



A preference for mathematical processing outweighs the selectivity for Arabic numbers in the inferior temporal gyrus

Mareike Grotheer^{a,*}, Brianna Jeska^a, Kalanit Grill-Spector^{a,b,c}

^a Psychology Department, Stanford University, Stanford, CA, 94305, USA

^b Neurosciences Program, Stanford University School of Medicine, Stanford, CA, 94305, USA

^c Stanford Neurosciences Institute, Stanford University, Stanford, CA, 94305, USA



ARTICLE INFO

Keywords:

Number form area
Visual word form area
Inferior temporal gyrus
Mathematical cognition
Numerosity
Vision

ABSTRACT

A region in the posterior inferior temporal gyrus (ITG), referred to as the number form area (NFA, here ITG-numbers) has been implicated in the visual processing of Arabic numbers. However, it is unknown if this region is specifically involved in the visual encoding of Arabic numbers per se or in mathematical processing more broadly. Using functional magnetic resonance imaging (fMRI) during experiments that systematically vary tasks and stimuli, we find that mathematical processing, not preference to Arabic numbers, consistently drives both mean and distributed responses in the posterior ITG. While we replicated findings of higher responses in ITG-numbers to numbers than other visual stimuli during a 1-back task, this preference to numbers was abolished when participants engaged in mathematical processing. In contrast, an ITG region (ITG-math) that showed higher responses during an adding task vs. other tasks maintained this preference for mathematical processing across a wide range of stimuli including numbers, number/letter morphs, hands, and dice. Analysis of distributed responses across an anatomically-defined posterior ITG expanse further revealed that mathematical task but not Arabic number form can be successfully and consistently decoded from these distributed responses. Together, our findings suggest that the function of neuronal regions in the posterior ITG goes beyond the specific visual processing of Arabic numbers. We hypothesize that they ascribe numerical content to the visual input, irrespective of the format of the stimulus.

Introduction

The ability to perform basic math is crucial for our daily lives, yet how our brain supports this skill is not fully understood. Recent research has identified an area in the human posterior inferior temporal gyrus (ITG), which responds more strongly to Arabic numbers relative to other visual stimuli (Shum et al., 2013; Grotheer et al., 2016b). This region has previously been referred to as the ‘number form area’ (NFA) (Shum et al., 2013). Since this name does not convey information about anatomical location, we will refer to these number-selective activations in the posterior ITG as ‘ITG-numbers’ (thereby indicating both anatomical location and functional preference, following the same naming convention we use for other category-selective regions (Grill-Spector and Weiner, 2014)). ITG-numbers is functionally connected to both areas in the visual system and areas implicated in numerosity processing (Abboud et al., 2015). It is currently thought to be involved in detecting (Grotheer et al., 2016a) and processing (Shum et al., 2013) Arabic numbers. However, localizing

ITG-numbers remains challenging, and in fact several studies failed to localize it based on higher responses to numbers vs. other stimuli (Libertus et al., 2009; Price and Ansari, 2011; Park et al., 2012; Peters et al., 2015; for a recent meta-analysis see Yeo et al., 2017).

In contrast, other recent studies report activations in the posterior ITG during a variety of mathematical tasks, even in the absence of visually presented numbers. For instance, a study in congenitally blind participants, found activations in the posterior ITG during an auditory numerosity task vs. color and letter discrimination tasks (Abboud et al., 2015). Additionally, higher responses in the posterior ITG were reported in expert mathematicians for auditorily presented mathematical relative to non-mathematical statements (Amalric and Dehaene, 2016). Further, electrocorticography (ECoG) data showed increased posterior ITG responses (Hermes et al., 2015) as well as increased functional coupling between the posterior ITG and areas involved in numerosity processing (Daitch et al., 2016) during mathematical calculations. These studies suggest an alternative view in which the posterior ITG is involved in

* Corresponding author. Psychology Department, Stanford University, Stanford, CA 94305, USA..

E-mail address: grotheer@stanford.edu (M. Grotheer).

<https://doi.org/10.1016/j.neuroimage.2018.03.064>

Received 11 November 2017; Received in revised form 24 March 2018; Accepted 27 March 2018

Available online 28 March 2018

1053-8119/© 2018 Elsevier Inc. All rights reserved.

stimulus-independent mathematical processing rather than the visual processing of Arabic numbers specifically. In the present study, we will refer to a region in the posterior ITG showing higher activations during a mathematical task relative to other tasks as ITG-math.

One possible resolution to this debate on the function of the posterior ITG is that it is functionally heterogeneous. That is, within this anatomical expanse there are different subregions, one involved in visual processing of Arabic numbers (ITG-numbers) and another involved in mathematical processing (ITG-math). Support for this hypothesis comes from a recent ECoG study (Daitch et al., 2016), which reported that different electrodes within the ITG show either (i) a preference for numbers vs. characters together with a preference for mathematical processing vs. working memory or (ii) a preference for mathematical processing vs. working memory but no preference for Arabic numbers. However, this finding does not explain why localizing ITG-numbers based on a preference for numbers is rarely successful (Yeo et al., 2017). Further, due to the sparse sampling by ECoG electrodes, the spatial nature of this putative heterogeneity of the ITG remains unknown.

Resolving this debate about the function of the posterior ITG critically requires (i) measuring ITG responses when task and stimulus are systematically varied within the same experiment and (ii) an experimental technique which simultaneously allows precise measurements and coverage of the entire posterior ITG in each participant. To fill this fundamental gap in knowledge, we conducted three fMRI experiments in which we systematically varied both the visual stimulus and the participants' task, and measured mean and distributed ITG responses. Crucially, in contrast to previous studies (Abboud et al., 2015; Amalric and Dehaene, 2016; Grotheer et al., 2016b), all analyses were conducted in the participants' native anatomical space and without spatial smoothing, as this approach improves the spatial precision of the acquired measurements (Weiner and Grill-Spector, 2013).

We tested the following hypotheses: H1: A region in the posterior ITG (ITG-numbers) is involved in the visual encoding of Arabic numbers, but not other visual stimuli (as proposed by Shum et al., 2013). This predicts (i) higher responses for numbers than other stimuli across experiments, independently of the participant's task and (ii) that the number stimulus can be decoded from distributed responses. H2: A region in the ITG (ITG-math) is involved in stimulus-independent mathematical processing (as suggested by studies using auditory stimuli (Abboud et al., 2015; Amalric and Dehaene, 2016)). This predicts (i) higher responses during mathematical tasks than other tasks across experiments and independently of the stimulus, and (ii) that mathematical task can be decoded from distributed responses. H3: The posterior ITG is functionally heterogeneous (i.e. contains both ITG-numbers and ITG-math; as proposed by Daitch et al., 2016). This hypothesis predicts (i) distinct subregions within the ITG that consistently prefer either mathematical tasks, numbers or both and (ii) that both numbers and mathematical task can be decoded from distributed ITG responses. H4: A region in the ITG is involved in the visual encoding of Arabic numbers only in the context of mathematical processing. This hypothesis has not been considered previously and predicts (i) an interaction between task and stimulus, such that responses within the posterior ITG will be highest for number stimuli during mathematical tasks and (ii) that distributed responses across the ITG carry information about number stimuli only during mathematical tasks.

Materials and methods

Participants

15 volunteers (8 female, age: 21–48, 1 left-handed) were recruited from Stanford University and participated in all three experimental sessions. One experiment was repeated in one subject, as motion exceeded our threshold of 2 voxels within-run motion or 3.5 voxels between-run motion. Subjects had normal or corrected to normal vision and gave their informed written consent. The Stanford Internal Review Board on Human Subjects Research approved all procedures.

Data acquisition and preprocessing

Acquisition: Data was collected at the Center for Cognitive and Neurobiological Imaging at Stanford University, using a GE 3 tesla Signa Scanner with a 32-channel head coil. We acquired 48 slices covering the occipitotemporal and most of the frontal cortex using a T2*-sensitive gradient echo sequence (resolution: 2.4 mm × 2.4 mm × 2.4 mm, TR: 1000 ms, TE: 30 ms, FoV: 192 mm, flip angle: 62°, multiplexing factor of 3). A whole-brain, anatomical volume was also acquired, once for each participant, using a T1-weighted BRAVO pulse sequence (resolution: 1 mm × 1 mm × 1 mm, TI = 450 ms, flip angle: 12°, 1 NEX, FoV: 240 mm).

Processing: The anatomical brain volume of each subject was segmented into gray and white matter using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>), with manual corrections using ITKGray (<http://web.stanford.edu/group/vista/cgi-bin/wiki/index.php/ItkGray>), and each participant's cortical surface was reconstructed.

Functional data was analyzed using the mrVista toolbox (<http://github.com/vistalab>) for Matlab, as in previous work (Weiner and Grill-Spector, 2010, 2011). fMRI data from each experiment was motion-corrected, both within and between runs, and then aligned to the anatomical volume. To increase the spatial precision of our data compared to previous studies (Abboud et al., 2015; Amalric and Dehaene, 2016; Grotheer et al., 2016b), no smoothing was applied. For details on the effects of smoothing on fMRI data, see Weiner and Grill-Spector (2013). The time course of each voxel was high-pass filtered with a 1/20 Hz cutoff and it was transformed from arbitrary units to percentage signal change. In each experiment, a separate design matrix was created and convolved with the hemodynamic response function (HRF) implemented in SPM (<http://www.fil.ion.ucl.ac.uk/spm>) to generate predictors for each experimental condition. Regularized response coefficients (betas) were estimated for each voxel and each predictor using a general linear model (GLM) indicating the magnitude of response for that condition.

Experimental design

Experiment 1

This experiment was designed to detect voxels which show a preference for visually presented numbers vs. other visual stimuli. Similar to a previous study (Grotheer et al., 2016b), subjects were presented with greyscale images of numbers (N), letters (L), objects (O), Fourier phase scrambled numbers (SN), Fourier phase scrambled letters (SL), false numbers (FN), and false letters (FL) (Fig. 1-Exp 1). The false numbers and letters were created by relocating 1–4 components of the original numbers and letters to create unfamiliar shapes (similar to Price and Ansari, 2011; Grotheer and Kovacs, 2014; Grotheer et al., 2016b). The size of the stimuli alternated between ~5° and ~7° of visual angle and they were presented in the center of the screen. Stimuli repeated once across the experiment, but never within the same run (apart from 1-back targets). Subjects viewed 4 s long trials presenting 8 images of the same type at a rate of 2 Hz, as this rate has been shown to be the optimum for character selective regions (Stigliani et al., 2015). The trial order was counterbalanced, such that each stimulus category was equally likely to be preceded by each of the other stimulus categories. 22% of all trials were blank trials, during which no images were presented. Each participant completed eight 5 min long runs, which included 9 repetitions of each stimulus category. Participants were asked to fixate on a black dot in the center of the screen and to indicate immediate repetitions of an image (1-back task). Repetitions occurred once per trial in 1/3 of the trials.

Experiment 2

The second experiment was designed to compare brain responses to well-controlled character-like stimuli under different tasks (adding, reading, and remembering colors, Fig. 1-Exp 2).

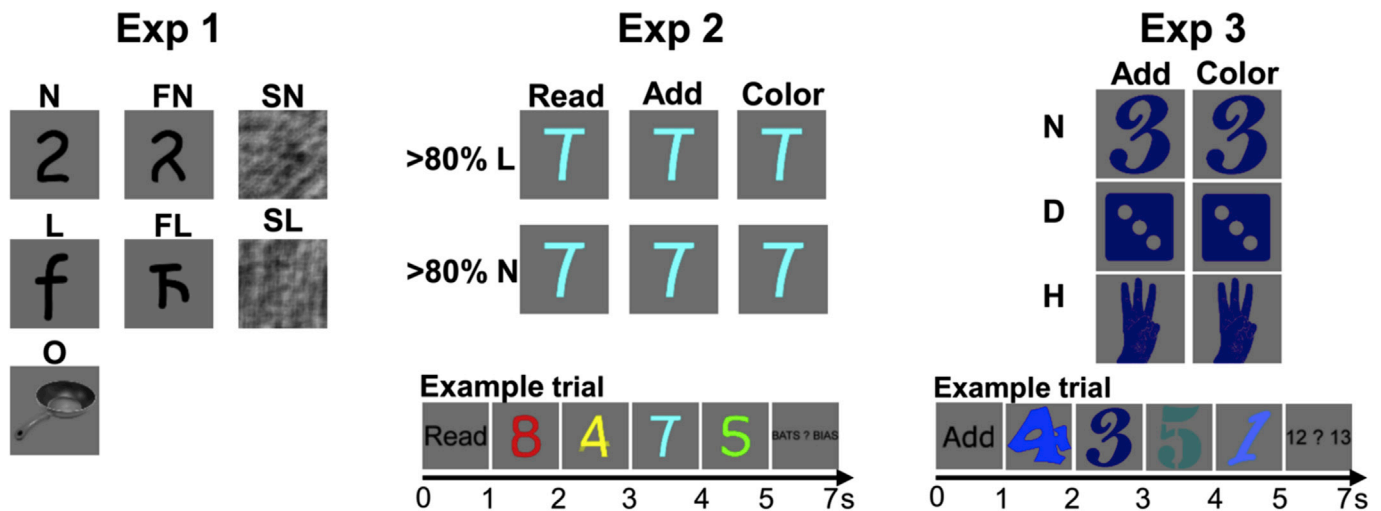


Fig. 1. Experimental design.

Exp 1: Subjects viewed numbers (N), letters (L), objects (O), false numbers (FN), false letters (FL), Fourier phase scrambled numbers (SN) and Fourier phase scrambled letters (SL) in 4 s long blocks, while performing a 1-back task.

Exp 2: Subjects viewed morphs between numbers and letters, containing either >80% letter (<20% number) or >80% number (<20% letter) information. They were instructed to read the word, add the numbers, or remember the colors of the stimuli. At the beginning of each trial, a cue (“Read”/“Add”/“Color”) indicated which task should be performed, then 4 stimuli of the same morph type appeared for 1 s each, followed by an answer screen presented for 2 s. Subjects indicated their answer with a button press. Identical stimuli were presented across tasks. Trial structure is shown at the bottom.

Exp 3: Subjects viewed numbers, dice or hands, indicating a numerical value. They were instructed to either add the numerical value of the stimuli or remember their colors. At the beginning of each trial, a cue (“Add”/“Color”) indicated which task should be performed, then 4 stimuli of the same category appeared for 1 s each, followed by an answer screen presented for 2 s. Subjects indicated their answer with a button press. Identical stimuli were presented across tasks.

Stimuli: Each image was a morph of a number and a letter. All images in a trial were either number morphs (N, >80% number + <20% letter) or letter morphs (L, >80% letter + <20% number), i.e. the stimuli mostly contained information from one category, but held just enough evidence from the other category to be recognizable as both letters and numbers. The morphs used in this experiment were created using FantaMorph (<http://www.fantamorph.com>; the pairings were: A-4, B-8, G-6, I-1, O-0, S-5, T-7 and Z-2). Amazon Mechanical Turk (<https://www.mturk.com>) was used to determine morph levels that were perceived as numbers more than 80% of the time and as letters more than 80% of the time. For this, 15 steps along each morph line were chosen and randomly grouped into 5 lists. Workers were instructed to decide if the presented stimulus resembles a number or a letter and each image was judged by 50 workers. To ensure that the task was performed accurately, a black/white judgment task as well as at least 2 original numbers and letters were included on each list and workers could make no more than 2 mistakes on these controls.

Stimuli in a trial were designed to form a 4-letter long English word. 24 such groups of stimuli were created. Crucially, each of the 24 groups (words) was presented twice in each of the task conditions, once presented with number morphs and once with letter morphs. Stimuli appeared in one of 8 colors: red [RGB: 255 0 0], orange [RGB: 255 128 0], yellow [RGB: 255 255 0], green [RGB: 0 255 0], cyan [RGB: 0 255 255], blue [RGB: 0 0 255], magenta [RGB: 255 0 255], and rose [RGB: 255 0 128].

Trial structure (Fig. 1-Exp2, bottom): At the beginning of each trial subjects were presented with a cue (“Add”, “Read” or “Color”), indicating which task they should perform. After the cue, 4 images were shown in a row, followed by an answer screen. Each image was presented in a different color for 0.9 s (0.1 s inter-stimulus interval) at the center of the screen, and subtended a visual angle of $\sim 7^\circ$. The answer screen was presented for 2 s and showed the correct answer as well as one incorrect answer at counterbalanced locations left and right of fixation.

Task: Participants performed three different tasks on the same stimuli: *Adding:* participants were asked to sum the values of the four presented stimuli and indicate which of the two answers on the answer

screen conveyed the correct sum. *Reading:* subjects were instructed to read the word in their head, and to indicate which of the two words on the answer screen had been presented. *Color:* participants were asked to memorize the color of the stimuli and to indicate which of the colors on the answer screen was shown during the trial. In 25% of the trials only the cue or only the answer screen was presented, so that blood-oxygen-level dependent (BOLD) responses induced by the images in the trial could be isolated. Participants performed six 6 min long runs. Each experimental condition was repeated 4 times per run, but each group of stimuli (word) was presented only once in each run. Prior to the experiment, subjects were given training to ensure that they could perform the task with at least 80% accuracy.

Experiment 3

This experiment was designed to compare brain responses to visually dissimilar stimuli (numbers, hands, and dice) under different tasks (adding and remembering colors, Fig. 1-Exp 3).

Stimuli: Each trial contained 4 images from a single category (numbers (N), dice (D) or hands (H)) which displayed numerical values between 1 and 5. To reduce the impact of low-level differences between the stimuli, their size and their location on the screen was jittered (size: jittered in 5 steps between ~ 2.85 and ~ 3.50 degrees visual angle; position: center position as well as center position with a shift of 1/6 stimulus size in upper, lower, right and left direction). To match task difficulty between the adding and color tasks, stimuli were presented in shades of blue (midnight [RGB: 0 0 128], true blue [RGB: 0 0 255], cyan [RGB: 0 255 255], teal [0 128 128] and light blue [RGB: 0 128 255]). 24 groups of numerical values and colors were created randomly in the beginning of the experiment and were presented once per run and once in each experimental condition. Colors and numerical values could repeat within a block, but the same image never repeated within the experiment.

Trial structure: the trial structure and timing was identical to Exp 2 (Fig. 1-Exp 3, bottom).

Task: *Adding:* participants were asked to sum the values of the four presented stimuli and indicate which of the two answers on the answer screen conveyed the correct sum. *Color:* participants were asked to

memorize the color of the stimuli and indicate which of the colors on the answer screen was shown during the trial.

Regions of interest definition

Functional Regions of Interest (fROIs) were defined on an individual participant level using both functional and anatomical criteria, and were named either per their anatomical location or in accordance with previously published naming conventions. We defined two sets of fROIs: (i) using data from all runs ($T = 3$, $p < 0.0015$, voxel level, uncorrected, $df = 2673$ in Exp 1, $df = 2309$ in Exp 2, $df = 2311$ in Exp 3) and (ii) using data from half of the runs (odd or even runs, $T = 2.5$, $p < 0.0063$, voxel level, uncorrected, $df = 1333$ in Exp 1, $df = 1145$ in Exp 2 and $df = 1147$ in Exp 3). We used a lower threshold for the latter, as it contains only half of the data, and therefore has lower statistical power to find significant effects. fROIs that were defined using all runs were used to analyze independent data from the other experiments (e.g. define fROI with Exp 1 data, and extract responses from Exp 2 and Exp 3). fROIs that were defined using half of the data were used to analyze the left-out, independent data from the same experiment (e.g. define with odd runs and extract responses from even runs, and vice versa). Crucially, since spatial smoothing and group based analyses can lead to spurious overlap between fROIs (Weiner and Grill-Spector, 2013), all our fROIs were defined in each participant's native anatomical space and no spatial smoothing was applied.

We used stringent anatomical criteria for defining the fROIs. ITG fROIs only included voxels within the posterior ITG and neighboring sulci. As we only evaluated a small part of the cortex, we did not perform statistical correction for multiple comparisons. We also did not set any size restriction for our fROIs as a previous study (Daitch et al., 2016) suggested that number selective responses arise from a smaller region within the ITG than math selective responses, and any a priori size restriction could have therefore biased the outcome.

We defined two sets of fROIs within the posterior ITG:

- (1) *ITG-numbers* was defined based on stimulus preference. In accordance with a previous study (Grotheer et al., 2016b), it included all voxels in the posterior ITG and adjacent sulci that showed significantly higher responses to numbers than all other stimuli in Exp 1. It was detected bilaterally in 13 subjects when using data from all runs and in 11 subjects in the split-half analysis. An example ITG-numbers from a representative subject is shown in Fig. 3 and all individual subjects' ITG-numbers fROIs are shown in Figs S1 and S2. ITG-numbers could not be defined reliably using a more conservative conjunction analysis (see Fig S3). The MNI coordinates of ITG-numbers [MNI (\pm standard error (SE))]: right hemisphere: 57 (1), -54 (1), -14 (2); left hemisphere: -54 (2), -59 (2), -12 (2)] closely match the coordinates reported for the NFA in previous studies (e.g. Grotheer et al. (2016b): [MNI(\pm SE): right hemisphere: 55 (2), -50 (2), -15 (1); left hemisphere: -54 (2), -55 (2), -13 (2)]). To evaluate the reproducibility of the preference for numbers in the posterior ITG, we also defined additional ITG-numbers fROIs based on higher responses to numbers vs. dice and hands in Exp 3 ($T = 3$, voxel level,

uncorrected, bilateral $N = 9$). Then we evaluated the overlap between these fROIs and the ones defined based on Exp 1 data using a dice-coefficient analysis.

- (2) *ITG-math* was defined based on task preference and included voxels in the posterior ITG and adjacent sulci that showed higher responses in the adding task vs. the reading and the color task in Exp 2 (a contrast comparable to Abboud et al., 2015). We identified ITG-math bilaterally in 13 subjects using data from all runs, as well as in 13 subjects in the split-half analysis. An example of ITG-math in a representative subject is shown in Fig. 4, all individual subjects' ITG-math fROIs are shown in Fig S1 and S2. ITG-math could also have been defined using a more conservative conjunction analysis (Fig S3), but we chose to use a similar approach as has been used in previous studies (Grotheer et al., 2016b). The MNI coordinates of ITG-math are: [MNI(\pm SE): right hemisphere: 57 (1), -56 (2), -12 (2); left hemisphere: -54 (2), -60 (2), -11 (1)]. Please note that, even though the MNI coordinates of ITG-numbers and ITG-math are similar, individual subject analyses allowed us to show that these fROIs are not overlapping significantly (see section 3.4 and Fig. 5). To evaluate the reproducibility of the task preference, we also defined ITG-math using Exp 3 data, as voxels within the posterior ITG that showed significantly higher responses in the adding vs. the color task ($T = 3$, voxel level, uncorrected, bilateral $N = 14$). We then evaluated the overlap between these fROIs and the ones defined in Exp 2 using a dice-coefficient analysis. Additionally, we extracted Exp 2 responses from ITG-math defined from Exp 3 data (Fig S5).

In addition to these fROIs, we also defined constant size 3 mm-radius disk ROIs and an anatomical ITG ROI:

Disk (cluster) ROIs. To test if results depend on clustering and the spatial scale of the fROIs, we performed a complementary analysis using constant size, 3 mm-radius, disk ROIs (Fig. S4). Disks were automatically placed on the center of mass of ITG-numbers and ITG-math fROIs, respectively. In cases in which the center of mass was outside the fROI, the disk was placed on the center of the largest patch. This analysis was done separately for ITG-numbers defined using either the odd or even runs for the split-half analysis of Exp 1, and separately for ITG-math defined from either the odd or even runs for the split-half analysis of Exp 2.

Anatomical posterior ITG. This ROI was defined bilaterally from anatomical landmarks: anterior boundary: the posterior hippocampus, inferior boundary: the occipito-temporal sulcus (OTS), superior boundary: the middle temporal gyrus (MTG), and posterior boundary: the anterior occipital sulcus (AOS) also referred to as the posterior inferior temporal sulcus (pITS). This anatomical ROI enabled an unbiased analysis of multivoxel patterns of response. An example anatomical ITG ROI of a representative subject is shown in Fig. 7A. Anatomical ITG ROIs of all individual subjects are shown in Fig S1 and S2. The anatomical expanse of this posterior ITG ROI typically encompassed both ITG-numbers and ITG-math.

To determine the specificity of any task/stimulus preference observed, we also evaluated responses in the neighboring mid OTS (mOTS, which likely corresponds to the VWFA-2 (Cohen et al., 2000)) and clusters in the medial aspect of the fusiform gyrus that showed higher

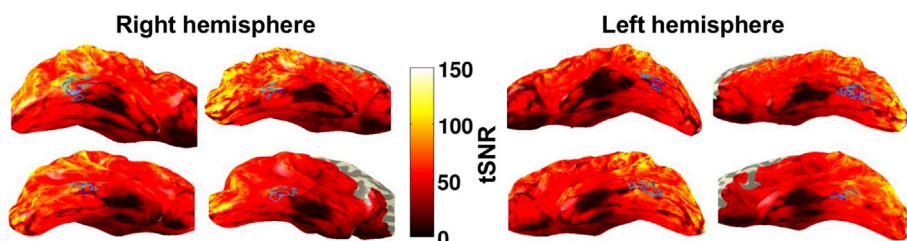


Fig. 2. ITG fROIs are located posterior to signal dropout.

The outline of ITG-math (cyan) and ITG-numbers (blue) is superimposed on four representative inflated brains showing the average temporal signal-to-noise ratio (tSNR) from one MRI session. These data show that the ITG fROIs are adjacent to, but not within, the signal dropout caused by the air/bone interface of the petrous bone.

responses in the color task than the other tasks (likely corresponding to color patches Ac, Cc, and Pc (Lafer-Sousa et al., 2016)). Based on prior research (Dehaene and Cohen, 2011), we expected regions in the mOTS to prefer letters over numbers and reading over adding and color tasks, whereas the color patches were expected to prefer the color task over the reading and the adding tasks. *mOTS-letters* was defined as voxels the left middle occipitotemporal sulcus that showed significantly higher responses for letters vs. other stimuli in Exp 1 (all data: $T = 3$, voxel level, uncorrected, $N = 13$; split-half: $T = 2.5$, voxel level, uncorrected, $N = 12$). An example mOTS-letters in a representative subject is shown in green in Fig. 6A. *mOTS-reading* was defined as voxels in the left middle occipitotemporal sulcus that showed significantly higher responses during the reading task compared to the other tasks in Exp 2 (all data: $T = 3$, voxel level, uncorrected, $N = 14$; split-half: $T = 2.5$, voxel level, uncorrected, $N = 14$). An example mOTS-reading in a representative subject is shown in green in Fig. 6D. The color patches (*Ac-color*, *Cc-color* and *Pc-color*) were defined as voxels in the medial aspect of the fusiform gyrus that showed higher responses during the color task vs. the adding and the reading tasks in Exp 2 (all data: $T = 3$, voxel level, uncorrected; split half: $T = 2.5$, voxel level, uncorrected). *Ac-color* was identified bilaterally in $N = 12$ subjects using all data and in $N = 7$ subjects in the split-half analysis. *Cc-color* was identified bilaterally in $N = 14$ subjects using all data and in $N = 13$ subjects in the split-half analysis. *Pc-color* was found bilaterally in $N = 11$ subjects using all data and in $N = 9$ subjects in the split-half analysis. An example *Cc-color* in a representative subject is shown in pink in Fig. 6D, all color patches in the same subject are shown in Fig S6.

A previous study suggested that the posterior ITG cannot be investigated using fMRI as it falls within the fMRI signal dropout zone near the petrous bone (Shum et al., 2013). To address this concern, we evaluated the temporal signal-to-noise ratio (tSNR; a measure of image quality that is related to the detectability of BOLD signal changes of interest (Murphy et al., 2007)) in the neighborhood of our fROIs. We created tSNR maps in each subject, by dividing the mean functional image of one session by the standard deviation of that image. We then visualized these tSNR maps and our fROIs on the inflated cortical surface to evaluate the spatial relation of the drop-out zone and our fROIs. Fig. 2 shows this visualization for four representative subjects. In all our participants, ITG-numbers and ITG-math were located posterior to the drop-out zone near the ear canal, suggesting that they can be investigated using fMRI.

Statistical analysis

Univariate analysis: We first extracted the average time course in percentage signal change for each condition from the respective fROI and then applied a GLM to estimate betas which indicate the magnitude of response for each condition. Accordingly, the bar graphs in Figs. 3, 4 and 6 show betas in units of % signal change \pm SEM. Crucially, we always used independent data for fROI definition and signal extraction. In cases in which fROIs were defined in the same experiment, the data was split into odd and even runs. We defined the fROIs once using only the odd and once using only the even runs, and extracted the signal from the other, independent, runs. Bar graphs show the average of these split-half analyses.

Analysis of variance: For the univariate data analyses, we used the response amplitudes parameter estimates (betas) from each functionally defined fROI to conduct repeated measures analyses of variance (ANOVAs), in which subject was the repeated measure, to determine which factors significantly modulated responses. **Exp 1:** ANOVAs used hemisphere and stimulus as factors; **Exp 2 and Exp 3:** ANOVAs used hemisphere, stimulus, and task as factors. For the mOTS, only the left hemisphere was evaluated, as fROIs could not be determined reliably in the right hemisphere (in accordance with the previously described left-hemisphere dominance of the VWFA (Dehaene and Cohen, 2011)).

Post-hoc tests testing differences between conditions showing a significant main effect or interactions in the ANOVA were performed using

Fisher's least significant difference (LSD) tests and we report all significant effects. Fisher's LSD was chosen as it allows comparison to results from a previous study (Grotheer et al., 2016b). We used this post-hoc test to assess which conditions contribute to effects found by the more conservative ANOVA. Since Fisher's LSD does not account for multiple comparisons, we additionally report Bonferroni corrected post-hoc results, where applicable.

Dice coefficient (DC): We quantified the spatial overlap between functionally defined ROIs using the dice coefficient $DC = \frac{2|A \cap B|}{|A| + |B|}$, where $|A|$ is one fROI, $|B|$ is the other fROI and $|A \cap B|$ is the intersection between these two fROIs. The dice coefficient (Dice, 1945; Sorensen, 1948) quantifies the similarity of two samples; a DC of 1 indicates complete overlap (i.e. each voxel in fROI A is also found in fROI B), while a DC of 0 indicates that there is no overlap between the fROIs.

Since the dice coefficient can be influenced by the size of the fROI if we consider spatial overlap within a confined anatomical territory, we estimated a chance level for each combination of fROIs and tested if the observed DC differs significantly from chance. To do this, we randomly positioned 2 disk ROIs, matched in size to the fROIs, within a square that matched the size of the combined ITG anatomical and fROIs and measured the DC. This process was repeated 10000 times within each subject, and the average DC was defined as the chance level DC for each combination of fROIs. We then used paired t-tests to evaluate if the measured DC was different from chance. For evaluating the overlap between the ITG and the mOTS fROIs, the size of the square was increased by the size of the mOTS fROIs.

Multivoxel pattern analysis (MVPA): We conducted MVPAs (Haxby et al., 2001) for distributed responses across the anatomical ITG ROI. A GLM was used to estimate the response to each experimental condition separately for each voxel. The responses were then normalized (by subtracting each voxel's mean response) and z-transformed. We calculated the correlation among each pair of multivoxel patterns (MVPs) for each condition in each experiment using a leave-one-run-out procedure, and summarized these correlations in representation similarity matrix (RSMs, Fig. 7). The leave-one-run-out cross-validation procedure allowed us to evaluate the reliability of the observed distributed patterns across runs. The RSMs were calculated independently within each subject and then averaged across subjects.

We used a winner-takes all (WTA) classifier to determine in each subject and experiment, if the stimulus and/or the task can be decoded from MVPs of the posterior ITG. The classifier used a leave-one-run-out procedure, where training data included all runs but one and testing data included the left-out run. Results report the average of all iterations of leave-one-out combinations. Paired t-tests were used to evaluate if encoding performance differed from chance level.

Analysis of behavioral responses: We tested for differences in subjects' performance (measured in % correct) as well as in their reaction times (RTs) across tasks and stimuli in Exp 2 and Exp 3 by conducting repeated measures ANOVAs of behavioral responses with task and stimulus as factors. Post-hoc tests on effects that were found to be significant in the ANOVA were performed using Fisher's LSD. As this post-hoc test does not account for multiple comparisons, we additionally report Bonferroni corrected post-hoc results.

Results

Behavioral performance

Subjects could perform all tasks in all experiments. Their average accuracy (\pm SE) was 73 (4)% in Exp 1 ($N = 14$; one participant's responses are missing due to technical difficulties) and $>85\%$ in Exp 2 and 3. In Exp 2, both accuracy and response times (RTs) differed across the reading, adding, and color tasks (main effect of task: accuracy: $F(2,28) = 10.68$, $p = 0.0004$, $\eta^2 = 0.43$; RTs: $F(2,28) = 25.69$, $p < 0.0001$, $\eta^2 = 0.65$). Accuracy was significantly higher in the reading task (mean (\pm SE):

97 ± 2% correct), relative to the other two tasks (all $p < 0.0009$ in uncorrected post-hoc comparisons, all $p < 0.003$ after Bonferroni correction), while accuracy in the adding (mean (±SE): 88 ± 3% correct) and the color (mean (±SE): 89 ± 2% correct) tasks did not differ ($p = 0.61_{\text{uncorrected}}$). Further, subjects were significantly slower in the color task (mean (±SE): 951 ± 43% ms) relative to the other two tasks (all $p < 0.0001_{\text{uncorrected}}$, all $p < 0.0001_{\text{Bonferroni}}$), while RTs in the adding (mean (±SE): 730 ± 27% ms) and the reading (mean (±SE): 783 ± 36% ms) tasks did not differ significantly ($p = 0.12_{\text{uncorrected}}$). In Exp 3 we found no difference in accuracy between the adding (mean (±SE): 93 ± 2% correct) and the color (mean (±SE): 90 ± 2% correct) tasks ($F(1,14) = 3.65$, $p = 0.08$, $\eta^2 = 0.21$), but RTs were slower for the color (mean (±SE): 1001 ± 37% ms) than the adding (mean (±SE): 776 ± 33% ms) task ($F(1,14) = 73.23$, $p < 0.0001$, $\eta^2 = 0.84$). We observed no other main effects or interactions in either study. Because accuracy and response times were not consistently different for the adding task relative to the other tasks, and we could identify task-selective fROIs not only for adding, but also for the reading and the color tasks (Fig. 6, and Fig S6), it is unlikely that performance differences across tasks drove responses in the ITG and nearby regions.

Does a region in the ITG respond more strongly to number vs. other stimuli across experiments?

We first asked whether we can reliably find in each subject a region in the posterior ITG that prefers Arabic numbers over other visual stimuli. We used similar stimuli (numbers, letters, objects, false letters, false numbers, Fourier phase scrambled letters, and Fourier phase scrambled numbers, Fig. 1-Exp1) as prior studies (Price and Ansari, 2011; Shum et al., 2013; Grotheer et al., 2016b), but analyzed the data using individual subject analyses without spatial smoothing. Critically, in contrast to prior work, we also tested if the preference for numbers remains under different tasks and when comparing to additional stimuli. Thus, the same

15 subjects participated in two additional experiments, in which they (i) viewed morphs of numbers and letters while performing adding, reading, and color tasks (Fig. 1-Exp 2) and (ii) viewed numbers, hands and dice while performing adding and color tasks (Fig. 1-Exp 3).

Consistent with prior results (Grotheer et al., 2016b), a region in the posterior ITG (referred to as ITG-numbers/NFA) showed significantly ($T = 3$, voxel-level, uncorrected) higher responses for numbers vs. letters, objects, false letters, false numbers, scrambled letters, and scrambled numbers (Fig. 3A). We refer to this fROI as ITG-numbers. ITG-numbers is located posterior to the fMRI signal dropout near the ear canal (Fig. 2). We identified it in 13 subjects in the right hemisphere and in 15 subjects in the left hemisphere. That is, it was bilateral in 13 subjects. Independent analyses of responses in ITG-numbers during Exp 1 using a split-half approach, revealed a main effect of stimulus ($F(6,60) = 7.22$, $p < 0.0001$, $\eta^2 = 0.42$) that did not differ across hemispheres (no significant interaction between hemisphere and stimulus, $p > 0.05$). As expected, numbers elicited significantly higher responses than each of the other stimuli (Fig. 3B, $p < 0.006_{\text{uncorrected}}$). These higher responses for numbers however did not survive correction for multiple comparisons (Bonferroni), indicating that this weak preference is not robust. Additionally, ITG-numbers could not be identified reliably across subjects using a more conservative conjunction analysis (Fig S3).

To test if the preference for numbers generalizes to other experiments using different tasks and stimuli, we extracted the response amplitudes of ITG-numbers in Exp 2 and Exp 3 in the same participants. Even though ITG-numbers was defined based on its preference to number stimuli in Exp 1, responses in both Exp 2 and Exp 3 showed no main effect of stimulus (Fig. 3 C,D; Exp 2: $F(1,12) = 1.28$, $p = 0.28$, $\eta^2 = 0.10$; Exp 3: $F(2,24) = 0.36$, $p = 0.70$, $\eta^2 = 0.03$) and no interaction between stimulus and task (Exp 2: $F(2,24) = 1.34$, $p = 0.28$, $\eta^2 = 0.10$; Exp 3: $F(2.24) = 0.17$, $p = 0.85$, $\eta^2 = 0.01$).

Surprisingly, responses of ITG-numbers in Exp 2 and Exp 3 were instead modulated by the task (main effect of task: Exp 2: F

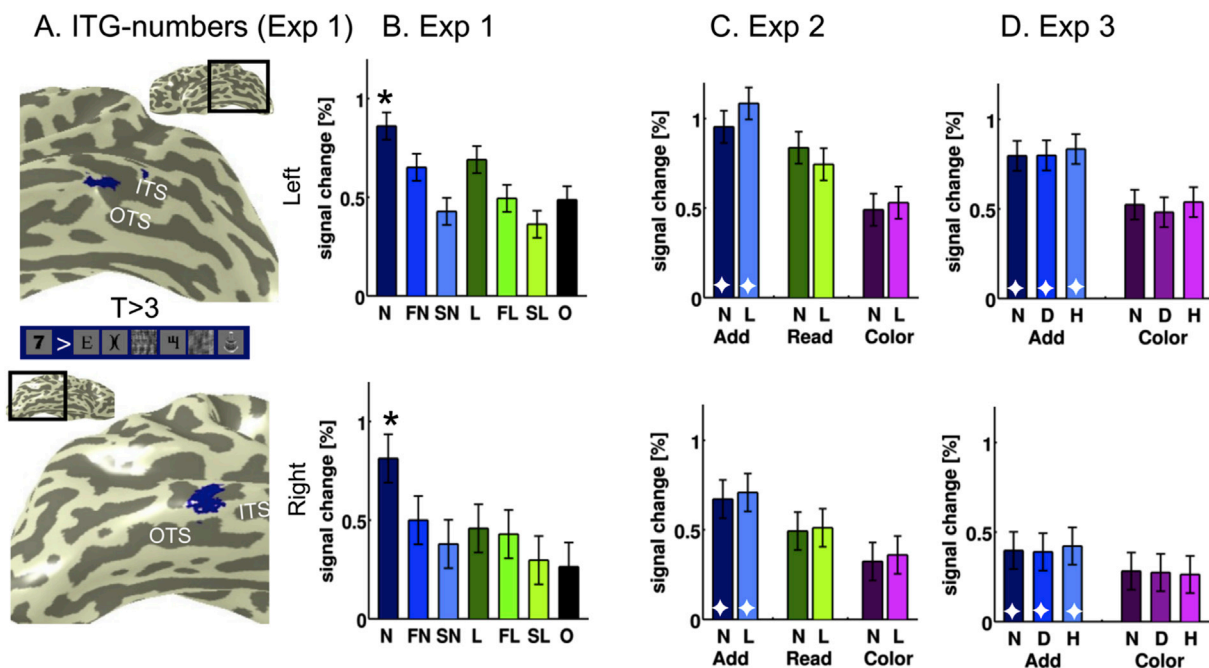


Fig. 3. An ITG region defined based on its preference to Arabic numbers also shows higher responses during a mathematical task compared to other tasks. A. ITG-numbers: voxels in the inferior temporal gyrus (ITG) and inferior temporal sulcus (ITS) that responded significantly more strongly to numbers than the other stimuli in Exp 1 ($T = 3$, voxel level, uncorrected). The figure shows inflated left and right hemispheres from a representative subject. B. Mean responses ± SEM across 11 subjects in ITG-numbers during Exp 1 using a split-half analysis. C. Mean responses ± SEM across 13 subjects from ITG-numbers in Exp 2. D. Mean responses ± SEM across 13 subjects from ITG-numbers in Exp 3. In all panels, *top row*: left hemisphere; *bottom row*: right hemisphere; * significantly higher responses for a stimulus, $p < 0.05_{\text{uncorrected}}$; ♦ significantly higher responses for one task vs. the other tasks, $p < 0.05_{\text{uncorrected}}$. Abbreviations: N = number/number-like morph, L = letter/letter-like morph, FN = false number, FL = false letter, SN = Fourier phase scrambled letter, O = object, D = dice, H = hand, OTS: occipito-temporal sulcus; ITS: inferior-temporal sulcus.

(2,24) = 12.65, $p = 0.0002$, $\eta^2 = 0.51$; Exp 3: $F(1,12) = 9.98$, $p = 0.008$, $\eta^2 = 0.45$). Interestingly, responses in ITG-numbers in both Exp 2 and Exp 3 were higher during the adding task than the other tasks (Fig. 3 C,D; Exp 2: $ps < 0.03_{\text{uncorrected}}$, $ps < 0.07_{\text{Bonferroni}}$; Exp 3: $p = 0.008$) irrespective of the stimuli used in the experiments. That is, not only responses to numbers, but also responses to number-letter morphs, hands, and dice were higher during a mathematical task than during a reading or a color task and responses to the various stimuli were not significantly different from each other within a given task. Results did not differ across hemispheres (no significant interaction between hemisphere and task/stimulus, all $ps > 0.05$) and were replicated in a complimentary analysis in which we used constant-sized disk ROIs centered on fROIs (Supplementary Fig S4 A-C).

While the lack of a stimulus effect in Exp 2 may be due to the visual similarity between the number-letter morphs, the lack of a stimulus effect in Exp 3, which used visually distinct stimuli (numbers, dice, and hands) cannot be explained by visual similarity. Taken together, these results show that, while we can replicate prior findings (Shum et al., 2013; Grotheer et al., 2016b) of a region in the ITG that responds more strongly to numbers vs. letters, objects, false letters, false numbers, scrambled numbers, and scrambled letters using similar stimuli as prior experiments, this preference to numbers is not resilient to changes in stimuli and task.

Does a region in the ITG respond more strongly to the mathematical task vs. other tasks across experiments?

Next, we tested if there is a region in the posterior ITG that is modulated by mathematical processing when stimuli are well controlled. Thus, we created morphs between numbers and letters (Fig. 1-Exp 2) and tested for a region in the ITG that shows stronger responses when subjects performed a mathematical task compared to a reading or a color task on the same stimuli. Then, we tested (i) if preference for mathematical processing remains when using visually distinct stimuli, rather than

number-letter morphs (Fig. 1-Exp 3), and (ii) if the region illustrating preference to mathematical processing in Exp 2, also exhibits preference for number stimuli in Exp 1 and Exp 3.

Analysis of Exp 2 data revealed a region in the posterior ITG that showed significantly ($T = 3$, voxel-level, uncorrected) higher responses during the adding task vs. the reading and the color tasks irrespective of stimulus (Fig. 4A). We refer to this fROI as ITG-math. ITG-math was identified in 13 participants in the right hemisphere and in 15 participants in the left hemisphere and it is located posterior to the fMRI signal dropout zone near the ear canal (Fig. 2). In contrast to ITG-numbers, ITG-math could also have been identified using a more conservative conjunction analysis (Fig S3).

Independent analysis of responses in ITG-math during Exp 2 using a split-half approach, revealed a main effect of task ($F(2,24) = 87.96$, $p < 0.0001$, $\eta^2 = 0.88$), which did not differ across hemispheres (no hemisphere by task or stimulus interactions, $ps > 0.05$). The adding task induced significantly higher responses than the reading and the color tasks (Fig. 4B, all $ps < 0.0001_{\text{uncorrected}}$, $ps < 0.0001_{\text{Bonferroni}}$). In Exp 2, ITG-math also showed a task by stimulus interaction ($F(2,24) = 6.22$, $p = 0.007$, $\eta^2 = 0.34$). That is, not only were responses for number-letter morphs higher when subjects engaged in an adding task compared to other tasks, but during the adding task ITG-math responses were higher ($p = 0.005_{\text{uncorrected}}$, $p = 0.08_{\text{Bonferroni}}$) for letter-like morphs (>80% letters/<20% numbers) than for number-like morphs (>80% numbers/<20% letters). Differences in responses to these stimuli were not significant for the other tasks (reading: $p = 0.18_{\text{uncorrected}}$; color: $p = 0.27_{\text{uncorrected}}$). Crucially, this interaction goes in the opposite direction of what would be expected for a region selective for numbers.

To test if the preference for mathematical processing transcends stimuli, we also evaluated responses of ITG-math in Exp 3, in which the same participants performed adding and color tasks on visually more diverse stimuli than in Exp 2. As in Exp 2, in Exp 3, ITG-math showed a main effect of task ($F(1,12) = 32.13$, $p = 0.0001$, $\eta^2 = 0.73$). Notably, responses were significantly higher in the adding task vs. the color task

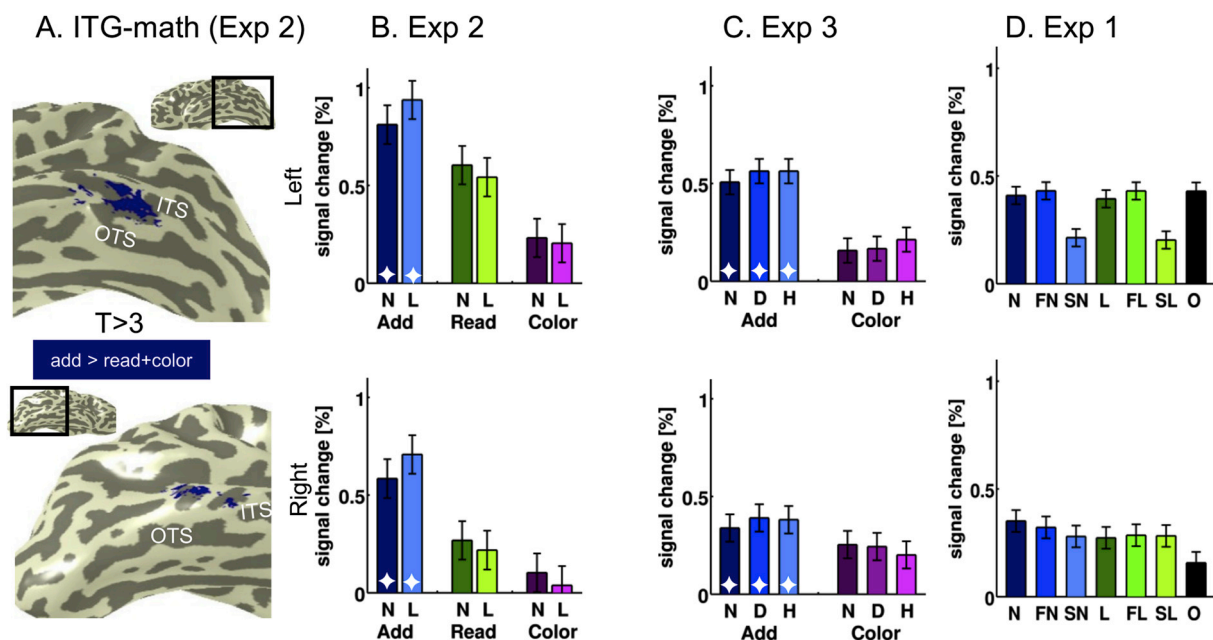


Fig. 4. An ITG region showing preference to a mathematical task does not show preference to Arabic numbers.

A. *ITG-math*: voxels in the ITG/ITS that showed significantly higher responses during adding vs. reading and color tasks in Exp 2 ($T = 3$, voxel level, uncorrected). The figure shows inflated left and right hemispheres from a representative subject. B. Mean responses \pm SEM across 13 subjects in ITG-math during Exp 2 using a split-half analysis. C. Mean responses \pm SEM across 13 subjects from ITG-math in Exp 3. D. Mean responses \pm SEM across 13 subjects from ITG-math in Exp 1. In all panels, *top row*: left hemisphere; *bottom row*: right hemisphere; * significantly higher responses for a stimulus, $p < 0.05_{\text{uncorrected}}$; \blacklozenge significantly higher responses for one task vs. the other tasks, $p < 0.05_{\text{uncorrected}}$; in Exp 2 ITG-math additionally showed a task by stimulus interaction ($F(2,24) = 6.22$, $p = 0.007$, $\eta^2 = 0.34$). Abbreviations: N = number/number-like morph, L = letter/letter-like morph, FN = false number, FL = false letter, SN = Fourier phase scrambled number, SL = Fourier phase scrambled letter, O = object, D = dice, H = hand, OTS: occipito-temporal sulcus; ITS: inferior-temporal sulcus.

(Fig. 4C, $p = 0.0001$), irrespective of whether the subjects viewed numbers, hands, or dice. There was no significant effect of stimulus ($F(2,24) = 0.52$, $p = 0.60$, $\eta^2 = 0.04$), no significant task by stimulus interaction ($F(2,24) = 0.78$, $p = 0.47$, $\eta^2 = 0.06$), and the task effect was stronger in the left than the right hemisphere (task by hemisphere interaction: $F(1,12) = 12.25$, $p = 0.004$, $\eta^2 = 0.51$; task effect in the left hemisphere: $p < 0.0001_{\text{uncorrected}}$, $p < 0.0001_{\text{Bonferroni}}$; task effect in the right hemisphere: $p = 0.01_{\text{uncorrected}}$, $p = 0.07_{\text{Bonferroni}}$). Thus, ITG-math showed higher responses during an adding task than non-mathematical tasks irrespective of the visual stimulus.

We also tested the alternative hypothesis, that ITG-math prefers number stimuli over other visual stimuli. Similar to the results described for ITG-numbers, ITG-math responses were not higher for numbers vs. other stimuli in either Exp 2 or Exp 3 (no main effect of stimulus: Exp 2: $F(1,12) = 0.11$, $p = 0.75$, $\eta^2 = 0.009$; Exp 3: $F(2,24) = 0.52$, $p = 0.60$, $\eta^2 = 0.04$; Fig. 4C), even though Exp 3 used a variety of stimuli that were visually more dissimilar compared to the stimuli used in Exp 2. Moreover, ITG-math did not show a preference for numbers in Exp 1 (no main effect of stimulus: $F(6,72) = 1.71$, $p = 0.13$, $\eta^2 = 0.12$; Fig. 4D). In Exp 1, ITG-math showed a differential pattern of response across hemispheres (significant stimulus by hemisphere interaction: $F(6,72) = 4.23$, $p = 0.001$, $\eta^2 = 0.26$), but numbers did not induce higher responses than all other stimuli in either of the hemispheres. In the right hemisphere, ITG-math responses to numbers were significantly different only from responses to objects ($p = 0.002_{\text{uncorrected}}$) and in the left hemisphere ITG-math responses to numbers were significantly different only from responses to scrambled numbers ($p = 0.002_{\text{uncorrected}}$) and scrambled letters ($p = 0.009_{\text{uncorrected}}$; Fig. 4D). None of these differences survived correction for multiple comparisons, suggesting that the preference for numbers in ITG-math is not robust. Results replicated in a

complimentary analysis in which we used disk ROIs of constant size (Supplementary Fig S4 D-F). Thus, an ITG fROI that showed preference for mathematical processing vs. other tasks did not show a preference for number stimuli over other visual stimuli.

Together these analyses indicate that both ITG-numbers and ITG-math reliably show higher responses for the adding task vs. the other task(s) across experiments. In contrast, neither ITG-numbers or ITG-math showed a consistent preference for numbers across experiments.

Do distinct subregions within the ITG show a preference for numbers or mathematical tasks?

The prior analyses revealed that when the ITG fROIs were defined based on number preferences they also exhibited a preference for mathematical tasks, but when they were defined by task preference, they do not exhibit a stimulus preference to numbers. This raises the possibility that within ITG-math there may be a subpopulation of voxels that is involved in visual processing of numbers, as has been suggested in a recent ECoG study (Daitch et al., 2016). To evaluate this possibility, we tested (i) whether ITG-math and ITG-numbers fROIs spatially overlap and (ii) if ITG-numbers and ITG-math are spatially reliable when defined independently across experiments. These analyses were performed both qualitatively, by visualizing the fROIs on the cortical surface of each participant, and quantitatively, by calculating the dice coefficient (DC) of spatial overlap.

We identified in each participant ITG-numbers separately for Exp 1 (as in the prior analyses) and Exp 3 (Fig. 5A), as voxels that responded more strongly to numbers vs. other stimuli ($T = 3$, voxel level, uncorrected). We identified ITG-math separately in Exp 2 (as in the prior analyses) and Exp 3 (Fig. 5B), as voxels that responded more strongly during the adding task than during the other task(s) ($T = 3$, voxel level,

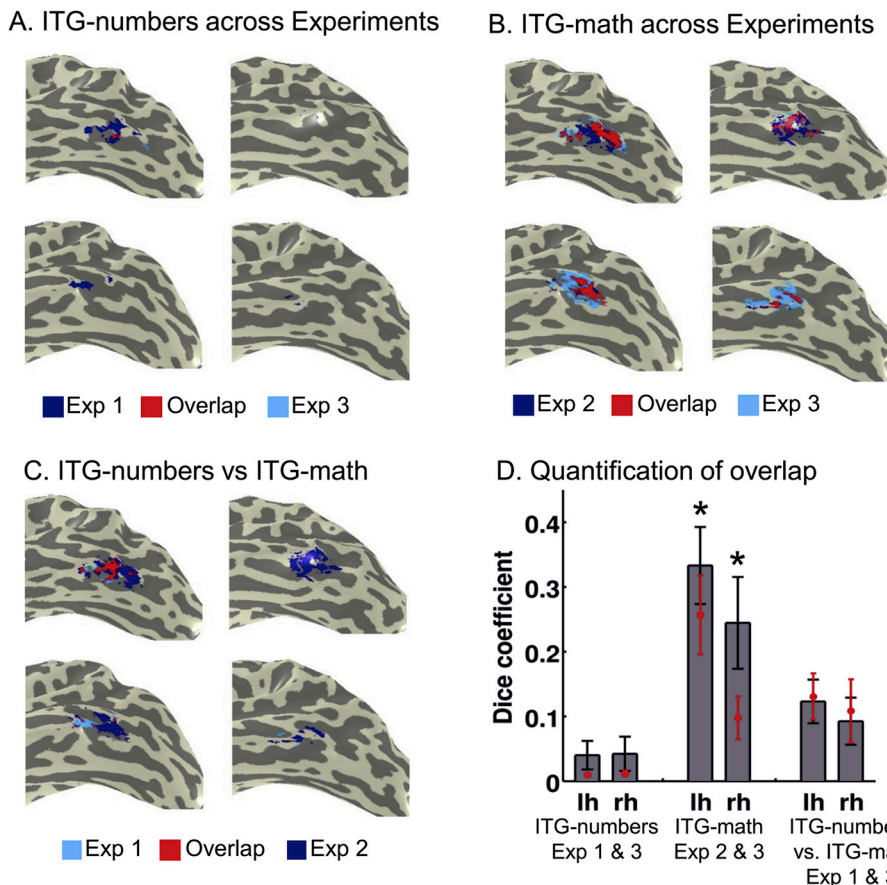


Fig. 5. ITG fROIs are reliable across experiments when defined by task preference, but not when defined based on a preference for numbers.

A-C show the inflated cortical surface of the same 4 participants zoomed on the posterior occipito-temporal cortex. A. Comparison of the location of ITG fROIs based on stimulus preferences to numbers vs. control stimuli ($T = 3$, voxel level, uncorrected) in Exp 1 and Exp 3. B. Comparison of the location of ITG fROIs based on a preference to adding vs. other task(s) ($T = 3$, voxel level, uncorrected) in Exp 2 and 3. C. Comparison of fROIs based on a preference to Arabic numbers in Exp 1 (same as A) with those based on a preference to the mathematical task in Exp 2 (same as B). D. Quantification of the overlap between fROIs across experiments using the dice coefficient (DC), mean across subjects \pm SEM. The DC quantifies the similarity of two samples: a DC of 1 indicates complete overlap between two fROIs, while a DC of 0 indicates no overlap between fROIs. Red dots show chance level based on ROI sizes \pm SEM. * DC is significantly higher than chance, $p < 0.05$.

uncorrected).

Not only were the ITG-numbers and ITG-math fROIs of different sizes, their reliability across experiments also differed. The average size of ITG-numbers across Exp 1 and Exp 3 was $90 \pm 41 \text{ mm}^3$ (average \pm SE) in the right hemisphere and $89 \pm 32 \text{ mm}^3$ in the left hemisphere. In comparison, ITG-math defined in Exp 2 and Exp 3 was more than 3 times larger, extending $364 \pm 83 \text{ mm}^3$ in the right hemisphere and $618 \pm 133 \text{ mm}^3$ in the left hemisphere. Further, ITG-math was reliable across experiments, while ITG-numbers was not. ITG-math defined in Exp 2 and ITG-math defined in Exp 3 overlapped ($N=12$; $DC \pm SE$: right hemisphere: 0.24 ± 0.07 , left hemisphere: 0.33 ± 0.06 ; Fig. 5 B,D), and this overlap was significantly above chance (right hemisphere: $p < 0.009$; left hemisphere: $p < 0.04$; Fig. 5D). In contrast, there was almost no overlap between ITG-numbers across Exp 1 and Exp 3 ($N=8$; $DC \pm SE$: right: 0.04 ± 0.03 , left: 0.04 ± 0.02 ; Fig. 5A,D), and this overlap was not different from chance (right hemisphere: $p = 0.24$; left hemisphere: $p = 0.18$; Fig. 5D).

Examination of the spatial overlap between ITG-numbers and ITG-math across experiments also revealed little overlap, evident in the paucity of red voxels in Fig. 5C. While there was an occasional subject that showed overlap between these fROIs, the average dice coefficient of the overlap between ITG-numbers and ITG-math was less than 0.13 ($DC \pm SE$: ITG-numbers from Exp 1 with ITG-math from Exp 2 ($N=11$): right: 0.09 ± 0.04 , left: 0.12 ± 0.03 ; ITG-numbers from Exp 1 with ITG-math from Exp 3 ($N=13$): right: 0.09 ± 0.04 , left: 0.11 ± 0.03 ; ITG-numbers from Exp 3 with ITG-math from Exp 2 ($N=7$): right: 0.01 ± 0.01 , left: 0.02 ± 0.01). Furthermore, the overlap between ITG-numbers and ITG-math was not different from chance (right hemisphere: $ps > 0.25$, left hemisphere: $ps > 0.33$).

The overlap between ITG-math defined across experiments (i) was larger than the overlap between ITG-numbers defined across experiments, (ii) was larger than the overlap between ITG-math and ITG-numbers across experiments and (iii) was the only overlap that was significantly above chance. Therefore, rather than supporting the

hypothesis that there are distinct neuronal populations within the posterior ITG involved in visual processing of numbers, mathematical processing, or both, our results suggest that the only reliable voxels within the ITG are those involved in mathematical processing but not the visual processing of numbers.

Do regions involved in processing numbers and mathematical tasks overlap with regions involved in processing letters and reading?

A related question is whether the same or different regions are involved in processing visually presented numbers and letters. To address this question, we identified regions that showed higher responses to letters than other stimuli ($T=3$, voxel level, uncorrected) using the data from Exp 1. These regions were anatomically segregated from our ITG fROIs as they were located on a different anatomical structure, namely the occipital temporal sulcus (OTS). We focus on a letter-selective region in the middle occipito-temporal sulcus (mOTS-letters, also referred to as VWFA2) as it was most proximal to the ITG fROIs. mOTS-letters was identified in 13 subjects in the left hemisphere, but only in 8 subjects in the right hemisphere. Thus, in subsequent analyses we only used the data from the left hemisphere (Fig. 6A).

An independent split-half analysis of the left mOTS-letters responses in Exp 1 revealed a main effect of stimulus ($F(6,66) = 28.40$, $p < 0.0001$, $\eta^2 = 0.73$, Fig. 6B), whereas responses were significantly higher for letters than the other stimuli (all $ps < 0.03_{\text{uncorrected}}$). This preference for letters in mOTS-letters did not survive Bonferroni correction for multiple comparisons. Similar to our findings in ITG-numbers, mOTS-letters responses did not differ across stimuli in Exp 2 (no significant main effect of stimulus: $F(1,12) = 0.93$, $p = 0.35$, $\eta^2 = 0.07$), but responses significantly varied across tasks (significant main effect of task: $F(2,24) = 20.33$, $p < 0.0001$, $\eta^2 = 0.63$, Fig. 6C). In contrast to ITG-numbers, in left mOTS-letters reading rather than adding induced significantly higher responses than the other tasks ($ps < 0.02_{\text{uncorrected}}$, $ps < 0.04_{\text{Bonferroni}}$). In Exp 2, mOTS-letters also exhibited a stimulus by

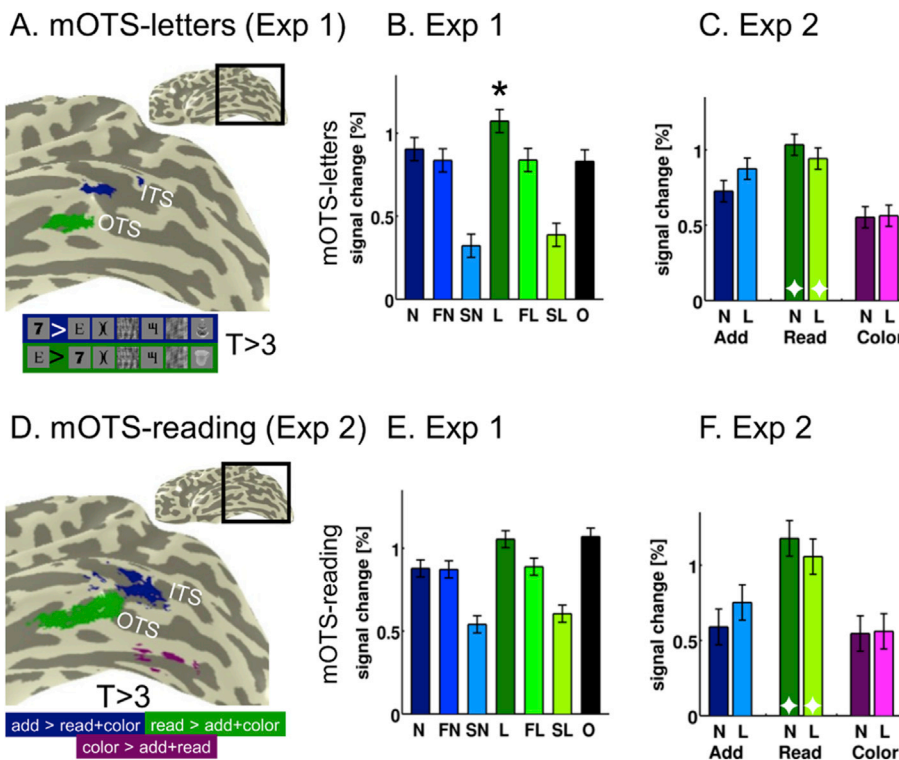


Fig. 6. Response properties of fROIs in the mOTS defined by their preference for letters or reading.

A. mOTS-letters (green) and ITG-numbers (blue) fROIs in a representative inflated left hemisphere. *mOTS-letters*: voxels in the left hemispheric middle portion of the OTS that showed significantly higher responses to letters vs. all other stimuli in Exp 1 ($T=3$, voxel level, uncorrected). B. Mean responses \pm SEM across 12 subjects in mOTS-letters during Exp 1 using a split-half analysis. C. Mean responses \pm SEM across 13 subjects from mOTS-letters in Exp 2. D. mOTS-reading (green), ITG-math (blue), and Cc-color (pink, mean responses in Fig S6) fROIs in a representative inflated left hemisphere. *mOTS-reading*: voxel that responded significantly more strongly to reading vs. adding and color tasks in Exp 2 ($T=3$, voxel level, uncorrected). E. Mean responses \pm SEM across 14 subjects from mOTS-reading in Exp 1. F. Mean responses \pm SEM across 14 subjects in mOTS-reading during Exp 2 using a split-half analysis. * significantly higher responses for a stimulus, $p < 0.05_{\text{uncorrected}}$; \blacklozenge significantly higher responses for one task vs. the other tasks, $p < 0.05_{\text{uncorrected}}$; in Exp 2 mOTS-letters and mOTS-reading additionally showed a task by stimulus interaction ($F(2,24) = 10.39$, $p = 0.0006$, $\eta^2 = 0.46$ and $F(2,26) = 8.29$, $p = 0.002$, $\eta^2 = 0.39$, respectively). Abbreviations: N = number/number-like morph, L = letter/letter-like morph, FN = false number, FL = false letter, SN = Fourier phase scrambled number, SL = Fourier phase scrambled letter, O = object, OTS: occipito-temporal sulcus; ITS: inferior-temporal sulcus.

task interaction ($F(2,24) = 10.39$, $p = 0.0006$, $\eta^2 = 0.46$). Adding letter-like morphs induced higher responses than adding number-like morphs ($p = 0.0006_{\text{uncorrected}}$, $p = 0.009_{\text{Bonferroni}}$), and reading number-like morphs induced higher responses than reading letter-like morphs ($p = 0.02_{\text{uncorrected}}$, n.s. after Bonferroni correction), but there was no difference between the morphs in the color task ($p = 0.79_{\text{uncorrected}}$). Thus, there is a double-dissociation between responses to task and stimulus across ITG-numbers and mOTS-letters, suggesting that these regions are functionally dissociable.

To further test if regions involved in reading are distinct from regions involved in adding, we identified in each subject fROIs that showed higher responses during reading than during the other tasks in Exp 2 (Fig. 6D). In 14 subjects in the left hemisphere, and in 7 subjects in the right hemisphere, a region in the mOTS (mOTS-reading) showed higher responses during reading than adding or remembering colors in Exp 2 (Fig. 6F). An independent split-half analysis of mOTS-reading in Exp 2, showed a main effect of task ($F(2,26) = 40.99$, $p < 0.0001$, $\eta^2 = 0.76$), with higher responses during the reading task relative to the adding ($p < 0.0001_{\text{uncorrected}}$, $p < 0.0001_{\text{Bonferroni}}$) and the color ($p < 0.0001_{\text{uncorrected}}$, $p < 0.0001_{\text{Bonferroni}}$) tasks (Fig. 6F). We also found an interaction between task and stimulus in Exp 2 ($F(2,26) = 8.29$, $p = 0.002$, $\eta^2 = 0.39$), as letter-like morphs induced higher responses in mOTS-reading than number-like morphs in the adding task ($p = 0.003_{\text{uncorrected}}$, $p = 0.04_{\text{Bonferroni}}$) and number-like morphs induced higher responses than letter-like morphs in the reading task ($p = 0.02_{\text{uncorrected}}$; n.s. after Bonferroni correction), while responses did not differ across stimuli in the color task ($p = 0.76_{\text{uncorrected}}$). In contrast to ITG-math, mOTS-reading further showed differential responses to the stimuli in Exp 1 (significant main effect of stimulus: $F(6,78) = 11.50$, $p < 0.0001$, $\eta^2 = 0.47$; Fig. 6E) as letters induced higher responses than the other stimuli ($p \leq 0.05$ for all uncorrected post-hoc comparisons, except for objects $p = 0.86_{\text{uncorrected}}$). This weak preference did not survive

Bonferroni correction for multiple comparisons. Thus, different regions in the ventral visual stream show a preference for reading, adding, and judging the color (example color fROI in Fig. 6, mean responses shown in Fig S6) of visually presented stimuli.

Even though fROIs involved in processing numbers/math or letters/reading are centered on different anatomical structures (ITG and OTS, respectively), these regions are neighboring. Thus, it remains possible that there is spatial overlap between these functionally distinct fROIs. To assess this possibility, we evaluated the overlap between these regions using the dice coefficient. In the left hemisphere, mOTS-letters (size: $201 \pm 67 \text{ mm}^3$; average \pm SE) and ITG-numbers (size: $255 \pm 65 \text{ mm}^3$) defined in Exp 1 were of comparable sizes. However, there was virtually no overlap between these number- and letter-selective fROIs ($N = 13$; $DC \pm SE$: 0.03 ± 0.02), and this negligible overlap was not significantly different from chance ($p = 0.21$). Likewise, in the left hemisphere, there was almost no overlap between ITG-math (size: $540 \pm 108 \text{ mm}^3$) and mOTS-reading (size: $239 \pm 63 \text{ mm}^3$) defined in Exp 2 ($N = 14$; $DC \pm SE$: 0.003 ± 0.002), and this lack of overlap was significantly below chance ($p = 0.002$). Overall, our analyses indicate that regions involved in adding and number processing and regions involved in reading and letter processing are both functionally and spatially distinct.

Do distributed responses across the ITG contain reliable stimulus information, task information or both?

The univariate analyses revealed that the ITG exhibits a reliable preference for a mathematical task, but no reliable preference for visually presented numbers. However, it remains possible that a preferential encoding of Arabic numbers is present in distributed responses across the posterior ITG. To test this hypothesis, we evaluated the information of distributed response patterns within an anatomically-defined, independent, posterior ITG ROI (Fig. 7A). The average size (\pm SE) of this

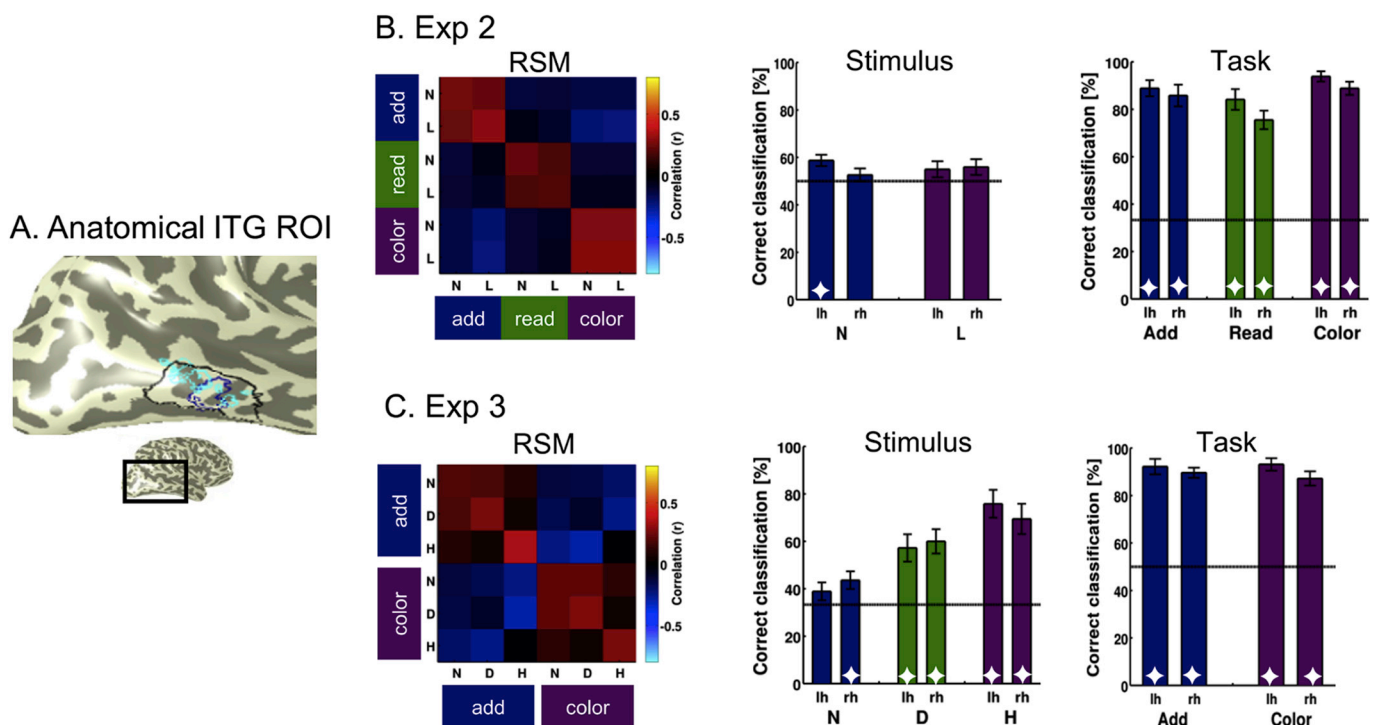


Fig. 7. Multivoxel pattern analyses show that a mathematical task can be more consistently read out than a number stimulus from the posterior ITG. A. Example anatomical posterior ITG ROI (in black) used in these analyses, presented together with ITG-math (cyan) and ITG-numbers (blue) in a representative subject. B. *Left*: Mean representational similarity matrix (RSM) for Exp 2 data, across both hemispheres and all subjects. Conditions are arranged by stimulus and grouped by task. *Right*: Mean \pm SEM winner-takes-all (WTA) classification performance in Exp 2 for stimulus and task. C. Same as B for Exp 3. In B, C: *Dotted line*: chance classification level; \blacklozenge $p < 0.05$ above chance classification. *Abbreviations*: N = number/number-like morph, L = letter/letter-like morph, FN = false number, FL = false letter, SN = Fourier phase scrambled number, SL = Fourier phase scrambled letter, O = object, D = dice, H = hand.

anatomical ITG ROI was $2825 (\pm 279) \text{ mm}^3$ in the right hemisphere and $2994 (\pm 297) \text{ mm}^3$ in the left hemisphere. We first calculated representational similarity matrices (RSMs) indicating the correlation among multivoxel patterns (MVPs) of distributed responses across runs and conditions in each experiment. Then we used a winner-takes-all (WTA) classifier, separately for stimulus and task, to determine which type of information can be reliably decoded from distributed ITG responses. We focused our analyses on Exp 2 and 3, as Exp 1 only contained one type of task (see Fig S7 for Exp 1 MVPA results).

In all experiments, across runs, right and left hemisphere RSMs were strongly correlated even when removing the diagonal of the RSM that contains the within-category correlations (Exp 2: $r = 0.81$; Exp 3: $r = 0.77$). Therefore, we present these RSMs averaged across hemispheres. Importantly, across experiments, the distributed patterns for a given stimulus and task was reproducible across runs, as illustrated in the significant and positive on-diagonal correlations (Exp 2: $r = 0.26 \pm 0.02$; Exp 3: $r = 0.26 \pm 0.02$, Fig. 7B-left and C-left, $p_s < 0.0001$). This indicates that our experiments generated reproducible MVPs across an anatomical ITG expanse.

In Exp 2 and Exp 3, in which both stimulus and task varied, a task effect would manifest as positively correlated MVPs to different stimuli under the same task, which would be significantly higher than the correlations to the same stimuli under different tasks. In contrast, a stimulus effect would manifest as positively correlated MVPs for different tasks performed on the same stimulus, which would be significantly higher than the correlations for the same task performed on different stimuli. Additionally, a classifier can be used to determine if task, stimulus, or both can be read out from distributed ITG responses. Examination of the RSMs of Exp 2 and Exp 3 revealed a clear task effect: MVPs for different stimuli under the same task were positively correlated (mean \pm SE: Exp 2: 0.23 ± 0.02 , Exp 3: 0.12 ± 0.02), but MVPs for the same stimulus across tasks were negatively correlated (mean \pm SE: Exp 2: -0.13 ± 0.03 , Exp 3: -0.08 ± 0.02 ; Fig. 7B-left, C-left). Further, within task correlations across different stimuli were significantly higher than within stimulus correlations across tasks (all $p_s < 0.004$).

Accordingly, we observed better decoding accuracy for task relative to stimulus. In Exp 2, the three tasks were classified with an average accuracy (\pm SE) of $86\% \pm 4\%$ (Fig. 7B-right), which was significantly above the 33.3% chance level (t -test against chance: $p < 0.0001$ for the average classification as well as the classification for each task individually). The classifier could successfully determine if subjects were adding with $87\% \pm 4\%$ accuracy. In contrast, decoding whether the stimulus was a number or a letter had an average accuracy of only $56\% \pm 3\%$ (t -test against 50% chance: $p = 0.01$ for the average classification; above chance classification of numbers was only significant for the left ITG: $59\% \pm 2\%$, t -test against chance: $p = 0.003$, Fig. 7B-middle).

A similar pattern of results was observed in Exp 3, in which the different tasks were decoded from MVPs with an average accuracy of $91\% \pm 3\%$ (t -test against 50% chance: $p < 0.0001$, Fig. 7C-right), and the classifier could determine if participants were adding with $91 \pm 3\%$ accuracy. In contrast, stimulus was decoded with an average accuracy of only $58\% \pm 5\%$ (t -test against 33.3% chance: $p < 0.0001$; Fig. 7C-middle). Of the three stimulus categories used in Exp 3, hands rather than numbers were most accurately decoded from MVPs across posterior ITG. Decoding of numbers was significantly above chance only in the right hemisphere (accuracy: $44\% \pm 4\%$; t -test against 33.3% chance: $p = 0.01$).

These results suggest that an independent classifier can reliably decode from distributed responses in the posterior ITG if participants are involved in a mathematical task across a variety of stimuli (number-letter morphs, numbers, hands, and dice) but it cannot reliably decode if participants are viewing numbers vs. other stimuli under different tasks. Taken together, the MVPAs complement the univariate analyses, and lead to the same conclusion: the influence of task outweighs the influence of the visual stimulus in driving responses across the posterior ITG.

Discussion

In the current study, we addressed a central debate regarding the function of the posterior inferior temporal gyrus (ITG): is it involved in the visual processing of Arabic numbers or stimulus-independent mathematical processing? Our data provides strong evidence supporting the second view for three reasons: (i) Responses in ITG fROIs were higher during mathematical processing than other tasks irrespective of whether subjects viewed numbers or other visual stimuli (such as letter-morphs, hands, or dice). (ii) This preference for mathematical tasks was robust and reliable across experiments. In contrast, while responses in ITG-numbers were higher for numbers relative to other visual stimuli in Exp 1, this preference did not generalize across experiments that used different control stimuli and tasks. (iii) Task, but not number stimulus, was successfully and consistently decoded from distributed responses in the posterior ITG. Our data therefore suggest that neural populations within the posterior ITG are involved in mathematical processing of various stimuli, rather than being specifically involved in the visual processing of Arabic numbers.

Neural populations within the posterior ITG are involved in mathematical processing not task-independent visual processing of numbers

The current study shows that, even though the posterior ITG is activated by visual stimuli, it does not show reliable selectivity for Arabic numbers. While numbers induced higher responses than other stimuli in Exp 1 in a small region within the ITG (ITG-numbers), this preference was not significant after Bonferroni correction for multiple comparisons. Furthermore, we did not observe significantly higher responses to number stimuli compared to other stimuli in ITG-numbers in either Exp 2 or Exp 3. One could argue that the lack of a preference for numbers in Exp 2 is not surprising, given the visual similarity between the number and letter morphs. However, this argument does not hold for Exp 3, which used visually dissimilar stimuli. Together, the lack of a preference for numbers in ITG-numbers in Exp 1 following correction for multiple comparisons and the lack of higher responses for numbers than other stimuli in Exp 2 and Exp 3, suggest that any preference to visual numbers in the ITG is not robust.

Another interesting aspect of our results is that in Exp 2, adding letter-morphs induced higher responses in ITG-math than adding number-morphs. This effect, which goes in the opposite direction of what would be expected of an area selective for numbers, provides further compelling evidence that mathematical processing outweighs the stimulus preference to numbers in the ITG. Given that participants' behavior did not show a task by stimulus interaction in either accuracy or reaction times, it is likely that task difficulty did not drive this differential response. Nonetheless, it is still possible that this pattern of results is driven by the novelty of using letter-like stimuli in a mathematical task. These possibilities can be examined in future experiments that systematically vary task difficulty and stimulus novelty.

A recent meta-analysis on the location of the NFA across studies, suggests that controlling task demands across conditions is important (Yeo et al., 2017). Thus, an open question is whether the higher responses in ITG-math and ITG-numbers during mathematical processing are due to increased attentional load during adding than the other tasks. However, our data suggest that attention cannot explain this task preference for two reasons. First, there was no clear relationship between the participants' performance and meanresponses in either ITG-math or ITG-numbers. For example, despite the fact that performance did not differ between the adding and the color tasks in Exps 2 and 3, we found differential responses during color and adding tasks across ventro-temporal cortex: ITG-math and ITG-numbers showed higher responses during the adding than the color task, while color patches in the medial fusiform gyrus showed higher responses during the color task than the adding task. Likewise, there was no relationship between performance and distributed ITG responses. Participants' performance was

higher in the reading than in the adding and the color tasks in Exp 2, yet MVPA showed similar decoding accuracy across all three tasks. Second, enhanced responses during adding than other tasks were not pervasive across the brain, as predicted by a general attentional account. A complementary whole brain analysis examining which regions showed higher responses during adding than reading and color tasks, did not reveal a global enhancement of responses, but rather enhancement only in a sparsely-distributed set of regions (Fig S8). In addition to ITG-math, this included punctate activations in the intraparietal sulcus (IPS, activated bilaterally in all subjects) and the lateral prefrontal cortex (activated bilaterally in 14 subjects). These results are consistent with prior findings showing the involvement of IPS (Piazza et al., 2004; Eger et al., 2009; Harvey et al., 2013) and prefrontal regions (Harvey and Dumoulin, 2017) in numerosity processing.

Our findings are consistent with prior results that showed (i) activations in the right ITG during mathematical calculations using both Arabic numbers (e.g. $5 + 7$) and number words ('five plus seven') (Hermes et al., 2015) and (ii) higher ITG activations in expert mathematicians when they evaluated mathematical statements compared to non-mathematical statements (Amalric and Dehaene, 2016). An interesting direction for future research would be to directly compare response in ITG-math during different "types" of mathematical tasks, e.g. comparing responses during a passive viewing task, a numerosity judgment task, simple math task (e.g. addition), and complex mathematical calculations (e.g. division). Such studies would expand our understanding of the role of ITG-math in mathematical cognition.

Critically, the methodological advances of our study can resolve the discrepancy between prior reports of heterogeneous compared to homogeneous properties of the posterior ITG. A previous ECoG study reported spatial heterogeneity across the posterior ITG (Daitch et al., 2016) and suggested that number selectivity (or the NFA) constitutes a subregion of a larger ITG region involved in mathematical processing. In contrast, a prior fMRI study using group analyses reported a single region within the ITG that is engaged in both mathematical processing and number processing (Amalric and Dehaene, 2016). Notably, the current study goes beyond this prior work in three critical ways: (1) we are the first to manipulate the participants' task and the visual stimulus independently within the same experiment, (2) different than prior studies, we tested the robustness of results across experiments, and (3) we resolve methodological limitations in prior studies. On the one hand, the prior ECoG study reported data from electrodes that were placed sparsely and in varied locations across participants, which makes it hard to assess spatial heterogeneity. On the other hand, the prior fMRI study used group analyses and spatial smoothing which may have generated spurious overlap among number and math selective activations (Weiner and Grill-Spector, 2013). Thus, our approach of measuring task and stimulus effects across the ITG with fMRI within each participant and without spatial smoothing provides a more accurate and comprehensive method to determine task and stimulus related responses compared to previous studies.

At first glance, our data may seem consistent with the spatial heterogeneity reported by Daitch et al. (2016), as we found both a region in the ITG that shows a preference for numbers (ITG-numbers) and a region in the ITG that shows a preference for mathematical tasks (ITG-math). However, an innovative aspect of our study is that we tested if these regions are reliable across experiments. That is, we explicitly tested if they are spatially consistent across experiments more than what would be expected by chance. This analysis revealed that the spatial reproducibility of ITG-numbers was not different from chance, while the spatial reproducibility of ITG-math was significantly above chance. Therefore, our results do not support the spatial heterogeneity hypothesis, as we could not find a region that reliably prefers numbers across experiments.

In summary, our data show that neural populations within the posterior ITG are involved in mathematical processing of various stimuli, not the visual processing of Arabic numbers.

Difficulties in localizing ITG-numbers are not due to fMRI signal dropout

Previous studies that aimed to identify the ITG based on a preference for numbers over other visual stimuli (i.e. ITG-numbers/NFA) found mixed results; while some studies identified ITG-numbers, many studies failed to find it (see recent meta-analysis by Yeo et al., 2017). Our data shows that there is no reliable preference for numbers in the ITG, which leads to the question why some studies report such a preference (e.g. Shum et al., 2013; Hermes et al., 2015; Daitch et al., 2016; Grotheer et al., 2016b) and why we find such a preference in Exp 1, but not in Exps 2 and 3.

One possibility is that the preference for numbers compared to other stimuli observed in some prior studies and in Exp 1, is related to the fact that numbers, but not the control stimuli in these experiments, contain numerical information. In the context of some tasks, subjects may extract this numerical information and therefore show increased responses to numbers vs. control stimuli in this region. While this interpretation is speculative, it would explain why in the present data the preference for numbers vs. other stimuli disappears (i) when subjects actively engage in a mathematical task, i.e. extract numerical content from all visual stimuli, and (ii) when participants actively extract other visual information from the stimulus (e.g. reading the characters or remembering the color of the stimuli).

Another possibility, is that ITG-numbers is difficult to detect with fMRI due to signal dropout caused by the air/bone interface of the petrous bone near the ITG (Shum et al., 2013). However, close inspection of our ITG fROIs show that they are located posterior to the dropout zone (Fig. 2). Furthermore, in Exp 1, which used similar stimuli and the same task as a prior high resolution fMRI study (Grotheer et al., 2016b), we could identify ITG-numbers (i.e. an NFA) in each of our individual participants with standard fMRI. In contrast, we could not reliably find such a region when we changed the stimuli and the task in Exp 2 and 3. Consistent with our data, recent ECoG measurements (Daitch et al., 2016), which are not affected by the air/bone interface, identified electrodes in the ITG that show a preference for numbers vs. characters only in 4 out of 13 patients, but identified electrodes in the ITG that showed higher responses during a math task than a memory task in all 13 patients. Together with the current study, this suggests that it is not poor fMRI signal quality in the ITG that led to previous failures to find ITG-numbers, but rather that the preference to numbers vs. other visual stimuli is not reliable across experimental scenarios. Thus, our findings suggest that contrasting responses during a mathematical task vs. other tasks rather than contrasting numbers with other visual stimuli will result in a more robust and reliable definition of fROIs involved in numerical processing in the ITG.

The ITG is part of the math processing network and should be considered in any study on mathematical cognition

The triple-code model for numerical cognition (Dehaene, 1992) proposes that there are three representations of numbers in the brain: (i) a visual number form, in which numbers are represented in a string of Arabic numerals in an internal visual spatial scratchpad (thought to be in occipito-temporal cortex, see for instance Shum et al., 2013; Grotheer et al., 2016b), (ii) a verbal word frame in which numbers are represented as a syntactic sequence of words (thought to be associated with language areas), and (iii) an analog magnitude representation which represents the quantity associated with the visual or verbal representation (thought to be in the parietal cortex, see for instance Piazza et al., 2004; Eger et al., 2009; Harvey et al., 2013). Our data suggests that (i) the ITG is likely the closest match for the proposed occipito-temporal region involved in extracting the numerical content of the visual input and (ii) provides important evidence that this ITG region is distinct from regions processing visual word forms (Cohen et al., 2000, 2002; Ben-Shachar et al., 2007; Dehaene and Cohen, 2011) found along the OTS (Fig. 6).

However, our data also suggest that the initial model of a specific

representation of Arabic numerals is too restrictive. The stimulus-independent preference for mathematical tasks in ITG fROIs defined by either a preference for numbers in Exp 1 or a preference for a mathematical task in Exps 2 and 3, observed in the current study, suggests that this region is involved in the extraction of numerical information from a variety of visual stimuli, not only Arabic numbers. This interpretation is consistent with recent findings of a topographically organized numerosity map in the temporo-occipital junction (Harvey and Dumoulin, 2017), which is anatomical proximal to the regions investigated here. Thus, we propose that the posterior ITG is recruited during mathematical processing and its goal may be to associate a numerical value to the current input, irrespective of the shape, format (e.g. pictorial vs. symbol) and possibly even modality (see Abboud et al., 2015; Amalric and Dehaene, 2016) of the stimulus. Such flexibility regarding the input would be advantageous considering that in our daily life mathematical problems are not always presented in the form of Arabic numbers.

In conclusion, our data shows that the posterior ITG supports mathematical processing irrespectively of the nature of the input. We suggest that it is recruited to associate numerical content to the current stimulus.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

This research was supported by the National Institutes of Health (NIH; 1R01EY02391501A1), the Deutsche Forschungsgemeinschaft (DFG; GR 4850/1-1) and an Innovation Grant from the Stanford Center for Cognitive and Neurobiological Imaging (CNI).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.neuroimage.2018.03.064>.

References

- Abboud, S., Maidenbaum, S., Dehaene, S., Amedi, A., 2015. A number-form area in the blind. *Nat. Commun.* 6, 6026.
- Amalric, M., Dehaene, S., 2016. Origins of the brain networks for advanced mathematics in expert mathematicians. *Proc. Natl. Acad. Sci.* 113, 4909–4917.
- Ben-Shachar, M., Dougherty, R.F., Deutsch, G.K., Wandell, B.A., 2007. Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cereb. Cortex* 17, 1604–1611.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., Michel, F., 2000. The visual word form area. *Brain* 123, 291–307.
- Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., Dehaene, S., 2002. Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain* 125, 1054–1069.
- Daitch, A.L., Foster, B.L., Schrouff, J., Rangarajan, V., Kaşikçi, I., Gattas, S., Parvizi, J., 2016. Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. *Proc. Natl. Acad. Sci.* 113, E7277–E7286.
- Dehaene, S., 1992. Varieties of numerical abilities. *Cognition* 44, 1–42.
- Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262.
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.
- Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., Kleinschmidt, A., 2009. Deciphering cortical number coding from human brain activity patterns. *Curr. Biol.* 19, 1608–1615.
- Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–548.
- Grotheer, M., Ambrus, G.G., Kovács, G., 2016a. Causal evidence of the involvement of the number form area in the visual detection of numbers and letters. *Neuroimage* 132, 314–319.
- Grotheer, M., Herrmann, K.-H., Kovacs, G., 2016b. Neuroimaging evidence of a bilateral representation for visually presented numbers. *J. Neurosci.* 36, 88–97.
- Grotheer, M., Kovacs, G., 2014. Repetition probability effects depend on prior experiences. *J. Neurosci.* 34, 6640–6646.
- Harvey, B.M., Dumoulin, S.O., 2017. A network of topographic numerosity maps in human association cortex. *Nat. Hum. Behav.* 1, 36.
- Harvey, B.M., Klein, B.P., Petridou, N., Dumoulin, S.O., 2013. Topographic representation of numerosity in the human parietal cortex. *Science* 341 (80), 1123–1126.
- Hermes, D., Rangarajan, V., Foster, B.L., King, J.-R., Kasikci, I., Miller, K.J., Parvizi, J., 2015. Electrophysiological responses in the ventral temporal cortex during reading of numerals and calculation. *Cereb. Cortex* 27, bhv250.
- Lafer-Sousa, R., Conway, B.R., Kanwisher, N.G., 2016. Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *J. Neurosci.* 36, 1682–1697.
- Libertus, M.E., Brannon, E.M., Pelphrey, K.A., 2009. Developmental changes in category-specific brain responses to numbers and letters in a working memory task. *Neuroimage* 44, 1404–1414.
- Murphy, K., Bodurka, J., Bandettini, P.A., 2007. How long to scan? The relationship between fMRI temporal signal to noise ratio and necessary scan duration. *Neuroimage* 34, 565–574.
- Park, J., Hebrank, A., Polk, T.A., Park, D.C., 2012. Neural dissociation of number from letter recognition and its relationship to parietal numerical processing. *J. Cogn. Neurosci.* 24, 39–50.
- Peters, L., De Smedt, B., Op de Beeck, H.P., 2015. The neural representation of Arabic digits in visual cortex. *Front. Hum. Neurosci.* 9, 517.
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., Dehaene, S., 2004. Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555.
- Price, G.R., Ansari, D., 2011. Symbol processing in the left angular gyrus: evidence from passive perception of digits. *Neuroimage* 57, 1205–1211.
- Shum, J., Hermes, D., Foster, B.L., Dastjerdi, M., Rangarajan, V., Winawer, J., Miller, K.J., Parvizi, J., 2013. A brain area for visual numerals. *J. Neurosci.* 33, 6709–6715.
- Sorensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Det kong danske vidensk. Selesk Biol. Skr.* 5, 1–34.
- Stigliani, A., Weiner, K.S., Grill-Spector, K., 2015. Temporal processing capacity in high-level visual cortex is domain specific. *J. Neurosci.* 35, 12412–12424.
- Weiner, K.S., Grill-Spector, K., 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* 52, 1559–1573.
- Weiner, K.S., Grill-Spector, K., 2011. Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* 56, 2183–2199.
- Weiner, K.S., Grill-Spector, K., 2013. Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychol. Res.* 77, 74–97.
- Yeo, D.J., Wilkey, E.D., Price, G.R., 2017. The search for the number form area: a functional neuroimaging meta-analysis. *Neurosci. Biobehav. Rev.* 78, 145–160.