-numerals”) is derived from a meta-analysis of studies contrasting numerals and other symbols (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 & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008; 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 & Op de Beeck, 2016; Bracci, Ritchie, & Op de Beeck, 2017), and both types of information may be coded in pITG-numerals in terms of its sensitivity to shape and its structural and/or functional connectivity to parietal regions that are thought to subserve magnitude processing (Hannagan, Amedi,

Cohen, Dehaene-Lambertz, & Dehaene, 2015). 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., capable of 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).

3. Methods

3.1 Task and Datasets

3.1.1 Task. In each study, participants completed an identical fixation color change detection task (see symbol sets and example trials in Figure 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 Inline Supplementary Table S1 for more details). There are substantial differences in scan acquisition protocols and design parameters (e.g., additional factorial conditions examined) (see Section 3.3 Differences in task and neuroimaging acquisition parameters, and Inline Supplementary Table S1).

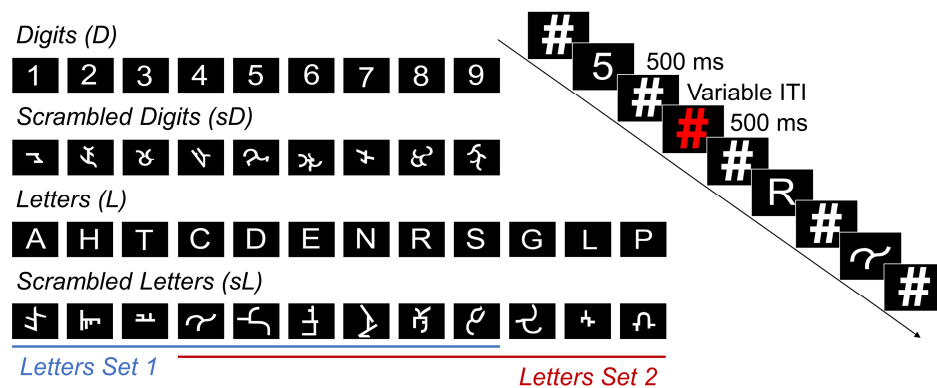


Figure 1. Stimulus sets and a schematic of the fixation color change detection task. The current study analyzed only the trials with these sets of stimuli that were only presented for 500 ms. ITI: Inter-stimulus interval. Letters Set 1 was used in Datasets 1 and 2. Letters Set 2 was used in Dataset 3.

[Insert Inline Supplementary Table S1 here]

3.1.2 Dataset 1 (Price & Ansari, 2011). Participants were 19 right-handed adults (6 females, mean age = 22.2 years, $SD = 1.7$, range = 20.5 – 27.2).

3.1.3 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, mean age = 17.8 years, $SD = 0.4$ years, range = 17.1 – 18.8 years, handedness: 3 left, 28 right, 1 unknown). As handedness was not a criterion for study eligibility, all participants regardless of handedness were included in the current analyses. Moreover, a recent study comparing left- and right-handed participants suggests no evidence of differences in the brain activation during passive viewing of numerals as a function of handedness (Goffin, Sokolowski, Slipenkyj, & Ansari, 2019). Seven additional participants were excluded due to head motion (see Section 3.1.5 Data Exclusion).

3.1.4 Dataset 3 (Merkley et al., 2019). Participants were 37 right-handed adults (26 females, mean age = 25.1 years, $SD = 5.9$, range = 18 – 39). 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 Section 3.1.5 Data Exclusion). In other words, 5 of 37 participants had usable data from only three runs.

3.1.5 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). Inline Supplementary Table S2 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.

[Insert Inline Supplementary Table S2 here]

3.2 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 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

1 set. The intact symbol sets were in Arial font (size 40), and the scrambled digits and letters were obtained
2 by segmenting and rearranging the parts into a unified, but novel curvilinear shape. The average visual
3 angles for each condition are reported in Inline Supplementary Table S3. Below, we provide further
4 characterization of the stimulus sets so as to rule out other low-level visual differences between any of
5 these categories that the pITG may be sensitive to.

6 Based on previous work (Schubert, 2017), we focused on two low-level visual parameters that
7 may underlie any categorical differences: luminance and perimetric complexity. Luminance was chosen
8 because Arial font is a proportional-width font – its characters do not take up the same horizontal space.
9 Hence, the digits set takes up less horizontal space than the letters set. Luminance was computed by
10 summing the intensity values of all pixels in each grayscale image. As the scrambled characters appeared
11 to be more visually complex than their intact counterparts, we wanted to quantify their complexity.
12 Perimetric complexity is commonly used to measure the size-invariant visual complexity of individual
13 characters (Pelli, Burns, Farell, & Moore-Page, 2006; Schubert, 2017; Shovman & Ahissar, 2006;
14 Ziegler, Pech-Georgel, Dufau, & Grainger, 2010), and has been shown to be highly correlated with the
15 efficiency of character identification and is mediated by the number of features (e.g., lines, curves,
16 terminations, etc.) (Pelli et al., 2006). We computed the perimetric complexity of each character using the
17 approach described by Pelli and colleagues (2006) using a custom MATLAB script: squared length of the
18 inner and outer perimeter divided by “inked” area of each shape traced from the binarized version of the
19 image. Pairwise comparisons of luminance and perimetric complexity showed that digits, letters and their
20 scrambled counterparts did not differ substantially in their perimetric complexity, however, digits on
21 average had lower luminance than letters in both Letters Sets 1 and 2 (see Inline Supplementary Table S3
22 – S4). Given the difference in luminance between the digits and letters sets, we directly assessed whether
23 the representational geometry in a region can be described by differences in luminance.

24 [Insert Inline Supplementary Tables S3 and S4 here]

25

26

3.3 Differences in Experimental Contexts and Neuroimaging Acquisition Parameters

Besides practical differences in MRI acquisition parameters with different scanner models (see Inline Supplementary Table S1), there are notable differences in the amount and nature of exposure to the stimuli.

3.3.1 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 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.

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

3.3.3 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, de Araujo Sanchez, & Adke, 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.

3.4 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 S1), 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.

3.5 Regions of Interest

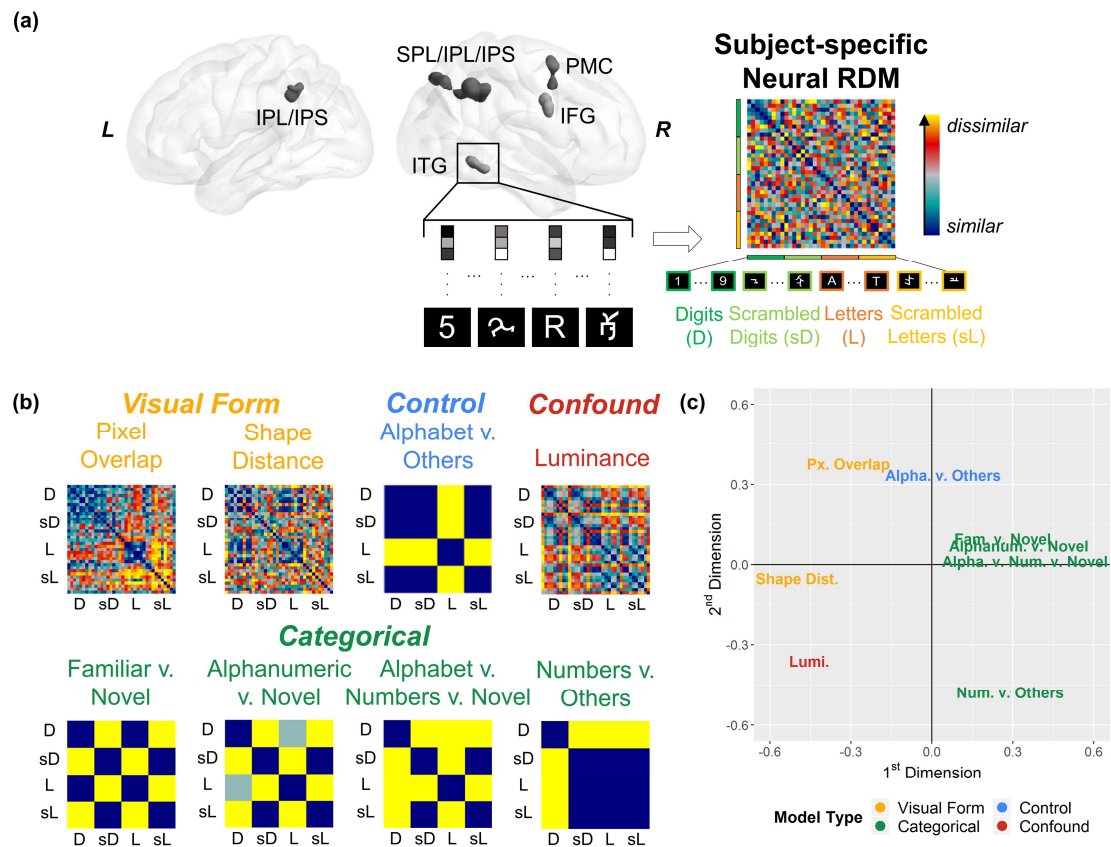
Regions of interest (ROIs) were obtained from a meta-analysis by Yeo and colleagues (2017) in which preferential activity to Arabic numerals than to other familiar symbols (e.g., Roman letters for English speakers or Chinese characters for Chinese speakers) was found to be convergent across 20

1 studies. Numeral preference was localized in the right ITG (55 3-mm isotropic voxels), as well as bilateral
2 parietal and right frontal regions (see Figure 2a and Supplementary Materials for more details of the
3 ROIs). For our a priori hypotheses, we focused on the cluster in the right ITG, as well as the left
4 homologue region because the left ITG also exhibits numeral preference (Amalric & Dehaene, 2016;
5 Bugden et al., 2018; Grotheer, Herrmann, et al., 2016; Grotheer et al., 2018; Pollack & Price, 2019; Roux
6 et al., 2008), but is possibly less robust to varying task contexts. Moreover, Pollack and Price (2019)
7 found that although the left (but not right) ITG showed, on average across participants, numeral
8 preference when detecting digits among letters, individual differences in the activation of the right (but
9 not left) ITG during digit detection correlated with calculation competence. To assess the specificity of
10 the ITG findings independent of correlated signal and/or noise across regions, as well as for
11 completeness, we also analyzed the representational geometries in the parietal and frontal regions and
12 reported these exploratory findings in the Supplemental Materials. Individual differences and group
13 means of the size and temporal signal-to-noise ratio in each ROI for each dataset are reported in Inline
14 Supplementary Figure S1.

15 [Insert Inline Supplementary Figure S1 here]

16

1



2

Figure 2. Regions of interest (ROIs), and neural and candidate representational dissimilarity models (RDMs). (a) Numeral-preferring ROIs from the meta-analysis by Yeo and colleagues (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 Inline Supplementary Figure S4 for Letters Set 2 from Dataset 3). All models presented are rescaled to [0, 1] for comparative visualization. (c) Multidimensional scaling plot of the correlational distance among the candidate models using Letters Set 1 (see Inline Supplementary Figure S4 for Letters Set 2 from Dataset 3).

[Insert Inline Supplementary Figure S4 here]

15

3.6 Representational Similarity Analyses

Figure 2a shows a schematic overview of the approach taken for the RSA.

3.6.1 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, Kim, Bandettini, & Kriegeskorte, 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, which are available on request. 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 data 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 2a).

3.6.2 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 2b).

3.6.2.1 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

$$d_{jk} = \frac{1}{\sqrt{N}} \sqrt{\sum_{x=1}^N [I_j(x) - I_k(x)]^2}, \text{ where } N \text{ is the number of pixels in the image, and } I_j(x) \text{ and } I_k(x) \text{ are}$$

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, Malik, & Puzicha, 2002) (see Supplemental Materials 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, Caramazza, & Peelen, 2015; Fairhall, Anzellotti, Pajtas, & Caramazza, 2011; Gotts, Milleville, Bellgowan, & Martin, 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 Inline Supplementary Figures S2 and S3.

[Insert Inline Supplementary Figures S2 and S3 here]

3.6.2.2 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 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.

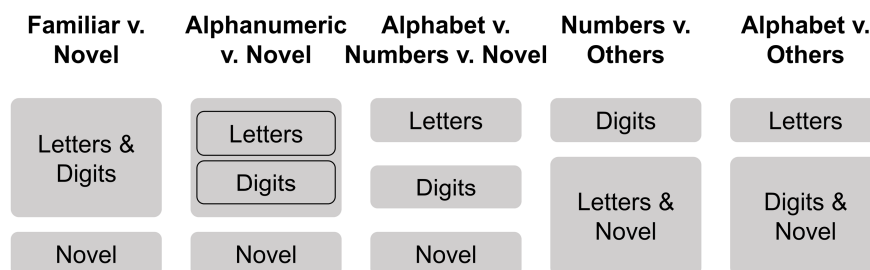


Figure 3. Schematic of the categorical and control models.

3.6.2.3 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). 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.

3.6.2.4 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 2c and Inline Supplementary Figure S4b show that the models are sufficiently different from one another, with each group of models roughly occupying separate quadrants in two-dimensional representational space (see Inline Supplementary Figure S4c for the pairwise rank correlations between the models).

3.6.3 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 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 (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.

3.6.4 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/>.

3.6.5 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, BF s tend to decrease with wider Cauchy priors, hence, all reported BF s 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 H_0 ($r_z = 0$) assuming that H_+ and H_0 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 H_1 (two-sided, r_z difference $\neq 0$) relative to H_0 ($r_z = 0$), to complement two-sided p -values.

Although we note that BF s 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).

3.6.6 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, Hall, & Rosenthal, 2016; Lakens & Etz, 2017; Valentine, Pigott, & Rothstein, 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 sound. 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).

3.7 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).

4. Results

4.1 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 4, and provide the detailed statistics only for the meta-analyses in Table 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.

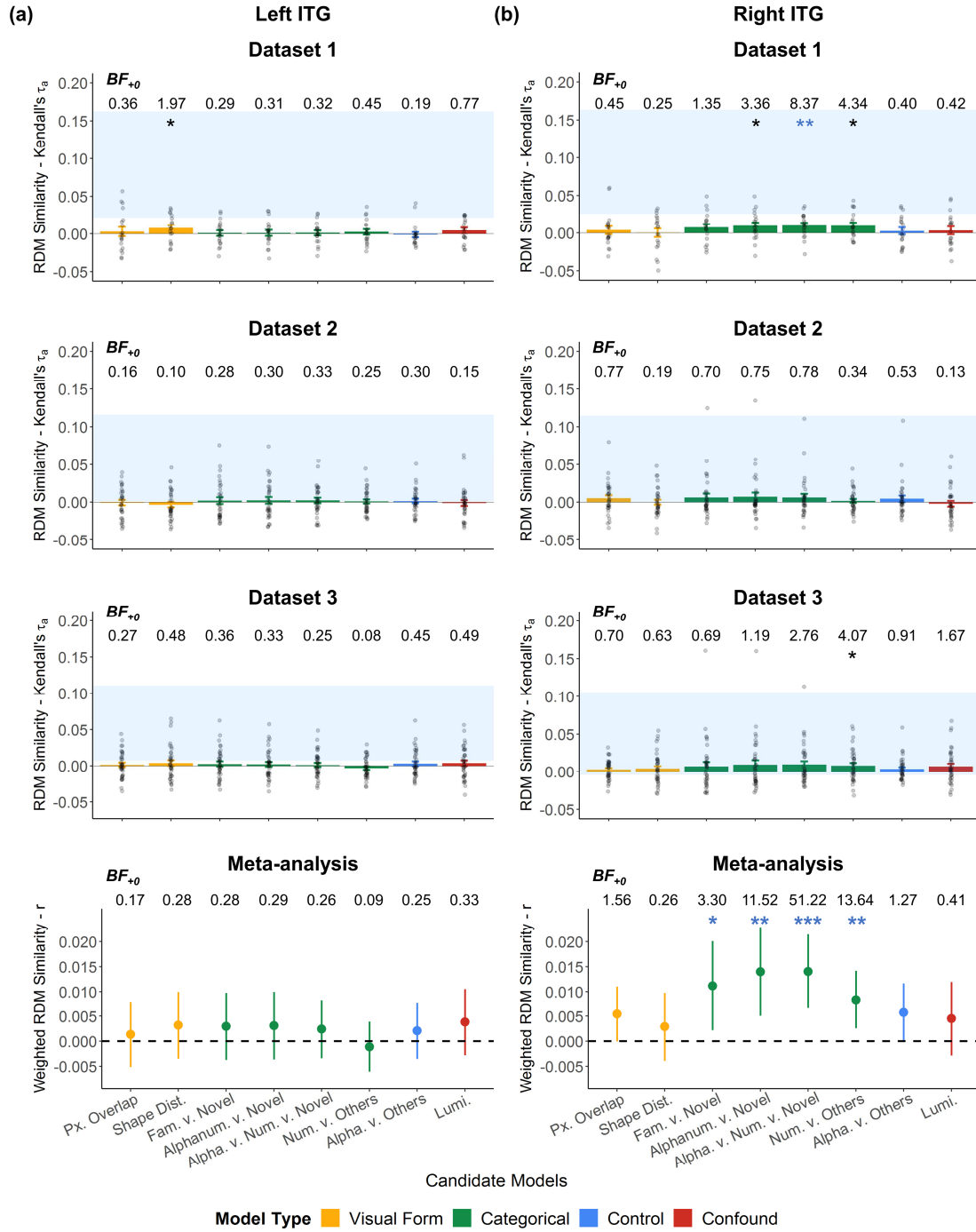


Figure 4. Similarity between neural and model representational dissimilarity matrices (RDMs) in the candidate numeral-prefering regions in (a) left and (b) right inferior temporal gyrus (ITG). Blue bars indicate the estimated noise ceiling. Group means and standard errors of the similarity are indicated 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 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)

4.1.1 Left ITG. Overall, there was conclusive meta-analytic evidence of a lack of similarity

between the neural RDMs and any model RDM (Figure 4a and Table 1). Below we report whether the meta-analytic findings were also observed in each dataset.

4.1.1.1 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 ($\tau_a = .0078$, $p = .044$, $BF_{+0} = 1.97$).

4.1.1.2 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.

4.1.2 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 4b and Table 1). The *BF* 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 *BF* 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). With error rates carefully controlled, mixed results across datasets could be interpreted as support for the compatibility between the Numbers v. Others model and the neural data (Lakens & Etz, 2017).

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 Inline Supplementary Table S5.

[Insert Inline Supplementary Table S5 here]

4.1.2.1 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.

4.1.2.2 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 p s > .828, BF_{10} s < 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 p s > .651, BF_{10} s < 0.31). Finally, it is important to note that the BF s 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 BF s 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} s > 1/3), suggesting that these results were not inconsistent with those of Dataset 1. These findings and the meta-analytic findings were qualitatively similar even when we restrict our analyses to only right-handed participants (i.e., $N = 28$) (see Inline Supplementary Figure S5 and Table S6).

[Insert Inline Supplementary Figure S5 and Table S6 here]

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.

We further assessed two models that describe weaker numeral sensitivity and letter sensitivity (see Inline Supplementary Figure S6), and found meta-analytic evidence for both models that were likely to be driven by an underlying three-way distinction among how numerals, letters, and novel characters are represented in the region. Nonetheless, the collective evidence is still consistent with the conclusion that this region distinguishes numerals from other character categories, and may even show a biased sensitivity towards numerals (see Inline Supplementary Figure S7 and Table S7).

[Insert Inline Supplementary Figures S6 and S7, and Table S7 here]

To also assess the possibility that over-representation of novel characters might bias the distinction between numerals and letters, we restricted our analyses to alphanumeric characters only (i.e., 18×18 RDMs) (see Inline Supplementary Figure S8). Although the substantial reduction of the size of the RDMs lowered the statistical power to detect any conclusive and robust evidence for the Numbers v. Alphabet model within each dataset, there was relatively robust meta-analytic evidence for the model (see Inline Supplementary Figure S9 and Table S8). Hence, it is unlikely that our findings are fully driven by the over-representation of novel characters in the main analyses.

[Insert Inline Supplementary Figures S8 and S9, and Table S8 here]

Figure 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 -medioids clustering analysis (Kaufman & Rousseuw, 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 Inline Supplementary Tables S9 – S11 and Figures S10 – S12).

[Insert Inline Supplementary Figures S10 – S12, and Tables S9 – S11 here]

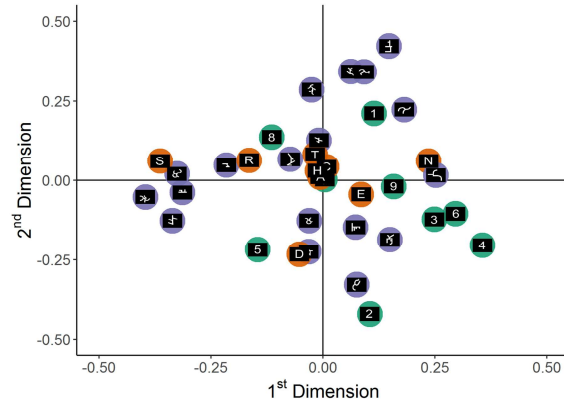
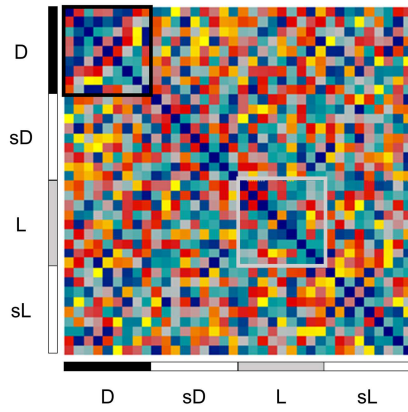
4.1.2.3 Control model. There was no evidence of similarity between the neural RDMs and Alphabet v. Others model RDM in any of the datasets.

- 1 **4.1.2.4 Confound model.** There was no evidence of similarity between the neural RDMs and
- 2 Luminance model RDM in any of the datasets.

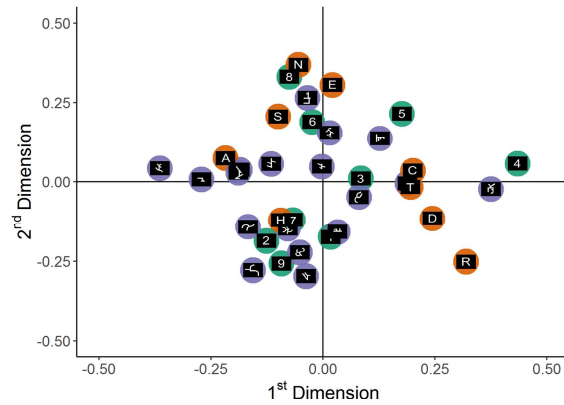
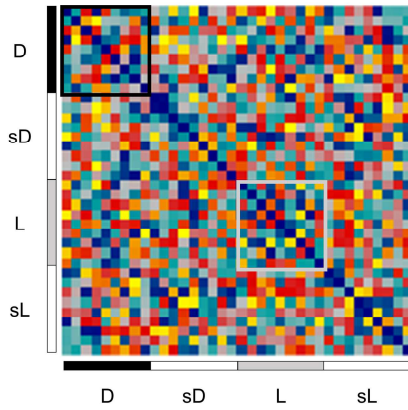
3

Right ITG

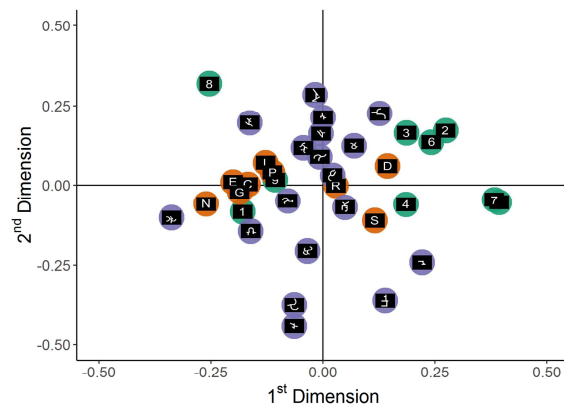
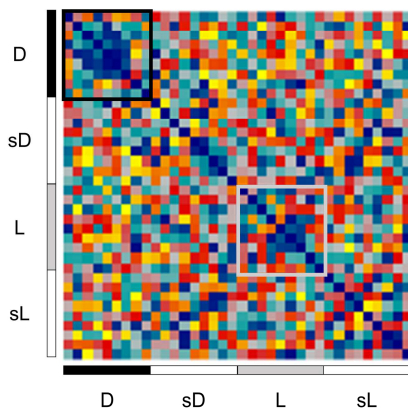
Dataset 1



Dataset 2



Dataset 3



similar dissimilar

Character Type ● Digits ● Letters ● Scrambled

Figure 5. 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/jwgk8/wiki/home>.

4.2 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 2a). 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 (Inline Supplementary Figures S13 – S14 and Table S12). 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 (Inline Supplementary Figures S15 – S16 and Table S13).

[Insert Inline Supplementary Figures S13 – S16, and Tables S12 – S13 here]

5. 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, spatial configurations, and written words (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 (Cohen & Dehaene, 2004; Flowers et al., 2004; James et al., 2005; Park et al., 2012; Polk & Farah, 1998; Polk et al., 2002; 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, Ambrus, et al., 2016; Grotheer et al., 2018). 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”).

5.1 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 conducted on 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 (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 (see Rothlein & Rapp, 2014, for a similar paradigm focusing on representations of letters).

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 estimate 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. Moreover, Lakens and Etz (2017) have demonstrated that “lines of studies with mixed results are relatively more likely when the H_1 is true than when the null hypothesis is true” (p. 880).

These results are also likely to be specific to the right pITG-numerals. We did not observe identical findings in any 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., 2018; Grotheer, Herrmann, et al., 2016; Grotheer et al., 2018), 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 (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, Chiang, Brannon, & Woldorff, 2014; Park, van den Berg, Chiang, Woldorff, & Brannon, 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

1 associating different recognition and computational goals with different categories of objects (for a recent
2 review, see Op de Beeck, Pillet, & Ritchie, 2019). It is therefore likely that literacy and numeracy lead to
3 the association of letters and numerals with habitually different goals (numerical and mathematical
4 relevance or not), which in turn lead to divergent neural processing pathways in the vOT that have a
5 preparatory or biased response for stimuli that potentially have numerical relevance or not. From this
6 perspective, not only does pITG-numerals encode subtle differences between visually similar objects,
7 such as numerals and letters, it could even encode similarities between objects that are visually dissimilar
8 (Gauthier, 2000), such as numerals, dice patterns, finger representations, mathematical formulas with
9 Greek and Roman letters (Amalric & Dehaene, 2016; Grotheer et al., 2018), or even from a different
10 sensory input, such as soundscapes associated with numerical content (Abboud et al., 2015), and auditory
11 mathematical statements (Amalric & Dehaene, 2016, 2017, 2019). In other words, it is likely that pITG-
12 numerals is recruited whenever the brain “predicts” that the stimulus has numerical relevance, through
13 automatic feedforward connections from posterior ventral (occipital cortex) regions and/or feedback
14 connections from parietal and frontal regions. This is also consistent with the interactive account that the
15 analogous “Visual Word Form Area” is involved in predictive coding through an experience-driven
16 automatic interaction of forward and backward connections, rather than word form detection per se (Price
17 & Devlin, 2011). In the current study, we are unable to disentangle the automatic, stimulus-driven
18 feedforward and feedback contributions, but only seek to exclude any modulatory contribution of
19 effortful, task-driven processing that may bias or amplify the representations along a particular dimension
20 (e.g., shape or conceptual domain). This exclusion is important because contemporary computational
21 models of category selectivity in the vOT suggest that at least for faces versus words, categorically
22 distinct representations can already be observed in category-selective vOT regions during passive
23 viewing, and those representations are further amplified by task-driven conceptual processing (Kay &
24 Yeatman, 2017).

25 Relatedly, it is noteworthy that the candidate numeral-preferring region in the right parietal lobule
26 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, Schel, & Klingberg, 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, Daitch, Pinheiro-Chagas, & Parvizi, 2018; Daitch et al., 2016; Pinheiro-Chagas, Daitch, Parvizi, & Dehaene, 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., 2018; 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.

5.2 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). However, it does not explain why pITG-numerals is spatially distinct from letter-preferring fusiform regions. Direct evidence for numeral visual form processing specifically has been lacking, because the shapes of digits have typically been contrasted with letters, false fonts, dice patterns, or finger representations, which are confounded by the object/conceptual category given that exemplars from the same category tend to have very similar shapes. 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. So, is the spatial dissociation of numeral-preferring and letter-preferring regions simply due to clustering of digit-shape and letter-shape “neural detectors”? This is unlikely as digits and letters are indistinguishable solely by the sets of features they comprise (Schubert, 2017). 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 in 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 “I”, “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 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”.

5.3 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 find 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 may 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 ($\tau_{as} < .02$; overall weighted $r = .008$) (Figure 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 ($\tau_{as} < .17$) 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.

6. 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

1 abstract shapes (i.e., not visual form detection per se). It is likely that literacy and numeracy experiences
2 may associate letters and digits with distinct processing goals (e.g., numerical relevance), and that the
3 numeral-preferring pITG is part of a neural pathway that has been developed for automatic processing
4 biases for stimuli with potential numerical relevance. In other words, “2” recruits the region because the
5 brain predicts based on past experiences that this character is likely to be numerically relevant.

Funding

This work was supported by the National Science Foundation grants (DRL 1660816 and DRL 1750213) awarded to G.R.P; the Humanities, Arts, and Social Sciences International PhD Scholarship (Nanyang Technological University and the Ministry of Education: Singapore) awarded to D.J.Y; the Canada First Research Excellence Fund award to BrainsCAN, a Brain Canada and NeuroDevNet Developmental Neuroscience Training award to R.M., and operating grants from the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), the Canada Research Chairs Program, and an E.W.R Steacie Memorial Fellowship from NSERC to D.A.

Acknowledgements

We are grateful to Benjamin Conrad for his assistance with quality assessment of Dataset 3 across neuroimaging packages and for his valuable comments on a previous version of the manuscript. We would also like to thank Armin Heinecke and Hester Berman of Brain Innovation, Inc., for their assistance with the implementation of analyses in BrainVoyager.

Conflict of Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Ethics Statement

For Dataset 1 (Price & Ansari, 2011) and Dataset 3 (Merkley et al., 2019), all participants gave informed consent and the research procedures were approved by the Health Sciences Research Ethics Board at the University of Western Ontario. For Dataset 2, all participants gave informed consent and the research procedures were approved by the Institutional Review Board at the Johns Hopkins University.

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>
- 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. (2017). 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(January), 19–31. <https://doi.org/10.1016/J.NEUROIMAGE.2019.01.001>
- 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. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 104(21), 9087–9092. <https://doi.org/10.1073/pnas.0703300104>
- 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>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful

- 1 approach to multiple testing. *Journal of the Royal Statistical Society*.
- 2 <https://doi.org/10.2307/2346101>
- 3 Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in
- 4 the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*, 143,
- 5 128–140. <https://doi.org/10.1016/j.neuroimage.2016.08.068>
- 6 Bracci, S., Caramazza, A., & Peelen, X. M. V. (2015). Representational Similarity of Body Parts in
- 7 Human Occipitotemporal Cortex. *Journal of Neuroscience*, 35(38), 12977–12985.
- 8 <https://doi.org/10.1523/JNEUROSCI.4698-14.2015>
- 9 Bracci, S., Daniels, N., & Op de Beeck, H. (2017). Task Context Overrides Object- and Category-Related
- 10 Representational Content in the Human Parietal Cortex. *Cerebral Cortex*, (January), 310–321.
- 11 <https://doi.org/10.1093/cercor/bhw419>
- 12 Bracci, S., & Op de Beeck, H. (2016). Dissociations and associations between shape and category
- 13 representations in the two visual pathways. *Journal of Neuroscience*, 36(2), 432–444.
- 14 <https://doi.org/10.1523/JNEUROSCI.2314-15.2016>
- 15 Bracci, S., Ritchie, J. B., & Op de Beeck, H. (2017). On the partnership between neural representations of
- 16 object categories and visual features in the ventral visual pathway. *Neuropsychologia*, 105(October
- 17 2016), 153–164. <https://doi.org/10.1016/j.neuropsychologia.2017.06.010>
- 18 Bugden, S., Woldorff, M. G., & Brannon, E. M. (2018). Shared and distinct neural circuitry for
- 19 nonsymbolic and symbolic double-digit addition. *Human Brain Mapping*, (October), hbm.24452.
- 20 <https://doi.org/10.1002/hbm.24452>
- 21 Cantlon, J. F., Pinel, P., Dehaene, S., & Pelphrey, K. A. (2011). Cortical Representations of Symbols,
- 22 Objects, and Faces Are Pruned Back during Early Childhood. *Cerebral Cortex*, 21(1), 191–199.
- 23 <https://doi.org/10.1093/cercor/bhq078>
- 24 Chouinard, P. A., Morrissey, B. F., Köhler, S., & Goodale, M. A. (2008). Repetition suppression in
- 25 occipital-temporal visual areas is modulated by physical rather than semantic features of objects.
- 26 *NeuroImage*, 41(1), 130–144. <https://doi.org/10.1016/j.neuroimage.2008.02.011>

- 1 Cohen, L., & Dehaene, S. (1991). Neglect Dyslexia for Numbers? A Case Report. *Cognitive*
- 2 *Neuropsychology*, 8(April), 39–58. <https://doi.org/10.1080/02643299108253366>
- 3 Cohen, L., & Dehaene, S. (1995). Number processing in pure alexia: The effect of hemispheric
- 4 asymmetries and task demands. *Neurocase*, 1(2), 121–137.
- 5 <https://doi.org/10.1080/13554799508402356>
- 6 Cohen, L., & Dehaene, S. (1996). Cerebral networks for number processing: Evidence from a case of
- 7 posterior callosal lesion. *Neurocase*, 2(November), 155–174.
- 8 <https://doi.org/10.1080/13554799608402394>
- 9 Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure
- 10 alexia. *Cognitive Neuropsychology*, 17(6), 563–583. <https://doi.org/10.1080/02643290050110656>
- 11 Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word
- 12 form area. *NeuroImage*, 22(1), 466–476. <https://doi.org/10.1016/j.neuroimage.2003.12.049>
- 13 Daitch, A. L., Foster, B. L., Schrouff, J., Rangarajan, V., Kaşıkçı, I., Gattas, S., & Parvizi, J. (2016).
- 14 Mapping human temporal and parietal neuronal population activity and functional coupling during
- 15 mathematical cognition. *Proceedings of the National Academy of Sciences*, 113(46), E7277–E7286.
- 16 <https://doi.org/10.1073/pnas.1608434113>
- 17 Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). *The emergence of the visual word form:*
- 18 *Longitudinal evolution of category-specific ventral visual areas during reading acquisition. PLOS*
- 19 *Biology* (Vol. 16). <https://doi.org/10.1371/journal.pbio.2004103>
- 20 Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing.
- 21 *Mathematical Cognition*, 1(1), 83–120.
- 22 Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in*
- 23 *Cognitive Sciences*, 15(6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>
- 24 Diedrichsen, J., & Kriegeskorte, N. (2017). *Representational models: A common framework for*
- 25 *understanding encoding, pattern-component, and representational-similarity analysis. PLoS*
- 26 *Computational Biology* (Vol. 13). <https://doi.org/10.1371/journal.pcbi.1005508>

- 1 Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*,
2 5(July), 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- 3 Dienes, Z., & Mclatchie, N. (2017). Four reasons to prefer Bayesian analyses over significance testing.
4 *Psychonomic Bulletin and Review*, 1–12. <https://doi.org/10.3758/s13423-017-1266-z>
- 5 Fairhall, S. L., Anzellotti, S., Pajtas, P. E., & Caramazza, A. (2011). Concordance between perceptual and
6 categorical repetition effects in the ventral visual stream. *Journal of Neurophysiology*, 106(1), 398–
7 408. <https://doi.org/10.1152/jn.01138.2010>
- 8 Flowers, D. L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T. A., Wood, F. B., & Eden, G. F. (2004).
9 Attention to single letters activates left extrastriate cortex. *NeuroImage*, 21(3), 829–839.
10 <https://doi.org/10.1016/j.neuroimage.2003.10.002>
- 11 Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex? *Trends in Cognitive*
12 *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)
- 13 Gilpin, A. R. (1993). Table for Conversion of Kendall's Tau to Spearman's Rho Within the Context of
14 Measures of Magnitude of Effect for Meta-Analysis. *Educational and Psychological Measurement*,
15 53(1), 87–92. <https://doi.org/10.1177/0013164493053001007>
- 16 Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for Highly Selective Neuronal Tuning to
17 Whole Words in the “Visual Word Form Area.” *Neuron*, 62(2), 199–204.
18 <https://doi.org/10.1016/j.neuron.2009.03.017>
- 19 Glezer, L. S., & Riesenhuber, M. (2013). Individual Variability in Location Impacts Orthographic
20 Selectivity in the “Visual Word Form Area.” *Journal of Neuroscience*, 33(27), 11221–11226.
21 <https://doi.org/10.1523/JNEUROSCI.5002-12.2013>
- 22 Goffin, C., Sokolowski, H. M., Slipenkyj, M., & Ansari, D. (2019). Does writing handedness affect
23 neural representation of symbolic number? An fMRI adaptation study. *Cortex*, 121, 27–43.
24 <https://doi.org/10.1016/j.cortex.2019.07.017>
- 25 Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some
26 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>
- 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)
- 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(March), 188–200. <https://doi.org/10.1016/j.neuroimage.2018.03.064>
- 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>
- 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>
- 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>
- JASP Team. (2019). JASP (Version 0.10.0) [Computer software]. Retrieved from <https://jasp-stats.org/>

- 1 Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional
2 architecture of the mind. *Proceedings of the National Academy of Sciences*, 107(25), 11163–11170.
3 <https://doi.org/10.1073/pnas.1005062107>
- 4 Karipidis, I. I., Pleisch, G., Röthlisberger, M., Hofstetter, C., Dornbierer, D., Stämpfli, P., & Brem, S.
5 (2017). Neural initialization of audiovisual integration in prereaders at varying risk for
6 developmental dyslexia. *Human Brain Mapping*, 38(2), 1038–1055.
7 <https://doi.org/10.1002/hbm.23437>
- 8 Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective
9 cortex. *ELife*, 6, 1–29. <https://doi.org/10.7554/elife.22341>
- 10 Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients.
11 *Communications for Statistical Applications and Methods*, 22(6), 665–674.
12 <https://doi.org/10.5351/CSAM.2015.22.6.665>
- 13 Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation,
14 and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412.
15 <https://doi.org/10.1016/j.tics.2013.06.007>
- 16 Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the
17 branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2(November), 4.
18 <https://doi.org/10.3389/neuro.06.004.2008>
- 19 Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008).
20 Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey.
21 *Neuron*, 60(6), 1126–1141. <https://doi.org/10.1016/j.neuron.2008.10.043>
- 22 Lakens, D., & Etz, A. J. (2017). Too True to be Bad: When Sets of Studies With Significant and
23 Nonsignificant Findings Are Probably True. *Social Psychological and Personality Science*, 8(8),
24 875–881. <https://doi.org/10.1177/1948550617693058>
- 25 Lochy, A., & Schiltz, C. (2019). Lateralized Neural Responses to Letters and Digits in First Graders.
26 *Child Development*, 90(6), 1866–1874. <https://doi.org/10.1111/cdev.13337>

- 1 Lyons, I. M., & Beilock, S. L. (2018). Characterizing the neural coding of symbolic quantities.
2 *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2018.05.062>
- 3 Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumati, R. I., Caramazza, A., & Martin, A. (2007).
4 Action-Related Properties Shape Object Representations in the Ventral Stream. *Neuron*, 55(3), 507–
5 520. <https://doi.org/10.1016/j.neuron.2007.07.011>
- 6 Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. (2018). Integrative and
7 distinctive coding of visual and conceptual object features in the ventral visual stream. *ELife*, 7,
8 e31873. <https://doi.org/10.7554/eLife.31873>
- 9 Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics
10 learning disability in the primary school-age years. *Annals of Dyslexia*, 53(1), 218–253.
11 <https://doi.org/10.1007/s11881-003-0011-7>
- 12 Merkley, R., Conrad, B., Price, G., & Ansari, D. (2019). Investigating the visual number form area: a
13 replication study. *Royal Society Open Science*, 6(10), 182067. <https://doi.org/10.1098/rsos.182067>
- 14 Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers
15 and response normalizations for pattern-information fMRI. *NeuroImage*, 53(1), 103–118.
16 <https://doi.org/10.1016/j.neuroimage.2010.05.051>
- 17 Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs.
18 R package version 0.9.12-4.2. Retrieved from <https://cran.r-project.org/package=BayesFactor>
- 19 Nemmi, F., Schel, M. A., & Klingberg, T. (2018). Connectivity of the Human Number Form Area
20 Reveals Development of a Cortical Network for Mathematics. *Frontiers in Human Neuroscience*,
21 12(November), 1–15. <https://doi.org/10.3389/fnhum.2018.00465>
- 22 Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox
23 for Representational Similarity Analysis. *PLoS Computational Biology*, 10(4), e1003553.
24 <https://doi.org/10.1371/journal.pcbi.1003553>
- 25 Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt
26 multivariate fMRI analyses? *NeuroImage*, 49(3), 1943–1948.

- 1 <https://doi.org/10.1016/j.neuroimage.2009.02.047>
- 2 Op de Beeck, H. P., Pillet, I., & Ritchie, J. B. (2019). Factors Determining Where Category-Selective
- 3 Areas Emerge in Visual Cortex. *Trends in Cognitive Sciences*, 23(9), 784–797.
- 4 <https://doi.org/10.1016/j.tics.2019.06.006>
- 5 Op de Beeck, H. P., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar
- 6 objects and the organization of the human object vision pathway. *The Journal of Neuroscience : The*
- 7 *Official Journal of the Society for Neuroscience*, 28(40), 10111–10123.
- 8 <https://doi.org/10.1523/JNEUROSCI.2511-08.2008>
- 9 Park, J., Chiang, C., Brannon, E. M., & Woldorff, M. G. (2014). Experience-dependent Hemispheric
- 10 Specialization of Letters and Numbers Is Revealed in Early Visual Processing. *Journal of Cognitive*
- 11 *Neuroscience*, 26(10), 2239–2249. https://doi.org/10.1162/jocn_a_00621
- 12 Park, J., Hebrank, A., Polk, T. A., & Park, D. C. (2012). Neural Dissociation of Number from Letter
- 13 Recognition and Its Relationship to Parietal Numerical Processing. *Journal of Cognitive*
- 14 *Neuroscience*, 24(1), 39–50. https://doi.org/10.1162/jocn_a_00085
- 15 Park, J., van den Berg, B., Chiang, C., Woldorff, M. G., & Brannon, E. M. (2018). Developmental
- 16 trajectory of neural specialization for letter and number visual processing. *Developmental Science*,
- 17 21(3), 1–14. <https://doi.org/10.1111/desc.12578>
- 18 Parviainen, T. (2006). Cortical Sequence of Word Perception in Beginning Readers. *Journal of*
- 19 *Neuroscience*, 26(22), 6052–6061. <https://doi.org/10.1523/jneurosci.0673-06.2006>
- 20 Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter
- 21 identification. *Vision Research*, 46(28), 4646–4674. <https://doi.org/10.1016/j.visres.2006.04.023>
- 22 Pinheiro-Chagas, P., Daitch, A., Parvizi, J., & Dehaene, S. (2018). Brain Mechanisms of Arithmetic: A
- 23 Crucial Role for Ventral Temporal Cortex. *Journal of Cognitive Neuroscience*, 30(12), 1757–1772.
- 24 https://doi.org/10.1162/jocn_a_01319
- 25 Pleisch, G., Karipidis, I. I., Brauchli, C., Röthlisberger, M., Hofstetter, C., Stämpfli, P., ... Brem, S.
- 26 (2019). Emerging neural specialization of the ventral occipitotemporal cortex to characters through

- 1 phonological association learning in preschool children. *NeuroImage*, 189, 813–831.
- 2 <https://doi.org/10.1016/j.neuroimage.2019.01.046>
- 3 Polk, T. A., & Farah, M. J. (1998). The neural development and organization of letter recognition:
- 4 Evidence from functional neuroimaging, computational modeling, and behavioral studies.
- 5 *Proceedings of the National Academy of Sciences*, 95(3), 847–852.
- 6 <https://doi.org/10.1073/pnas.95.3.847>
- 7 Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D’Esposito, M., Detre, J. A., & Farah, M. J.
- 8 (2002). Neural Specialization for Letter Recognition. *Journal of Cognitive Neuroscience*, 14(2),
- 9 145–159. <https://doi.org/10.1162/089892902317236803>
- 10 Pollack, C., & Price, G. R. (2019). Neurocognitive mechanisms of digit processing and their relationship
- 11 with mathematics competence. *NeuroImage*, 185(May 2018), 245–254.
- 12 <https://doi.org/10.1016/j.neuroimage.2018.10.047>
- 13 Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to
- 14 reading. *Trends in Cognitive Sciences*, 15(6), 246–253. <https://doi.org/10.1016/j.tics.2011.04.001>
- 15 Price, G. R., & Ansari, D. (2011). Symbol processing in the left angular gyrus: Evidence from passive
- 16 perception of digits. *NeuroImage*, 57(3), 1205–1211.
- 17 <https://doi.org/10.1016/j.neuroimage.2011.05.035>
- 18 Price, G. R., Mazzocco, M. M. M., & Ansari, D. (2013). Why Mental Arithmetic Counts: Brain
- 19 Activation during Single Digit Arithmetic Predicts High School Math Scores. *Journal of*
- 20 *Neuroscience*, 33(1), 156–163. <https://doi.org/10.1523/JNEUROSCI.2936-12.2013>
- 21 R Core Team. (2018). R: A Language and Environment for Statistical Computing [Computer software].
- 22 Vienna, Austria. Retrieved from <https://cran.r-project.org/>
- 23 Rothlein, D., & Rapp, B. (2014). The similarity structure of distributed neural responses reveals the
- 24 multiple representations of letters. *NeuroImage*, 89, 331–344.
- 25 <https://doi.org/10.1016/j.neuroimage.2013.11.054>
- 26 Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem’s ESP claim. *Psychonomic*

- Bulletin and Review*, 18(4), 682–689. <https://doi.org/10.3758/s13423-011-0088-7>
- 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>
- 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>
- 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., ... 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>
- 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>
- 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>
- 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.

1 <https://doi.org/10.1016/j.neuroimage.2015.12.012>

2 Wilkey, E. D., Barone, J. C., Mazzocco, M. M. M., Vogel, S. E., & Price, G. R. (2017). The effect of
3 visual parameters on neural activation during nonsymbolic number comparison and its relation to
4 math competency. *NeuroImage*, 159(August), 430–442.

5 <https://doi.org/10.1016/j.neuroimage.2017.08.023>

6 Yeo, D. J., Wilkey, E. D., & Price, G. R. (2017). The search for the number form area: A functional
7 neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 78(April), 145–160.

8 <https://doi.org/10.1016/j.neubiorev.2017.04.027>

9 Zeithamova, D., de Araujo Sanchez, M. A., & Adke, A. (2017). Trial timing and pattern-information
10 analyses of fMRI data. *NeuroImage*, 153(November 2016), 221–231.

11 <https://doi.org/10.1016/j.neuroimage.2017.04.025>

12 Zhang, B., He, S., & Weng, X. (2018). Localization and functional characterization of an occipital visual
13 word form sensitive area. *Scientific Reports*, 8(1), 1–11. [https://doi.org/10.1038/s41598-018-25029-](https://doi.org/10.1038/s41598-018-25029-z)
14 [z](https://doi.org/10.1038/s41598-018-25029-z)

15 Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, 57(3),
16 271–295. [https://doi.org/10.1016/0010-0277\(95\)00674-3](https://doi.org/10.1016/0010-0277(95)00674-3)

17 Ziegler, J. C., Pech-Georgel, C., Dufau, S., & Grainger, J. (2010). Rapid processing of letters, digits and
18 symbols: What purely visual-attentional deficit in developmental dyslexia? *Developmental Science*,
19 13(4), 8–14. <https://doi.org/10.1111/j.1467-7687.2010.00983.x>

20

21

Yeo, Pollack, Merkley, Ansari, and Price (2019)

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

--