



# Representations within the Intraparietal Sulcus Distinguish Numerical Tasks and Formats

Griffin E. Koch<sup>ID</sup>, Melissa E. Libertus, Julie A. Fiez, and Marc N. Coutanche

## Abstract

■ How does our brain understand the number *five* when it is written as an Arabic numeral, and when presented as five fingers held up? Four facets have been implicated in adult numerical processing: semantic, visual, manual, and phonological/verbal. Here, we ask how the brain represents each, using a combination of tasks and stimuli. We collected fMRI data from adult participants while they completed our novel “four number code” paradigm. In this paradigm, participants viewed one of two stimulus types to tap into the visual and manual number codes, respectively. Concurrently, they completed one of two tasks to tap into the semantic and phonological/verbal number codes, respectively.

Classification analyses revealed that neural codes representing distinctions between the number comparison and phonological tasks were generalizable across format (e.g., Arabic numerals to hands) within intraparietal sulcus (IPS), angular gyrus, and precentral gyrus. Neural codes representing distinctions between formats were generalizable across tasks within visual areas such as fusiform gyrus and calcarine sulcus, as well as within IPS. Our results identify the neural facets of numerical processing within a single paradigm and suggest that IPS is sensitive to distinctions between semantic and phonological/verbal, as well as visual and manual, facets of number representations. ■

## INTRODUCTION

Numerical information can be conveyed through many means. For example, it can take both visual and verbal forms as in the symbol 5 written on a whiteboard in a classroom and the teacher labeling the symbol by saying the word *five*. In addition, numerical information carries semantic weight through its representation of quantity, such as determining that “5” represents a greater quantity than “3.” This is not always the case, however. For example, a student could be assigned to Group “5” and just as easily be assigned to the “green” or “E” group. Varied differences in the format of numerical information lead to questions about how our brains distinctly or concurrently organize representations of such information.

The three previously described aspects of numerical information have traditionally been studied in conjunction with one another, together forming the *triple-code model* (Dehaene, 1992). The three codes include verbal and phonological components of the names of numbers (Dehaene, 1992), a visual code (oftentimes as an Arabic numeral; Abboud, Maidenbaum, Dehaene, & Amedi, 2015), and a semantic component of quantity (Hubbard, Piazza, Pinel, & Dehaene, 2005) that is used during estimation and comparison tasks (Dehaene, 1992). Typically, a combination of dot stimuli (representing quantities) and number comparison tasks are used to better understand this semantic component.

The triple-code model has been useful for framing our understanding of the neural regions involved in complex

number-related tasks. It presents three individual neural codes relating to numerical processing (Myers & Szücs, 2015), conveying the idea that these isolated functions are subserved by distinct neural underpinnings (Skagenholt, Träff, Västfjäll, & Skagerlund, 2018). The intraparietal sulcus (IPS) and (less commonly) insula have been implicated in calculation and the semantic code (Arsalidou, Pawliw-Levac, Sadeghi, & Pascual-Leone, 2018; Skagenholt et al., 2018; Arsalidou & Taylor, 2011), whereas the fusiform, middle temporal, and inferior temporal gyri have been associated with the visual code (Grotheer, Ambrus, & Kovács, 2016; Grotheer, Herrmann, & Kovács, 2016), and language areas within the left hemisphere and left angular gyrus (Dehaene, Piazza, Pinel, & Cohen, 2003), as well as the inferior frontal gyrus (IFG; Schmithorst & Brown, 2004), have been associated with the verbal code. In line with the theories of the triple-code model, some recent work has identified notation and format independence for numerical magnitude representations (for a review, see Knops, 2017). On the one hand, some fMRI studies showed that viewing digits generates neural patterns of activity, especially in IPS, that are similar to those generated while viewing quantities of dots (Eger et al., 2009; Piazza, Pinel, Le Bihan, & Dehaene, 2007), while another fMRI study suggested that notation independence may be limited to the left IPS (Kadosh, Kadosh, Kaas, Henik, & Goebel, 2007). On the other hand, other studies question the notion of format independence (Cohen Kadosh, 2008) and suggest that neural representations for different number formats may be distinct (Lyons & Beilock, 2018; Bulthé, De Smedt, & Op de Beeck, 2014, 2015; Lyons, Ansari, & Beilock, 2015; Shuman & Kanwisher, 2004).

An additional fourth facet, a manual code, could potentially play an important functional role in providing a symbolic representation of number (Di Luca & Pesenti, 2011; Butterworth, 1999). In keeping with the previous examples, the teacher might hold up five fingers to signal that students have 5 minutes remaining to finish their exams, showing how fingers can signify quantities or values. In the same way that the components within the triple-code model are subserved by neural codes, so is the proposed manual code. The precentral gyrus is involved in representing number as part of this potential manual code (Kaufmann et al., 2008), and others have further identified areas within primary somatosensory cortex as representing digits (Schweisfurth, Frahm, Farina, & Schweizer, 2018). This manual code may plausibly be considered an addition to the triple-code model and is one of the oldest methods used for counting (Andres & Pesenti, 2015). Nevertheless, there is still much we do not know about how the brain represents and processes this type of information, for example, whether hands are processed similarly to nonsymbolic or symbolic numerical stimuli or constitute a separate form of number representation altogether. Furthermore, more research must be done to better understand whether intentional use of fingers, such as for counting, is a requisite for elucidating neural representations of numerical processing.

Although the triple-code model provides a conceptual lens through which to view the facets of numerical processing, to the best of our knowledge, no studies have intentionally investigated all four facets using a single paradigm. Instead, studies have typically targeted neural regions supporting numerical processing more generally or focused on a limited subset of facets. One particular paradigm has been used extensively in recent years to localize number processing regions both in adults and in children (Emerson & Cantlon, 2015; Cantlon & Li, 2013). This paradigm identifies regions in the brain that are more active when people compare dot quantities and Arabic numerals (relative to other conditions), thus providing a glimpse into the neural underpinnings of a combination of visual and semantic facets of the number-processing network. However, such a task does not allow the dissociation of these facets or a comparison to the verbal and manual facets. Recently, a first attempt has been made to explore the three facets of the triple-code model in the adult brain using a single paradigm (Skagenholt et al., 2018) in which participants decided which of two presented dot quantities, Arabic numerals, or number words was larger in quantity. The results supported the neural correlates of the triple-code model previously identified through separate paradigms, specifically the right middle temporal gyrus in representing Arabic numerals, left IFG in representing verbal number processing, and right IPS in processing of semantic number information (Skagenholt et al., 2018). Although this study utilized a novel method to provide support for the triple-code model, the method does not entirely discern the differences between the codes. For

example, in the number word condition, participants are still required to access the semantic facet when performing the quantity judgment. Thus, the task does not separate the phonological/verbal aspect of the number word and the semantic quantity it represents.

Here, we ask how the four facets of number are represented in the brain. In addition, although the triple-code model and recent research into a potential manual code usefully identify neural regions involved in their respective codes, little is known about the underlying neural representations that give rise to such numerical processing. Importantly, we draw on a forward inference point of view to ask how neural representations of number might remain constant across a variety of formats and tasks.

To accomplish this, we utilize a novel paradigm (hereafter referred to as the *four number code [4NC] paradigm*) that incorporates a combination of tasks and formats to efficiently investigate all four facets of numerical processing (visual, phonological/verbal, semantic, and manual). In this 4NC paradigm, participants viewed one of two formats indicating a quantity: Arabic numerals or hands, to tap into the visual and manual number codes, respectively. At the same time, they completed one of two tasks: deciding if a quantity was greater than another number or if the word denoting the quantity contained a long vowel sound, to tap into the semantic and phonological/verbal number codes, respectively. Participants performed the 4NC paradigm while undergoing an fMRI scan.

The current study addresses three key questions relating to representations of numerical information: First, how does the brain represent task information, and are these representations different between the semantic and phonological/verbal codes? Instead of more traditional approaches that use dot and verbal stimuli, here, we employ quantity and phonological comparison tasks consisting of numerical stimuli to better understand these representations. Second, how does the brain represent information about format, and are these representations different between Arabic numerals and hand stimuli? And finally, is it possible for a neural region to contain distinct representations of both task and visual format information? If so, can we delve deeper and dissociate how those representations are anatomically organized? Through a multivariate analysis generalization framework (Coutanche & Thompson-Schill, 2015), we provide new evidence of the types of representations stored within the brain that draw upon the four facets of number.

## METHODS

### Participants

Twenty-two right-handed, native English speakers without a learning disorder were recruited. Two participants' data were excluded from further analyses: one for excessive head movement during the scan (overall movement > 5 mm; Emerson & Cantlon, 2015; Cantlon & Li, 2013), one for

not following instructions (verbally indicated misunderstanding of stimuli comparison). In addition, the fifth (and final) functional run was excluded from two otherwise usable participants: one for excessive head movement (between-images displacement within the run exceeded the resolution of one functional voxel: 3.125 mm) and one for a technical disruption during the run. Based on prior sample sizes for studies examining a similar effect (Emerson & Cantlon, 2015; Cantlon & Li, 2013), we concluded recruitment once 20 participants' fMRI data were obtained and deemed usable. Of these 20 participants, two participants' data had to be removed because of poor behavioral performance (i.e., mean accuracies on the judgment tasks more than 2 *SDs* below the mean of the group). Thus, all analyses reported below were conducted on the remaining 18 participants (11 women, 7 men, *M* age = 22.5 years, *SD* = 4.1 years). Participants provided written consent and were compensated for their participation. The institutional review board approved all measures before data collection. Data presented here will be made available upon reasonable request that complies with policies of the institutional review board.

### Stimuli, Task, and Procedure

Before beginning the scanning session, participants completed a long vowel practice task in which they identified words containing a long vowel (i.e., selecting *ape* out of a list containing act, ape, bed, and dig). This practice task ensured that participants were familiar (and comfortable) with making judgments relating to long and short vowels. After discussing the safety procedures, participants underwent an anatomical scan, followed by five functional runs of the 4NC paradigm. Interleaved within the functional runs were four runs of an additional numerical judgment paradigm (see Appendix for full details and analyses). The session ended with three functional resting-state runs (not analyzed here).

Participants viewed stimuli consisting of values from 1 to 9 represented as either Arabic numerals or hands. See Table 1 for additional details concerning average number of trials for each value in each condition. Participants were first presented with instructions indicating their task for the subsequent three trials: either making a numerical or phonological judgment. For *numerical judgment* trials, participants judged whether the presented quantity was greater than the value presented during the instructions (e.g., "Is the quantity greater than 6?"). Across different blocks, the comparison value changed between 3, 4, 6, and 7. For *phonological judgment* trials, participants judged whether the presented quantity's name contained a long vowel sound (i.e., "Does the quantity's name contain a long vowel sound?"). For both judgments, participants indicated their response to each trial by pressing the left or right button (counterbalanced across participants). The yes/no button press assignment for the 4NC paradigm was

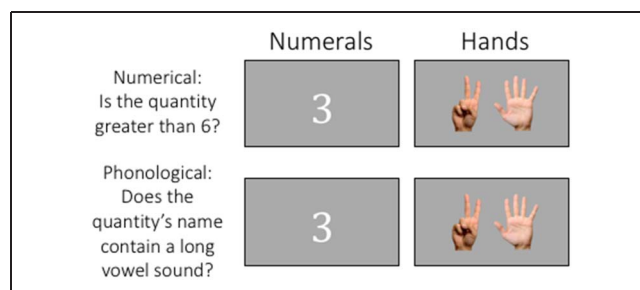
**Table 1.** Number of Trials for Each of the Nine Possible Quantity Values Used in the 4NC Paradigm for Each of the Four Conditions, across All Five Runs

	<i>Phonological Judgment</i>		<i>Numerical Judgment</i>	
	<i>Hands</i>	<i>Numerals</i>	<i>Hands</i>	<i>Numerals</i>
1	2.75 (1.77)	3.06 (1.18)	2.63 (1.20)	5.06 (1.12)
2	3.69 (1.85)	2.69 (1.14)	3.00 (1.15)	3.56 (1.03)
3	3.63 (1.59)	4.06 (1.48)	2.25 (1.61)	3.19 (1.38)
4	3.19 (1.72)	3.75 (1.81)	3.25 (1.84)	2.94 (1.34)
5	3.69 (1.35)	4.25 (1.98)	2.38 (1.02)	3.13 (1.41)
6	3.06 (1.77)	3.06 (1.81)	3.13 (1.02)	4.38 (1.54)
7	3.19 (1.38)	3.50 (1.51)	3.69 (1.45)	2.75 (1.34)
8	3.44 (0.96)	2.50 (1.55)	4.94 (1.73)	2.75 (1.18)
9	3.38 (1.41)	3.13 (1.67)	4.75 (1.53)	2.25 (0.86)

Means and standard deviations (in parentheses) were calculated from all participants who completed all five runs of usable data. There were 30 trials per condition across the five runs for each participant.

consistent with the match/nonmatch button assignment throughout the scanning session such that yes and match responses were always paired. Approximately 50% of the trials required positive responses (i.e., greater quantity than the value in the instructions; containing a long vowel), whereas the others required negative responses. Our 2 × 2 factorial design resulted in four different conditions presented to participants: numerical judgment presented in numeral format; numerical judgment presented in hand format; phonological judgment presented in numeral format; and phonological judgment presented in hand format, shown in Figure 1.

Stimuli were presented in a blocked design across five functional runs. All runs consisted of two blocks per condition, each consisting of three judgment trials (same condition). Blocks were pseudorandomized such that all conditions were presented once before a second presentation of any condition. Within each block, there was at least one trial requiring a positive response and one requiring a negative response. Judgment trials were presented for 2 sec, followed by a 2-sec intertrial interval (fixation cross). Each block was followed by 8 sec of instructions, informing participants of the condition



**Figure 1.** Stimulus conditions presented as part of the 4NC paradigm.

(either numerical or phonological judgment) for the next block of trials.

All experimental stimuli were presented using MATLAB (R2017a) and the Psychophysics Toolbox Version 3 (Kleiner et al., 2007; Brainard, 1997), which synchronized stimulus onset times with fMRI data acquisition.

### Behavioral Analyses

The behavioral data were analyzed by comparing accuracy as well as RTs for each of the four format/task conditions in the 4NC paradigm. RTs faster than 300 msec were removed. Trials with no responses were not included in the RT calculations but were categorized as incorrect in the accuracy calculations.

We conducted separate linear mixed-effects models (Baayen, Davidson, & Bates, 2008) predicting mean accuracy and mean RTs. We were interested in differences between format (hands and numerals) and judgment tasks (phonological and numerical), as well as their interaction, while including a random effect term for participants.

### fMRI Acquisition

Participants were scanned at the Neuroscience Imaging Center using a Siemens Allegra 3-T head-only magnet and standard radio frequency coil equipped with a mirror device to allow for fMRI stimuli presentation. The scanning session first consisted of a T1-weighted anatomical scan (repetition time [TR] = 1540 msec, echo time = 3.04 msec, voxel size = 1.00 × 1.00 × 1.00 mm), followed by T2-weighted functional scans that collected BOLD signals using a one-shot EPI pulse. Slices were collected in interleaved, ascending order (from foot to head), with no skips between slices (TR = 2000 msec, echo time = 25 msec, flip angle = 70°, isotropic voxel size = 3.125 × 3.125 × 3.125 mm, 36 slices, in-plane resolution = 64 × 64, field of view = 200 mm × 200 mm × 112.5 mm). The

functional scans for the 4NC paradigm were collected in five functional runs of 84 volumes each. Total scanning time was approximately 55 min.

### fMRI Preprocessing

Preprocessing was performed using the Analysis of Functional NeuroImages software (Cox, 1996) and consisted of the following: motion correction registration to the mean functional volume, high-pass filtering, scaling voxel activation values to have a mean of 100 (maximum limit of 200), and detrending. Because of the block design, we did not apply slice-time correction. Structural and functional images were warped to standardized space (Talairach, 1988) using a nonlinear transformation. Data were not smoothed. The unsmoothed functional data were imported into MATLAB using the Princeton Multi-Voxel Pattern Analysis (MVPA) toolbox (Detre et al., 2006). Custom MATLAB scripts were used to implement a series of multivariate analyses.

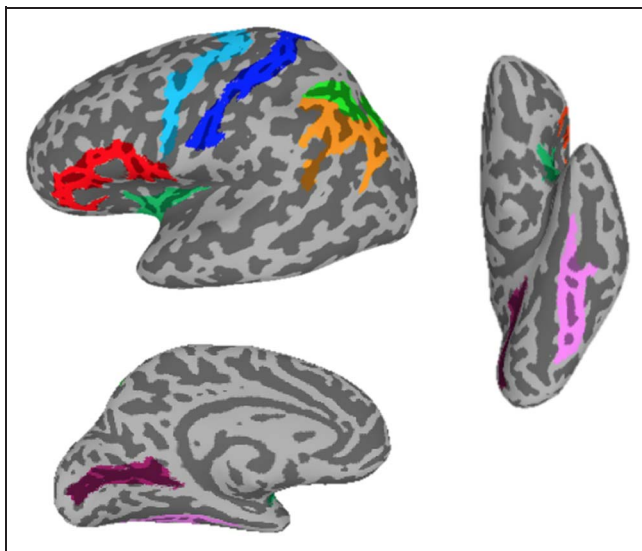
### ROIs

To investigate regions involved in numerical processing, we identified ROIs in the brain corresponding to each of the four facets previously described in the introduction: semantic (IPS and insula), phonological/verbal (left angular gyrus and left IFG), visual (fusiform gyrus and calcarine sulcus), and manual (precentral gyrus and postcentral gyrus). To individually isolate these ROIs, bilateral anatomical masks for each region were defined within each participant's native space using FreeSurfer's automated segmentation procedure (<https://surfer.nmr.mgh.harvard.edu>; Fischl et al., 2002, 2004). Participants' anatomical masks were then standardized to Talairach space to be used in all subsequent analyses. To see a summary of the size of each ROI, see Table 2. For a visual depiction of the ROIs placed on a standard template brain, see Figure 2. In

**Table 2.** Size of Each ROI, as Measured in Functional Voxels, Represented by Means and Standard Deviations (in Parentheses), As Well As Minimum and Maximum Number of Voxels across the Sample of Participants

	<i>M (SD)</i>	<i>Range (Minimum, Maximum)</i>
IPS	236.83 (33.01)	176–302
Left IPS	120.89 (24.79)	79–165
Right IPS	115.94 (21.44)	76–169
Insula	251.56 (15.4)	218–277
Left angular gyrus	162.00 (39.30)	99–254
Left IFG	258.89 (38.82)	183–308
Fusiform gyrus	314.83 (53.55)	207–459
Calcarine sulcus	214.61 (28.24)	170–263
Precentral gyrus	321.61 (36.97)	238–387
Postcentral gyrus	197.06 (37.94)	153–266





**Figure 2.** ROIs displayed on the left hemisphere of a standard (TT\_N27 template) brain. Semantic facet regions shown in green: IPS = light green, insula = dark green; phonological/verbal in orange: left angular gyrus = light orange, left IFG = dark orange; visual in purple: fusiform gyrus = light purple, calcarine sulcus = dark purple; manual in blue: precentral gyrus = light blue, postcentral gyrus = dark blue.

addition, we used the *Decoding* tool through Neurosynth.org (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) to confirm that the selected ROIs encompass regions in the brain associated with their respective facets. Each of the four facet maps showed particular similarity with neural maps of relevant terms within the Neurosynth database (i.e., relevant terms listed below found within top 2.5% most similar of all term entries): semantic: arithmetic ( $r = .211$ ), symbolic ( $r = .198$ ), calculation ( $r = .171$ ); verbal: language ( $r = .115$ ), word ( $r = .114$ ), phonological ( $r = .112$ ); visual: objects ( $r = .195$ ), visual stream ( $r = .158$ ), visual ( $r = .152$ ); manual: motor cortex ( $r = .292$ ), hand ( $r = .229$ ), finger ( $r = .217$ ).

## fMRI Multivariate Analyses

### *Task and Format Representations within the 4NC Paradigm*

The design of our 4NC paradigm allowed us to ask how representations of numerical and phonological processing might be different (or the same) across presentations of different formats. Because of the binary nature of the 4NC paradigm, we trained and tested using a support vector machine classifier model, one of the most often used classifiers in MVPAs (Floren, Naylor, & Miikkulainen, 2015; Pereira, Mitchell, & Botvinick, 2009), with a linear kernel, as is the default in MATLAB's implementation with two-class learning, within a leave-one-run-out cross-validation approach. The training and testing data consisted of the unsmoothed neural data ( $z$ -scored BOLD response within each run, after shifting time points by three TRs to account for the hemodynamic delay) from each voxel within each





of the ROIs. To control for confounds in this classification method, we followed the cross-validated confound regression approach outlined by Snoek, Miletic, and Scholte (2019) that accounts for differences in RTs between conditions. This approach involves first calculating the variance related to possible confounds (in this case, RTs) within only the training data set and then removing this variance from both the training and testing data sets before classification. For trials with no RT (because of no response made during the scan), we used the participant's average RT for that given condition. RTs were  $z$  scored within each run, before being included in the analyses. Accuracy for the classifier for each ROI and participant was calculated as the average accuracy across the entire leave-one-run-out cross-validation approach.

Because of our 2 (task)  $\times$  2 (format) design, we were able to investigate both task representations and then format representations. To investigate representations of numerical and phonological processing, the classifier decoded the task information contained within the trials. First, we trained and tested this classifier on trials with the same format presented (e.g., only the hands or only the numerals). We refer to these analyses, which use the same type of training and testing data, as *consistent*. In addition to these consistent analyses, we were also able to perform *generalizability* analyses. These differed from the consistent analyses in that we trained and tested on trials with a different format presented (e.g., train on only the hands, test on only the numerals, and vice versa). Next, we were able to ask how format representations (numerals and hands) might be different (or the same) regardless of their involvement in a certain cognitive task through both consistent classification (e.g., training and testing only on one task) and generalization classification approaches (e.g., train on only the numerical task, test on only the phonological task, and vice versa). See Figure 3 for an overview of our 2  $\times$  2 design and how each of the aforementioned analyses is implemented.

### *Group-level Statistical Tests*

To conduct group-level statistical analyses, we first calculated the accuracy for the classifier, within each ROI for each participant, across the entire leave-one-run-out cross-validation approach for each of the analyses separately (e.g., training/testing on numerical task; training/testing on phonological task). We then averaged these two classifier accuracies to give a final classifier accuracy. Group-level statistical analyses then determined whether classification was greater in a given ROI than would be expected by chance (0.50). These statistical findings were then further supported through nonparametric permutation testing, which has the added benefit of better controlling Type-1 errors (Valente, Castellanos, Hausfeld, De Martino, & Formisano, 2021). Instead of assuming a normal distribution, we generated a null distribution of 1000 classification accuracy values by randomly scrambling

**Figure 3.** Analytical approach reflecting our  $2 \times 2$  design. For *Consistent* analyses, the training and testing data consisted of trials using the same stimulus format (e.g., hands) when distinguishing between the two different tasks, and when distinguishing between the two different stimulus formats, the training and testing data consisted of trials using the same task (e.g., numerical judgment). For *Generalization* analyses, the testing data consisted of trials using the type of trial not used during training (e.g., when distinguishing between tasks, classifiers were trained using trials consisting of hands and then tested on trials using numerals; when distinguishing between stimulus formats, classifiers were trained using trials consisting of numerical judgments and then tested on trials using phonological judgments).

	Consistent		Generalization	
	Train	Test	Train	Test
Classifying Task (Numerical judgment vs. phonological judgment)	3	6	3	
				6
Classifying Format (Numerals vs. Hands)	Numerical judgment	Numerical judgment	Numerical judgment	Phonological judgment
	Phonological judgment	Phonological judgment	Phonological judgment	Numerical judgment

the true block labels within each run before training and testing the classifier. The true classification accuracy is then compared against the null distribution, and a one-tailed  $p$  value is calculated using the number of observed instances in the null distribution that are as, or more, extreme than the true classification accuracy. This approach allows us to assess how well the classifier performs when the underlying data are no longer in response to the tasks performed and stimuli observed by the participant, and instead are just assigned at random. When reporting results, we report  $p$  values derived from traditional null-hypothesis statistical testing (NHST) and from nonparametric permutation testing. We base our conclusions and interpretations on NHST to remain consistent with related literature, but also report nonparametric permutation results, showing that statistical effects do not depend on the requisite assumptions of NHST.

#### Organization of Representations within the 4NC Paradigm

To better understand how representations of the four facets of number are organized within regions, we investigated whether any regions that represent both format and tasks exhibit systematic commonalities across participants. To do this, we tested whether the voxels that are important for one aspect (format) overlap with those important in the other (task). We first identified the most important voxels for each contrast based on the weights resulting from each trained Support Vector Machine model within each participant, by calculating the absolute value for each weight, where larger values indicate a voxel's increased contribution toward performing the classification. We sought to identify approximately 50 voxels that were important for classifying task, and 50 voxels that were important for classifying format, as these would reflect a reasonable number of features from which to draw

conclusions (Davis et al., 2014; Hindy, Altmann, Kalenik, & Thompson-Schill, 2012), without risking overfitting by including all voxels in a given ROI. We identified the voxels important for classifying task using trials of both formats, and identified the voxels important for classifying format using trials of both tasks, while training on the whole data set in each case. We used one voxel identified as being important for classifying task, and calculated the Euclidean distance between it and another voxel identified as being important for classifying task. We repeated this process until all pairwise Euclidean distances had been calculated for all voxels important for classifying task. The same steps were then conducted using the voxels identified as being important for classifying format. Lastly, we returned to the voxel identified as being important for classifying task and calculated the Euclidean distance between it and one of the voxels identified as being important for classifying format. We repeated this process until all pairwise Euclidean distances had been calculated between task-important and format-important voxels. These three sets of calculations resulted in an average Euclidean distance for each participant for each of the following: within task (i.e., average distance between task voxels); within format (i.e., average distance between format voxels); between task and format (i.e., average distance between task voxels and format voxels). The average Euclidean distances were then analyzed in linear mixed-effects models, utilizing contrast coding to compare the within versus between, as well as the within-task versus within-format average Euclidean distances, while including a random effect term for participants.

## RESULTS

### Behavioral Performance

Before analyzing the 4NC accuracy data, we removed all phonological trials involving the number “two” as the correct response for these trials was deemed too ambiguous

( $M = 0.31$  vs.  $M = 0.83$  for all other numbers). In the 4NC paradigm, performance differences were observed for both accuracy and RTs. Participants were more accurate during the numerical judgment task ( $M = 0.96$ ,  $SD = 0.04$ ) than during the phonological judgment task ( $M = 0.83$ ,  $SD = 0.11$ ;  $\beta = -0.764$ ,  $p < .001$ ). No differences were observed between trials involving the numerals ( $M = 0.91$ ,  $SD = 0.09$ ) compared with the hand format ( $M = 0.88$ ,  $SD = 0.12$ ;  $\beta = 0.12$ ,  $p = .15$ ), and there was no statistically significant interaction between the task and format conditions ( $\beta = 0.04$ ,  $p = .61$ ). Participants were slower to respond during the phonological judgment task ( $M = 1411$  msec,  $SD = 249$  msec) compared with the numerical judgment task ( $M = 1006$  msec,  $SD = 199$  msec;  $\beta = 0.67$ ,  $p < .001$ ). Participants were also slower to respond to the hands ( $M = 1323$  msec,  $SD = 312$  msec) than to the numerals ( $M = 1094$  msec,  $SD = 248$  msec;  $\beta = -0.38$ ,  $p < .001$ ), which was more pronounced in the phonological judgment task than the numerical judgment task as reflected in a statistically significant interaction between the task and format conditions (phonological task and hands:  $M = 1563$  msec,  $SD = 215$  msec; phonological task and numerals:  $M = 1259$  msec,  $SD = 181$ ; numerical task and hands:  $M = 1083$  msec,  $SD = 179$  msec; numerical task and numerals:  $M = 929$  msec,  $SD = 191$  msec;  $\beta = -0.12$ ,  $p = .005$ ).

## fMRI Multivariate Results

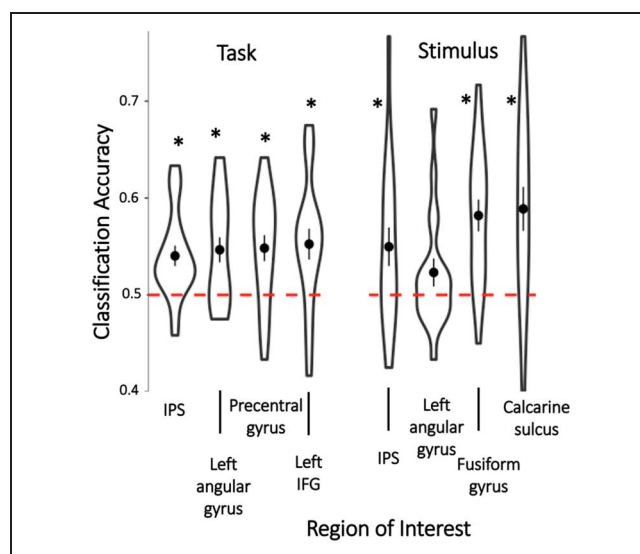
### Task Representations within the 4NC Paradigm

We investigated task representations in the 4NC paradigm using a classification approach across two analyses: consistent and generalization. For our consistent analyses, in which training and testing consisted of the same format type, classifiers trained on neural patterns from four ROIs were able to successfully decode the two different tasks (numerical and phonological) at a level above chance (0.50 because of two conditions): IPS ( $M = 0.54$ ,  $SD = 0.06$ ;  $t(17) = 2.94$ ,  $p = .009$ ;  $p$  (one-tailed permutation testing (perm.))  $< .001$ ), left angular gyrus ( $M = 0.55$ ,  $SD = 0.07$ ;  $t(17) = 3.73$ ,  $p = .002$ ;  $p$  (perm.)  $< .001$ ), left IFG ( $M = 0.53$ ,  $SD = 0.06$ ;  $t(17) = 2.22$ ,  $p = .040$ ;  $p$  (perm.) = .008), and precentral gyrus ( $M = 0.53$ ,  $SD = 0.05$ ;  $t(17) = 2.24$ ,  $p = .038$ ;  $p$  (perm.) = .029). No other ROIs reached statistical significance: insula ( $M = 0.50$ ,  $SD = 0.05$ ;  $t(17) = 0.18$ ,  $p = .863$ ;  $p$  (perm.) = .409), fusiform gyrus ( $M = 0.52$ ,  $SD = 0.06$ ;  $t(17) = 1.68$ ,  $p = .112$ ;  $p$  (perm.) = .066), calcarine sulcus ( $M = 0.49$ ,  $SD = 0.06$ ;  $t(17) = -0.85$ ,  $p = .409$ ;  $p$  (perm.) = .752), and postcentral gyrus ( $M = 0.51$ ,  $SD = 0.04$ ;  $t(17) = 0.65$ ,  $p = .524$ ;  $p$  (perm.) = .318).<sup>1</sup>

To further understand these task representations, we investigated whether the above-chance classification performances were driven by training and testing on one of the stimulus formats, or both. When training and testing only with the hands stimuli, the left IFG did not reach statistical significance ( $M = 0.54$ ,  $SD = 0.10$ ;  $t(17) = 1.48$ ,  $p = .158$ ,  $p$  (perm.) = .042), the IPS was at a trending level of statistical significance ( $M = 0.54$ ,  $SD = 0.10$ ;  $t(17) = 1.79$ ,

$p = .091$ ,  $p$  (perm.) = .019), and both left angular gyrus ( $M = 0.54$ ,  $SD = 0.07$ ;  $t(17) = 2.67$ ,  $p = .016$ ,  $p$  (perm.) = .016) and precentral gyrus ( $M = 0.55$ ,  $SD = 0.08$ ;  $t(17) = 2.76$ ,  $p = .013$ ,  $p$  (perm.) = .01) exceeded the conventional threshold of  $p < .05$  for statistical significance. When training and testing only with the numerals, left IFG could not classify above chance ( $M = 0.53$ ,  $SD = 0.10$ ;  $t(17) = 1.42$ ,  $p = .175$ ,  $p$  (perm.) = .044), nor could precentral gyrus ( $M = 0.51$ ,  $SD = 0.07$ ;  $t(17) = 0.29$ ,  $p = .776$ ,  $p$  (perm.) = .413), but IPS ( $M = 0.54$ ,  $SD = 0.06$ ;  $t(17) = 3.33$ ,  $p = .004$ ,  $p$  (perm.) = .014) and left angular gyrus could ( $M = 0.56$ ,  $SD = 0.08$ ;  $t(17) = 3.03$ ,  $p = .008$ ,  $p$  (perm.) = .002).

For the generalization test, we tested whether the neural representations would be robust enough to generalize to a new format during testing (which had not been presented to the classifier during training; i.e., training with numerals, testing with hands; training on hands, testing with numerals). We only conducted this follow-up analysis within the four ROIs that showed statistically significant classification accuracies in the above (orthogonal) consistent analysis. Classifiers for each of the ROIs showed an ability to generalize to an unseen format during testing at a level above chance: IPS ( $M = 0.54$ ,  $SD = 0.05$ ;  $t(17) = 3.79$ ,  $p = .001$ ;  $p$  (perm.) = .001), left angular gyrus ( $M = 0.54$ ,  $SD = 0.06$ ;  $t(17) = 2.93$ ,  $p = .009$ ;  $p$  (perm.)  $< .001$ ), left IFG ( $M = 0.55$ ,  $SD = 0.07$ ;  $t(17) = 3.30$ ,  $p = .004$ ;  $p$  (perm.)  $< .001$ ), and precentral gyrus ( $M = 0.55$ ,  $SD = 0.06$ ;  $t(17) = 3.56$ ,  $p = .002$ ;  $p$  (perm.)  $< .001$ ).<sup>2</sup>



**Figure 4.** Classification accuracy of generalization analyses. Left side depicts generalizability of task representations across format. Right side depicts generalizability of format representations across task. Regions depicted here were able to successfully classify within the consistent analyses at a level above chance. Red dotted line represents chance performance accuracy (.50). \*Indicates statistical significance based on traditional inferential statistics ( $p < .05$ ). Error bars reflect the standard error of the mean. Distribution reflected in violin plot, where length of each shape corresponds to the range from all participants, and width represents frequency of corresponding classification accuracy.



Classification accuracies for generalizability of task representations across format are shown in Figure 4.

To probe these generalization results further, we investigated directionality by asking whether the information contained within the ROIs would generalize in a symmetrical fashion. For example, could classifiers train and test equally well on the different stimulus formats? For the ROIs showing ability to successfully generalize representations relating to the task, both the left IFG and the precentral gyrus were consistent with symmetrical generalization: Hands  $\rightarrow$  Numerals (left IFG:  $M = 0.57$ ,  $SD = 0.10$ ;  $t(17) = 2.72$ ,  $p = .015$ ,  $p$  (perm.)  $< .001$ ; precentral gyrus:  $M = 0.54$ ,  $SD = 0.06$ ;  $t(17) = 2.94$ ,  $p = .009$ ,  $p$  (perm.)  $= .007$ ) and Numerals  $\rightarrow$  Hands (left IFG:  $M = 0.54$ ,  $SD = 0.07$ ;  $t(17) = 2.37$ ,  $p = .030$ ,  $p$  (perm.)  $= .012$ ; precentral gyrus:  $M = 0.56$ ,  $SD = 0.07$ ;  $t(17) = 3.25$ ,  $p = .005$ ,  $p$  (perm.)  $< .001$ ). The IPS and left angular gyrus did not show the same support (in that our statistical significance threshold was not met for *both* directions); however, the directions of the effects were the same: training on hands and testing on numerals (IPS:  $M = 0.55$ ,  $SD = 0.06$ ;  $t(17) = 3.94$ ,  $p = .001$ ,  $p$  (perm.)  $< .001$ ; left angular gyrus:  $M = 0.56$ ,  $SD = 0.06$ ;  $t(17) = 4.17$ ,  $p = .001$ ,  $p$  (perm.)  $< .001$ ), training on numerals and testing on hands (IPS:  $M = 0.53$ ,  $SD = 0.08$ ;  $t(17) = 1.59$ ,  $p = .129$ ,  $p$  (perm.)  $= .03$ ; left angular gyrus:  $M = 0.52$ ,  $SD = 0.10$ ;  $t(17) = 0.90$ ,  $p = .379$ ,  $p$  (perm.)  $= .095$ ).

#### Format Representations within the 4NC Paradigm

In addition to investigating task representations, our 4NC paradigm allowed us to investigate format representations using a classification approach across consistent and generalization analyses. For our consistent analyses, in which training and testing consisted of the same task type, classifiers trained on neural patterns from four ROIs were able to successfully decode the two different formats (numerals and hands) at a level above chance (0.50 because of two conditions): IPS ( $M = 0.57$ ,  $SD = 0.09$ ;  $t(17) = 3.35$ ,  $p = .004$ ;  $p$  (perm.)  $< .001$ ), left angular gyrus ( $M = 0.53$ ,  $SD = 0.07$ ;  $t(17) = 2.17$ ,  $p = .044$ ;  $p$  (perm.)  $= .009$ ), fusiform gyrus ( $M = .62$ ,  $SD = 0.08$ ;  $t(17) = 6.06$ ,  $p < .001$ ;  $p$  (perm.)  $< .001$ ), and calcarine sulcus ( $M = 0.59$ ,  $SD = 0.06$ ;  $t(17) = 5.91$ ,  $p < .001$ ;  $p$  (perm.)  $< .001$ ). No other ROIs reached statistical significance: insula ( $M = 0.49$ ,  $SD = 0.06$ ;  $t(17) = -0.73$ ,  $p = .474$ ;  $p$  (perm.)  $= .707$ ), left IFG ( $M = 0.49$ ,  $SD = 0.08$ ;  $t(17) = -0.60$ ,  $p = .555$ ;  $p$  (perm.)  $= .784$ ), precentral gyrus ( $M = 0.49$ ,  $SD = 0.07$ ;  $t(17) = -0.35$ ,  $p = .729$ ;  $p$  (perm.)  $= .664$ ), and post-central gyrus ( $M = 0.50$ ,  $SD = 0.06$ ;  $t(17) = 0.15$ ,  $p = .881$ ;  $p$  (perm.)  $= .447$ ).<sup>3</sup>

To further understand these format representations, we investigated whether the above-chance classification performances were driven by training and testing on one of the task types, or both. When training and testing only when making numerical judgments, the IPS ( $M = 0.61$ ,  $SD = 0.10$ ;  $t(17) = 4.74$ ,  $p < .001$ ,  $p$  (perm.)  $< .001$ ), fusiform gyrus ( $M = 0.64$ ,  $SD = 0.11$ ;  $t(17) = 5.19$ ,  $p <$

$.001$ ,  $p$  (perm.)  $< .001$ ), and calcarine sulcus ( $M = 0.61$ ,  $SD = 0.11$ ;  $t(17) = 4.40$ ,  $p < .001$ ,  $p$  (perm.)  $< .001$ ) could classify at a level above chance, but the left angular gyrus could not ( $M = 0.51$ ,  $SD = 0.09$ ;  $t(17) = 0.56$ ,  $p = .584$ ,  $p$  (perm.)  $= .266$ ). When training and testing only when making phonological judgments, the fusiform gyrus ( $M = 0.61$ ,  $SD = 0.10$ ;  $t(17) = 4.58$ ,  $p < .001$ ,  $p$  (perm.)  $< .001$ ), calcarine sulcus ( $M = 0.57$ ,  $SD = 0.07$ ;  $t(17) = 4.10$ ,  $p = .001$ ,  $p$  (perm.)  $= .002$ ), and left angular gyrus ( $M = 0.56$ ,  $SD = 0.09$ ;  $t(17) = 2.61$ ,  $p = .018$ ,  $p$  (perm.)  $= .001$ ) could classify at a level above chance, but IPS could not ( $M = 0.53$ ,  $SD = 0.10$ ;  $t(17) = 1.23$ ,  $p = .236$ ,  $p$  (perm.)  $= .049$ ).

We then tested whether the neural representations would be robust enough to generalize to a new task during testing (which had not been presented to the classifier during training; i.e., training on semantic task, testing on phonological task; training on phonological task, testing on semantic task). We applied this generalization approach within the four ROIs that showed statistically significant classification accuracies in the above orthogonal, consistent analysis. Classifiers for three of the four ROIs showed an ability to generalize to an unseen task type during testing at a level above chance: IPS ( $M = 0.55$ ,  $SD = 0.08$ ;  $t(17) = 2.49$ ,  $p = .023$ ;  $p$  (perm.)  $= .001$ ), fusiform gyrus ( $M = 0.59$ ,  $SD = 0.10$ ;  $t(17) = 3.90$ ,  $p = .001$ ;  $p$  (perm.)  $< .001$ ), and calcarine sulcus ( $M = 0.58$ ,  $SD = 0.07$ ;  $t(17) = 5.01$ ,  $p < .001$ ;  $p$  (perm.)  $< .001$ ). The left angular gyrus ROI did not show evidence of generalizability ( $M = 0.52$ ,  $SD = 0.05$ ;  $t(17) = 1.43$ ,  $p = .172$ ;  $p$  (perm.)  $= .087$ ).<sup>4</sup> Classification accuracies for generalizability of format representations across task types are shown in Figure 4.

We next asked whether our generalization accuracies for classifying format were driven by any commonalities within participants, as would be reflected in correlations between the three ROIs that generalized above chance. Generalization accuracies for the calcarine sulcus were correlated with both the IPS,  $r(16) = .49$ ,  $p = .038$ , and the fusiform gyrus,  $r(16) = .56$ ,  $p = .016$ ; however, the IPS was not correlated with the fusiform gyrus,  $r(16) = .18$ ,  $p = .467$ .

To probe these generalization results further, we again investigated directionality by asking whether the information contained within the ROIs would generalize in a symmetrical fashion, this time for representations relating to stimulus format. All three of the ROIs showing ability to successfully generalize were consistent with symmetrical generalization (although one of the IPS results was only marginally significant from the NHST). When training on numerical judgment task trials and testing on phonological task trials: IPS ( $M = 0.55$ ,  $SD = 0.09$ ;  $t(17) = 2.47$ ,  $p = .025$ ,  $p$  (perm.)  $< .001$ ), fusiform gyrus ( $M = 0.58$ ,  $SD = 0.11$ ;  $t(17) = 3.01$ ,  $p = .008$ ,  $p$  (perm.)  $< .001$ ), and calcarine sulcus ( $M = 0.57$ ,  $SD = 0.08$ ;  $t(17) = 3.84$ ,  $p = .001$ ,  $p$  (perm.)  $< .001$ ). When training on phonological task trials and testing on numerical judgment task trials: IPS ( $M = 0.55$ ,  $SD = 0.10$ ;  $t(17) = 2.01$ ,  $p = .060$ ,  $p$  (perm.)  $= .003$ ), fusiform gyrus ( $M = 0.60$ ,  $SD = 0.10$ ;  $t(17) = 4.15$ ,



$p = .001, p$  (perm.)  $< .001$ , and calcarine sulcus ( $M = 0.59, SD = 0.09; t(17) = 4.50, p < .001, p$  (perm.) = .001).

### Organization of Representations within the 4NC Paradigm

Because the IPS showed evidence of being involved in representing information about both tasks and format, we investigated how and where these representations might be organized within this neural region. As this bilateral IPS region contained voxels in both the left and right hemispheres, we investigated organization within each hemisphere separately, as that allowed us to reduce potential confounds of voxel comparisons from separate hemispheres. To achieve our goal of including approximately 50 voxels for this analysis, we identified the top 40% most important voxels within each IPS hemisphere pertaining to task and format representations to test whether the respective representations relied on overlapping or distinct voxels.

Within the left IPS, the average Euclidean distance within the set of voxels that were important for either classifying the task or the format ( $M = 16.90$  mm,  $SD = 1.66$ ) did not statistically significantly differ from voxels across task/format ( $M = 16.78$  mm,  $SD = 1.47; \beta = 0.04, p = .556$ ). Furthermore, the voxels representing task information ( $M = 17.35$  mm,  $SD = 1.40$ ) were even more distributed and exhibited a greater distance from each other than those voxels representing format information ( $M = 16.44$  mm,  $SD = 1.81; \beta = -0.24, p < .001$ ).

The right IPS showed the same patterns as above, as the average Euclidean distance within the set of voxels that were important for either classifying the task or the format ( $M = 16.24$  mm,  $SD = 1.56$ ) again did not statistically significantly differ from voxels across task/format ( $M = 16.08$  mm,  $SD = 1.10; \beta = 0.05, p = .431$ ). The voxels representing task information ( $M = 17.10$  mm,  $SD = 0.79$ ) were more distributed compared with those representing format information ( $M = 15.37$  mm,  $SD = 1.67; \beta = -0.51, p < .001$ ). See Figure 5 for a visual depiction

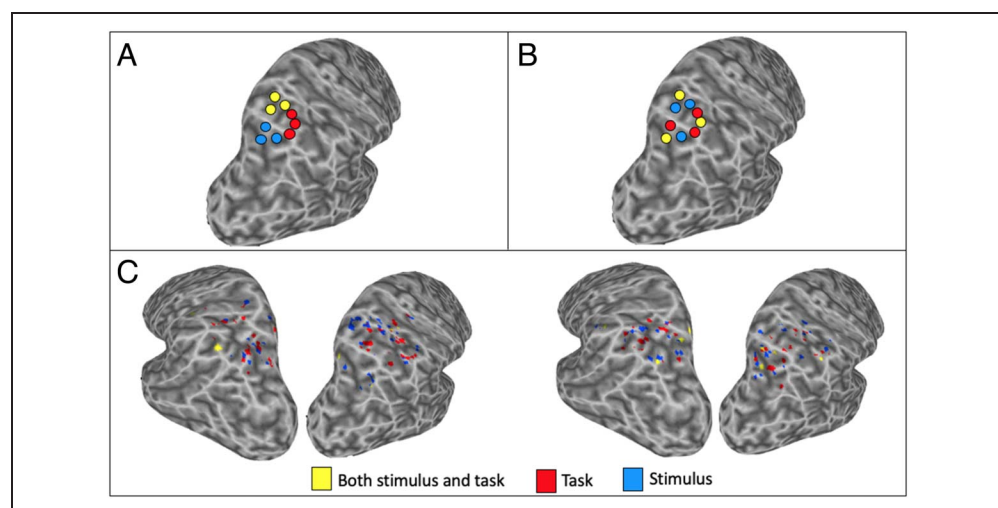
of the hypothetical possible organization, as well as data from representative participants.

### DISCUSSION

In this study, we investigated how the brain represents both number-related task and format information, by contrasting the semantic with the phonological/verbal number codes, and the visual with the proposed manual number codes, respectively. While semantic and phonological/verbal regions were able to decode the tasks, and visual regions were able to decode the format, IPS was able to do both. In addition, representations within IPS were robust enough to generalize in both instances. Thus, we provide evidence that IPS contains representations that are informative about both numerical task and numerical format. Critically, our findings extend previous work examining format independence by additionally examining representations derived from semantic and phonological tasks, in contrast to such previous work that primarily manipulated stimulus format (e.g., dots vs Arabic numerals; Bulthé et al., 2014, 2015; Piazza et al., 2007; Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003).

Why might such representations be concurrently contained within IPS? Although often considered an important hub for the semantic component of numerical processing (Nieder & Dehaene, 2009), our evidence supports claims that IPS is also involved in other facets, even without explicit direction to use the semantic component (e.g., visual: Eger et al., 2003). In addition, these findings might suggest an anatomical location (IPS) in which the integration of the distinct number codes can occur within the triple-code model. The model posits that the three codes are separate, yet interconnected (Myers & Szűcs, 2015); therefore, it is possible that both types of representations we observed in IPS (format and task) reflect a type of integration of the facets. To better understand the underlying organization of the representations within IPS, we examined Euclidean distances between voxels, resulting in evidence suggesting that these representations are

**Figure 5.** Organization of representations within the 4NC paradigm. (A) Hypothetical illustration of IPS voxels important for task, format, or both format and task, depicting the possibility of systematic localization. (B) Hypothetical illustration of IPS voxels important for task, format, or both format and task, depicting the possibility of distributed organization. (C) Data from two representative participants, showing a distributed organization of IPS voxels.



distributed throughout the IPS, separately in both the left and right hemispheres (and do not find support that these representations are different between hemispheres), rather than being contained within distinct, localized subregions within the IPS. These findings do not align with evidence suggesting numerical processing differs between anatomically distinct regions within IPS (Bugden, Price, McLean, & Ansari, 2012; Cappelletti, Barth, Fregni, Spelke, & Pascual-Leone, 2007). Future studies may seek to better understand the organization of these representations with a more fine-grained approach to the IPS.

In addition, task representations were also identified within the left angular gyrus. These representations were robust enough to generalize across multiple formats, thus suggesting that this region represents the distinction between semantic and phonological/verbal processing of numerical information in a manner that is invariant to format (see Regev, Honey, Simony, and Hasson [2013] for similar findings of format-invariance in nonnumber spoken and written words).

By contrasting the visual and potential manual codes, we detected format representations within left angular gyrus, fusiform gyrus, and calcarine sulcus. Representations within these latter two visual regions were robust enough to generalize across multiple tasks, providing evidence that the format representations are task invariant, in contrast to other visual regions (i.e., inferior temporal gyrus), which have been proposed as having a preference for math tasks that outweighs a preference for Arabic numerals (Grotheer, Jeska, & Grill-Spector, 2018). Representations within left angular gyrus were not robust enough to generalize across the task in which they were accessed, thus suggesting the format representations may be particularly tied to a certain task, possibly when invoking phonological/verbal processing (Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002).

Our findings of representations that can generalize across formats and tasks contributes to recent evidence showing similar findings within parietal regions, and to a lesser degree within left IFG (Wilkey, Conrad, Yeo, & Price, 2020) across different formats and tasks, although we only found generalization across tasks within left IFG, not formats. Wilkey and colleagues show that neural patterns of activity representing number magnitudes can both generalize across dots and Arabic numerals, and separately across identification and number comparison tasks, yet our findings extend further into additional facets of numerical processing including manual number format and phonological/verbal processing of numerical information.

As part of this work, we introduced a novel paradigm to better understand multiple facets of numerical processing. This paradigm was specifically designed to target all four proposed facets, and the results were analyzed within neural regions identified in advance for their relevance to numerical processing. As opposed to traditional univariate analyses, we utilized a multivariate approach that has the

advantage of being sensitive to percepts and cognitive states at higher levels of specificity (Coutanche, 2013). Analyzing our 4NC paradigm through this approach makes it more likely that informative representations will be gleaned by directly contrasting the facets of numerical processing.

Of note for future studies, because of the efficiency and flexibility provided by our 4NC paradigm, one could potentially utilize it in a more exploratory capacity to possibly localize additional subcomponents of numerical processing. Whereas the current study applied the 4NC paradigm to ROIs selected for their relevance to numerical processing, future research may use the paradigm in a way to better understand additional neural components and whether they, too, contain information that distinguishes tasks and formats. Our paradigm can be easily adapted for use with children and other populations requiring special considerations when conducting fMRI research. Prior work investigating the verbal and semantic facets of numerical processing in children with and without mathematical learning disability utilized functional localizers during their fMRI scans (Berteletti, Prado, & Booth, 2014; see also Prado et al., 2011). During two different localizer scans employed by Berteletti and colleagues, participants completed a rhyming judgment task (nonnumerical words) and a numerosity judgment task (dot quantities). Although their localizer tasks proved useful, we suggest our 4NC paradigm could be a more efficient method for obtaining even more informative results, with added benefits in that it is a shorter task and requires responses to numerical stimuli in all conditions, while also addressing all four facets of number processing.

Because this study was particularly focused on information within neural regions associated with general representations of number, we were unable to more systematically investigate magnitude representations of number in the brain. Future research may wish to better control the frequency with which each numerical quantity is presented during the 4NC paradigm to allow for a deeper investigation into how representations of magnitude may generalize across tasks and formats. In addition, our study does not tell us whether representations may be found in other brain regions, and how regions that represent different facets of number are connected with each other. Future studies may wish to investigate connectivity between regions identified here, as a way to learn about crosstalk and possible similar representations for the same facet. Connectivity approaches that examine shared fluctuations in multivoxel representations might be particularly applicable and useful (Anzellotti & Coutanche, 2018; Coutanche & Thompson-Schill, 2013, 2014).

In addition, in this study, we were unable to directly identify differences in representations between numerals and hands within the expected locations for a manual code: pre- and postcentral gyrus. One possible reason is that our sample of participants consisted of all adults (older than 18 years old). Perhaps children, who more

readily rely on hands as a form of number representation (e.g., when counting), would show the expected differences between numerals and hands (Geary, Hoard, Byrd-Craven, & DeSoto, 2004). It could also be because of our hand stimuli, because it is unclear whether our images of hands are processed similarly to nonsymbolic or symbolic numerical stimuli. By using only one specific, canonical combination of fingers to represent each numerical quantity, participants may have treated our hand stimuli like symbolic stimuli (similar to patterns on a dice). Interestingly, representations within the precen-tral gyrus did show evidence of differences between numerical and phonological tasks, even across format. One possible reason for this may be the prompt used in our phonological task. Because we asked participants about a feature of a vowel within the number name, it is possible that this focus on particular phonemes could have driven the representational difference between numerical and phonological processing (Fiez et al., 1995). The precen-tral gyrus has previously been shown to be involved in speech perception at the phoneme level, possibly as part of a link between phonological perception and production (Pulvermüller et al., 2006).

In this article, we have identified regions in which numerical task or format representations are present, and in the case of IPS, a location containing both. These results speak to the importance of investigating how the brain stores information relating to the semantic, visual, verbal, and manual facets of numerical processing, while also providing evidence of a viable paradigm for future research.

## APPENDIX

During the functional scans, participants also completed four functional runs of an additional paradigm, based on a paradigm that has been used previously to isolate number-specific brain regions without distinguishing between different facets in adults and children (Emerson & Cantlon, 2015; Cantlon & Li, 2013).

## Methods

In this paradigm, participants viewed pairs of stimuli consisting of numbers, words, faces, and shapes (either resembling geometric or tool shapes, counterbalanced across participants). Participants judged whether the pairs of stimuli were matches or nonmatches (left vs. right

buttons were counterbalanced across participants). Approximately 50% of the trials were matches. Unlike in the original paradigm, we changed the color of the back-ground behind the stimuli (gray instead of green) and required participants to press a button for both matching and nonmatching stimulus pairs (instead of only for matching pairs). For additional details concerning the stimuli used, please see Emerson and Cantlon (2015) and Figure A1.

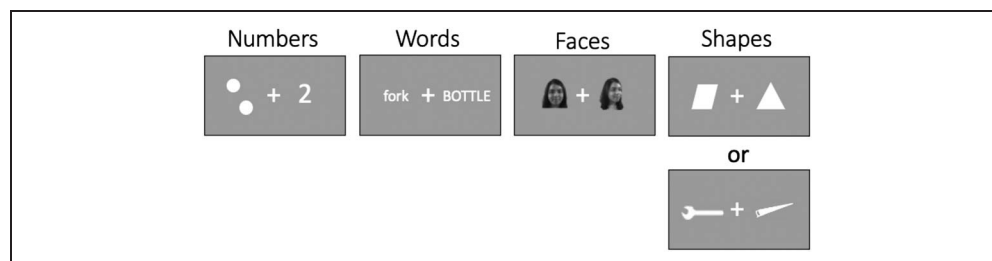
The paradigm consisted of presenting stimuli in a blocked design across four functional runs. In our version, all runs consisted of two blocks per condition (compared with three per run in Emerson & Cantlon, 2015), with three picture comparison trials (same condition) in each block. As in the original version of the paradigm, blocks were pseudorandomized such that all conditions were presented once before a second presentation of any condition. Picture comparison trials were presented for 2 sec, followed by a 2-sec intertrial interval (fixation cross). Each block was followed by 8 sec of fixation, collected in four functional runs of 80 volumes each.

We conducted separate linear mixed-effects models (Baayen et al., 2008) predicting mean accuracy and mean RTs for each of the four conditions in this paradigm, while including a random effect term for participants. We were specifically interested in differences between the numbers condition and each of the other three conditions (words, faces, and shapes).

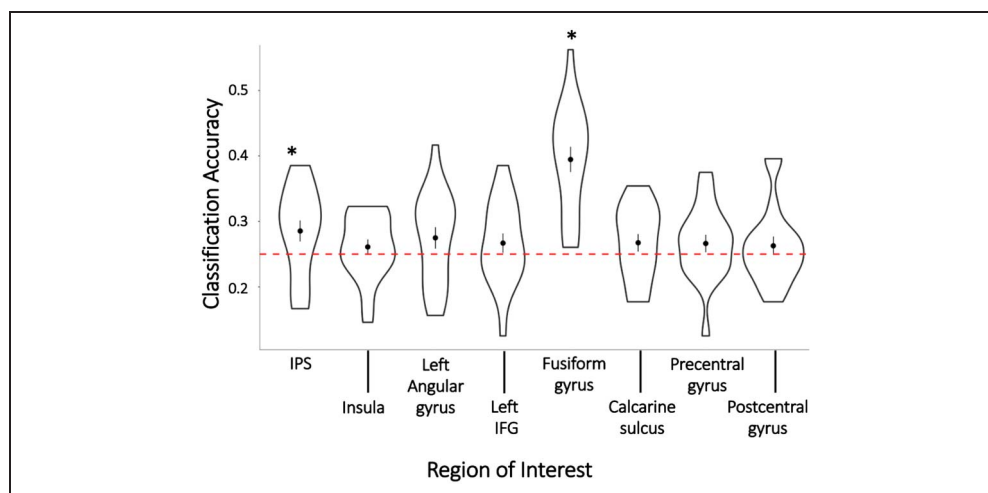
Two participants' data were removed from this para-digm because of poor behavioral performance (i.e., mean accuracies more than 2 *SDs* below the mean of the group). All analyses reported below were conducted on the remaining 18 participants for this paradigm (12 females, 6 males,  $M_{\text{age}} = 22.7$ ,  $SD = 4.1$ ).

To determine how different types of stimulus formats and comparison tasks evoke different neural responses, we investigated the representations within each of our bilateral ROIs in response to four conditions: numbers, words, faces, and shapes. In contrast to traditionally imple-mented univariate approaches for this paradigm, we implemented a multivariate approach in which we trained and tested a Gaussian Naïve Bayes classifier using a leave-one-run-out cross-validation approach to predict one of four possible categorizations (the conditions of the stim-uli). The training and testing data consisted of the unsmoothed neural data ( $z$ -scored BOLD response within each run, after shifting time points by three TRs to account

**Figure A1.** Four stimulus conditions presented, inspired by Emerson and Cantlon (2015).



**Figure A2.** Classification accuracy from the Cantlon localizer paradigm. Red dotted line represents chance performance accuracy (0.25). \*Indicates statistical significance ( $p < .05$ ). Error bars reflect the standard error of the mean. Distribution is reflected in violin plot, where length of each shape corresponds to the range from all participants, and width represents frequency of corresponding classification accuracy.



for the hemodynamic delay) from each voxel within each of the ROIs. To control for confounds in this classification method, we followed the cross-validated confound regression approach outlined by Snoek and colleagues (2019) that accounts for differences in RTs between the four conditions. This approach involves first calculating the variance related to possible confounds (in this case, RTs) within only the training data set, and then removing this variance from both the training and testing data sets before classification. For trials with no RT (because of no response made during the scan), we used the participant's average RT for that given condition. RTs were z-scored within each run, before being included in the above analyses. Accuracy for the classifier for each ROI and participant was calculated as the average accuracy across the entire leave-one-run-out cross-validation approach. This classifier accuracy was then used within group-level statistical analyses to determine whether classification was greater in a given ROI than would be expected by chance (0.25).

## Results

Participants did not differ in accuracy in the numbers condition ( $M = 0.97$ ,  $SD = 0.04$ ) compared with the faces ( $M = 0.95$ ,  $SD = 0.04$ ), words ( $M = 0.97$ ,  $SD = 0.04$ ), and shapes ( $M = 0.96$ ,  $SD = 0.03$ ) conditions (all  $ps > .061$ ); however, they were slower to respond to the numbers ( $M = 1167$  msec,  $SD = 169$  msec) when separately compared with each of the other three conditions: faces ( $M = 1004$  msec,  $SD = 159$  msec;  $\beta = -0.42$ ,  $p < .001$ ), words ( $M = 1060$  msec,  $SD = 118$  msec;  $\beta = -0.28$ ,  $p < .001$ ), and shapes ( $M = 916$  msec,  $SD = 135$  msec;  $\beta = -0.64$ ,  $p < .001$ ).

We utilized a machine learning classification approach using the patterns of voxels within ROIs, to predict one of the four trial types on held-out data. Classifiers trained on neural patterns of activity within two of the ROIs successfully classified the conditions at a level above chance (0.25): IPS ( $M = 0.29$ ,  $SD = 0.07$ ;  $t(17) = 2.19$ ,  $p = .042$ ;

$p$  [perm.]  $< .003$ ) and fusiform gyrus ( $M = 0.39$ ,  $SD = 0.08$ ;  $t(17) = 7.43$ ,  $p < .001$ ;  $p$  [perm.]  $< .001$ ). No other ROIs reached statistical significance (all  $ps > .15$ ). Classification accuracies are shown in Figure A2.

## Acknowledgments

The authors thank Corrine Durisko and Ruizhe Liu for their assistance and discussions about developing the study, Heather Bruett and John Paulus for their helpful comments and suggestions on an earlier draft of this article, Claire Kollhoff and Aarya Wadke for their assistance with stimuli creation, and Mark Vignone and Jasmine Issa for their assistance in running participants. We also thank Jessica Cantlon for making stimuli and experimental procedures available to us and Lukas Snoek for consultation on the classification procedure used here.

Reprint requests should be sent to Griffin E. Koch, 3420 Forbes Avenue Office 512, Pittsburgh, PA, USA 15260, or via e-mail: griffinkoch@pitt.edu.

## Data Availability Statement

Data presented here will be made available upon reasonable request that complies with policies of the IRB.

## Funding Information

This work was supported by the National Sciences Foundation (grant number 1734735); and Behavioral Brain Research Training Program through the National Institutes of Health (grant number T32GM081760 to G. E. K.).

## Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle



and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .541; W/M = .189; M/W = .162; W/W = .108.

## Notes

1. In addition, we report results for the following individual hemispheres (as opposed to the full bilateral ROIs): left IPS ( $M = 0.51$ ,  $SD = 0.04$ ;  $t(17) = 0.75$ ,  $p = .461$ ;  $p$  (perm.) = .288), right IPS ( $M = 0.53$ ,  $SD = 0.05$ ;  $t(17) = 3.01$ ,  $p = .008$ ;  $p$  (perm.) = .010), left insula ( $M = 0.52$ ,  $SD = 0.07$ ;  $t(17) = 1.42$ ,  $p = .173$ ;  $p$  (perm.) = .052), right insula ( $M = 0.50$ ,  $SD = 0.04$ ;  $t(17) = 0.08$ ,  $p = .938$ ;  $p$  (perm.) = .456), left fusiform gyrus ( $M = 0.50$ ,  $SD = 0.06$ ;  $t(17) = 0.07$ ,  $p = .949$ ;  $p$  (perm.) = .489), right fusiform gyrus ( $M = 0.51$ ,  $SD = 0.05$ ;  $t(17) = 1.08$ ,  $p = .293$ ;  $p$  (perm.) = .199), left calcarine sulcus ( $M = 0.51$ ,  $SD = 0.07$ ;  $t(17) = 0.32$ ,  $p = .755$ ;  $p$  (perm.) = .333), right calcarine sulcus ( $M = 0.49$ ,  $SD = 0.07$ ;  $t(17) = -0.50$ ,  $p = .626$ ;  $p$  (perm.) = .716), left precentral gyrus ( $M = 0.52$ ,  $SD = 0.04$ ;  $t(17) = 1.65$ ,  $p = .118$ ;  $p$  (perm.) = .081), right precentral gyrus ( $M = 0.49$ ,  $SD = 0.05$ ;  $t(17) = -0.88$ ,  $p = .388$ ;  $p$  (perm.) = .747), left postcentral gyrus ( $M = 0.51$ ,  $SD = 0.05$ ;  $t(17) = 0.91$ ,  $p = .376$ ;  $p$  (perm.) = .209), right postcentral gyrus ( $M = 0.50$ ,  $SD = 0.05$ ;  $t(17) = 0.28$ ,  $p = .782$ ;  $p$  (perm.) = .377).

2. In addition, we report the right IPS classifying above chance during the *consistent* analyses ( $M = 0.53$ ,  $SD = 0.06$ ;  $t(17) = 2.02$ ,  $p = .060$ ;  $p$  (perm.) = .027).

3. In addition, we report results for the following individual hemispheres (as opposed to the full bilateral ROIs): left IPS ( $M = 0.55$ ,  $SD = 0.07$ ;  $t(17) = 2.87$ ,  $p = .011$ ;  $p$  (perm.) = .001), right IPS ( $M = 0.57$ ,  $SD = 0.08$ ;  $t(17) = 3.58$ ,  $p = .002$ ;  $p$  (perm.) < .001), left insula ( $M = 0.51$ ,  $SD = 0.05$ ;  $t(17) = 0.91$ ,  $p = .374$ ;  $p$  (perm.) = .212), right insula ( $M = 0.49$ ,  $SD = 0.06$ ;  $t(17) = -0.76$ ,  $p = .460$ ;  $p$  (perm.) = .768), left fusiform gyrus ( $M = 0.56$ ,  $SD = 0.09$ ;  $t(17) = 2.95$ ,  $p = .009$ ;  $p$  (perm.) < .001), right fusiform gyrus ( $M = 0.61$ ,  $SD = 0.07$ ;  $t(17) = 6.81$ ,  $p < .001$ ;  $p$  (perm.) < .001), left calcarine sulcus ( $M = 0.57$ ,  $SD = 0.08$ ;  $t(17) = 3.76$ ,  $p = .002$ ;  $p$  (perm.) < .001), right calcarine sulcus ( $M = 0.56$ ,  $SD = 0.07$ ;  $t(17) = 3.74$ ,  $p = .002$ ;  $p$  (perm.) < .001), left precentral gyrus ( $M = 0.49$ ,  $SD = 0.05$ ;  $t(17) = -0.86$ ,  $p = .401$ ;  $p$  (perm.) = .757), right precentral gyrus ( $M = 0.50$ ,  $SD = 0.04$ ;  $t(17) = 0.57$ ,  $p = .573$ ;  $p$  (perm.) = .383), left postcentral gyrus ( $M = 0.50$ ,  $SD = 0.07$ ;  $t(17) = -0.07$ ,  $p = .942$ ;  $p$  (perm.) = .541), right postcentral gyrus ( $M = 0.51$ ,  $SD = 0.06$ ;  $t(17) = 0.50$ ,  $p = .621$ ;  $p$  (perm.) = .314).

4. In addition, we report results for the following individual hemisphere, which showed ability to classify above chance during the *consistent* analyses: left IPS ( $M = 0.53$ ,  $SD = 0.05$ ;  $t(17) = 2.31$ ,  $p = .033$ ;  $p$  (perm.) = .011), right IPS ( $M = 0.54$ ,  $SD = 0.07$ ;  $t(17) = 2.56$ ,  $p = .020$ ;  $p$  (perm.) < .001), left fusiform gyrus ( $M = 0.55$ ,  $SD = 0.09$ ;  $t(17) = 2.46$ ,  $p = .025$ ;  $p$  (perm.) < .001), right fusiform gyrus ( $M = 0.58$ ,  $SD = 0.10$ ;  $t(17) = 3.58$ ,  $p = .002$ ;  $p$  (perm.) < .001), left calcarine sulcus ( $M = 0.57$ ,  $SD = 0.07$ ;  $t(17) = 4.90$ ,  $p < .001$ ;  $p$  (perm.) < .001), right calcarine sulcus ( $M = 0.55$ ,  $SD = 0.07$ ;  $t(17) = 3.04$ ,  $p = .007$ ;  $p$  (perm.) < .001).

## REFERENCES

Abboud, S., Maidenbaum, S., Dehaene, S., & Amedi, A. (2015). A number-form area in the blind. *Nature Communications*, 6, 6026. <https://doi.org/10.1038/ncomms7026>, PubMed: 25613599

Andres, M., & Pesenti, M. (2015). Finger-based representation of mental arithmetic. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 67–88). Oxford University Press. <https://doi.org/10.1093/oxfordhob/9780199642342.013.028>

Anzellotti, S., & Coutanche, M. N. (2018). Beyond functional connectivity: Investigating networks of multivariate representations. *Trends in Cognitive Sciences*, 22, 258–269. <https://doi.org/10.1016/j.tics.2017.12.002>, PubMed: 29305206

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., & Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. *Developmental Cognitive Neuroscience*, 30, 239–250. <https://doi.org/10.1016/j.dcn.2017.08.002>, PubMed: 28844728

Arsalidou, M., & Taylor, M. J. (2011). Is  $2 + 2 = 4$ ? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage*, 54, 2382–2393. <https://doi.org/10.1016/j.neuroimage.2010.10.009>, PubMed: 20946958

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>

Berteletti, I., Prado, J., & Booth, J. R. (2014). Children with mathematical learning disability fail in recruiting verbal and numerical brain regions when solving simple multiplication problems. *Cortex*, 57, 143–155. <https://doi.org/10.1016/j.cortex.2014.04.001>, PubMed: 24858066

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. <https://doi.org/10.1163/156856897X00357>, PubMed: 9176952

Bugden, S., Price, G. R., McLean, D. A., & Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Developmental Cognitive Neuroscience*, 2, 448–457. <https://doi.org/10.1016/j.dcn.2012.04.001>, PubMed: 22591861

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *Neuroimage*, 87, 311–322. <https://doi.org/10.1016/j.neuroimage.2013.10.049>, PubMed: 24201011

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual number beats abstract numerical magnitude: Format-dependent representation of Arabic digits and dot patterns in human parietal cortex. *Journal of Cognitive Neuroscience*, 27, 1376–1387. [https://doi.org/10.1162/jocn\\_a\\_00787](https://doi.org/10.1162/jocn_a_00787), PubMed: 25633646

Butterworth, B. (1999). *The mathematical brain*. Macmillan.

Cantlon, J. F., & Li, R. (2013). Neural activity during natural viewing of sesame street statistically predicts test scores in early childhood. *PLoS Biology*, 11, e1001462. <https://doi.org/10.1371/journal.pbio.1001462>, PubMed: 23300385

Cappelletti, M., Barth, H., Fregni, F., Spelke, E. S., & Pascual-Leone, A. (2007). rTMS over the intraparietal sulcus disrupts numerosity processing. *Experimental Brain Research*, 179, 631–642. <https://doi.org/10.1007/s00221-006-0820-0>, PubMed: 17216413

Cohen Kadosh, R. (2008). Numerical representation: Abstract or nonabstract? *Quarterly Journal of Experimental Psychology*, 61, 1160–1168. <https://doi.org/10.1080/17470210801994989>, PubMed: 18609377

Coutanche, M. N. (2013). Distinguishing multi-voxel patterns and mean activation: Why, how, and what does it tell us? *Cognitive, Affective, & Behavioral Neuroscience*, 13, 667–673. <https://doi.org/10.3758/s13415-013-0186-2>, PubMed: 23857415

Coutanche, M. N., & Thompson-Schill, S. L. (2013). Informational connectivity: Identifying synchronized

- discriminability of multi-voxel patterns across the brain. *Frontiers in Human Neuroscience*, 7, 15. <https://doi.org/10.3389/fnhum.2013.00015>, PubMed: 23403700
- Coutanche, M. N., & Thompson-Schill, S. L. (2014). Using informational connectivity to measure the synchronous emergence of fMRI multi-voxel information across time. *Journal of Visualized Experiments*, 89, 51226. <https://doi.org/10.3791/51226>, PubMed: 25046335
- Coutanche, M. N., & Thompson-Schill, S. L. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex*, 25, 2584–2593. <https://doi.org/10.1093/cercor/bhu057>, PubMed: 24692512
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29, 162–173. <https://doi.org/10.1006/cbmr.1996.0014>, PubMed: 8812068
- Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., & Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage*, 97, 271–283. <https://doi.org/10.1016/j.neuroimage.2014.04.037>, PubMed: 24768930
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N), PubMed: 1511583
- Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20, 487–506. <https://doi.org/10.1080/02643290244000239>, PubMed: 20957581
- Detre, G., Polyn, S. M., Moore, C., Natu, V., Singer, B., Cohen, J., et al. (2006). The Multi-Voxel Pattern Analysis (MVPA) toolbox. Paper presented at Annual Meeting of the Organization for Human Brain Mapping. Florence, Italy.
- Di Luca, S., & Pesenti, M. (2011). Finger numeral representations: More than just another symbolic code. *Frontiers in Psychology*, 2, 272. <https://doi.org/10.3389/fpsyg.2011.00272>, PubMed: 22069392
- Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. *Current Biology*, 19, 1608–1615. <https://doi.org/10.1016/j.cub.2009.08.047>, PubMed: 19781939
- Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., & Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. *Neuron*, 37, 719–726. [https://doi.org/10.1016/S0896-6273\(03\)00036-9](https://doi.org/10.1016/S0896-6273(03)00036-9), PubMed: 12597867
- Emerson, R. W., & Cantlon, J. F. (2015). Continuity and change in children's longitudinal neural response to numbers. *Developmental Science*, 18, 314–326. <https://doi.org/10.1111/desc.12215>, PubMed: 25051893
- Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience*, 7, 357–375. <https://doi.org/10.1162/jocn.1995.7.3.357>, PubMed: 23961866
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33, 341–355. [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X)
- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., et al. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23(Suppl. 1), S69–S84. <https://doi.org/10.1016/j.neuroimage.2004.07.016>, PubMed: 15501102
- Floren, A., Naylor, B., & Mikkiläinen, R. (2015). Accurately decoding visual information from fMRI data obtained in a realistic virtual environment. *Frontiers in Human Neuroscience*, 9, 327. <https://doi.org/10.3389/fnhum.2015.00327>, PubMed: 26106315
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, 88, 121–151. <https://doi.org/10.1016/j.jecp.2004.03.002>, PubMed: 15157755
- 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>, PubMed: 26940623
- Grotheer, M., Herrmann, K.-H., & Kovács, G. (2016). Neuroimaging evidence of a bilateral representation for visually presented numbers. *Journal of Neuroscience*, 36, 88–97. <https://doi.org/10.1523/JNEUROSCI.2129-15.2016>, PubMed: 26740652
- 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, 188–200. <https://doi.org/10.1016/j.neuroimage.2018.03.064>, PubMed: 29604456
- Hindy, N. C., Altmann, G. T., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *Journal of Neuroscience*, 32, 5795–5803. <https://doi.org/10.1523/JNEUROSCI.6294-11.2012>, PubMed: 22539841
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6, 435–448. <https://doi.org/10.1038/nrn1684>, PubMed: 15928716
- Kadosh, R. C., Kadosh, K. C., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-dependent and-independent representations of numbers in the parietal lobes. *Neuron*, 53, 307–314. <https://doi.org/10.1016/j.neuron.2006.12.025>, PubMed: 17224410
- Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L.-B., et al. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex*, 44, 376–385. <https://doi.org/10.1016/j.cortex.2007.08.003>, PubMed: 18387568
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36, 1–16.
- Knops, A. (2017). Probing the neural correlates of number processing. *Neuroscientist*, 23, 264–274. <https://doi.org/10.1177/1073858416650153>, PubMed: 27188457
- Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, 36, 475–488. <https://doi.org/10.1002/hbm.22641>, PubMed: 25238646
- Lyons, I. M., & Beilock, S. L. (2018). Characterizing the neural coding of symbolic quantities. *Neuroimage*, 178, 503–518. <https://doi.org/10.1016/j.neuroimage.2018.05.062>, PubMed: 29857048
- Myers, T., & Szűcs, D. (2015). Arithmetic memory is modality specific. *PLoS One*, 10, e0145614. <https://doi.org/10.1371/journal.pone.0145614>, PubMed: 26716692
- Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, 32, 185–208. <https://doi.org/10.1146/annurev.neuro.051508.135550>, PubMed: 19400715
- Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, 45, S199–S209. <https://doi.org/10.1016/j.neuroimage.2008.11.007>, PubMed: 19070668

- Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*, 293–305. <https://doi.org/10.1016/j.neuron.2006.11.022>, PubMed: 17224409
- Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., et al. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*, 1932–1947. <https://doi.org/10.1002/hbm.21159>, PubMed: 21246667
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 7865–7870. <https://doi.org/10.1073/pnas.0509989103>, PubMed: 16682637
- Regev, M., Honey, C. J., Simony, E., & Hasson, U. (2013). Selective and invariant neural responses to spoken and written narratives. *Journal of Neuroscience*, *33*, 15978–15988. <https://doi.org/10.1523/JNEUROSCI.1580-13.2013>, PubMed: 24089502
- Schmithorst, V. J., & Brown, R. D. (2004). Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group independent component analysis of the mental addition and subtraction of fractions. *Neuroimage*, *22*, 1414–1420. <https://doi.org/10.1016/j.neuroimage.2004.03.021>, PubMed: 15219612
- Schweisfurth, M. A., Frahm, J., Farina, D., & Schweizer, R. (2018). Comparison of fMRI digit representations of the dominant and non-dominant hand in the human primary somatosensory cortex. *Frontiers in Human Neuroscience*, *12*, 492. <https://doi.org/10.3389/fnhum.2018.00492>, PubMed: 30618677
- Shuman, M., & Kanwisher, N. (2004). Numerical magnitude in the human parietal lobe: Tests of representational generality and domain specificity. *Neuron*, *44*, 557–569. <https://doi.org/10.1016/j.neuron.2004.10.008>, PubMed: 15504334
- Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*, 475–487. [https://doi.org/10.1016/s0896-6273\(02\)00575-5](https://doi.org/10.1016/s0896-6273(02)00575-5), PubMed: 11832233
- Skagenholt, M., Träff, U., Västfjäll, D., & Skagerlund, K. (2018). Examining the triple code model in numerical cognition: An fMRI study. *PLoS One*, *13*, e0199247. <https://doi.org/10.1371/journal.pone.0199247>, PubMed: 29953456
- Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*, *184*, 741–760. <https://doi.org/10.1016/j.neuroimage.2018.09.074>, PubMed: 30268846
- Talairach, J. (1988). *Co-planar stereotaxic atlas of the human brain: 3-D proportional system: An approach to cerebral imaging* (1st ed.). Thieme.
- Valente, G., Castellanos, A. L., Hausfeld, L., De Martino, F., & Formisano, E. (2021). Cross-validation and permutations in MVPA: Validity of permutation strategies and power of cross-validation schemes. *Neuroimage*, *238*, 118145. <https://doi.org/10.1016/j.neuroimage.2021.118145>, PubMed: 33961999
- Wilkey, E. D., Conrad, B. N., Yeo, D. J., & Price, G. R. (2020). Shared numerosity representations across formats and tasks revealed with 7 tesla fMRI: Decoding, generalization, and individual differences in behavior. *Cerebral Cortex Communications*, *1*, tgaa038. <https://doi.org/10.1093/texcom/tgaa038>, PubMed: 34296107
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670. <https://doi.org/10.1038/nmeth.1635>, PubMed: 21706013