

## Intuitive physical reasoning is not mediated by linguistic nor exclusively domain-general abstract representations

Hope H. Kean <sup>a,b,\*</sup>, Alexander Fung <sup>a,b</sup>, R.T. Pramod <sup>a,b</sup>, Jessica Chomik-Morales <sup>a,b</sup>, Nancy Kanwisher <sup>a,b</sup>, Evelina Fedorenko <sup>a,b</sup>

<sup>a</sup> Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, 02139, United States

<sup>b</sup> McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, 02139, United States

### A B S T R A C T

The ability to reason about the physical world is a critical tool in the human cognitive toolbox, but the nature of the representations that mediate physical reasoning remains debated. Here, we use fMRI to illuminate this question by investigating the relationship between the physical-reasoning system and two well-characterized systems: a) the domain-general Multiple Demand (MD) system, which supports abstract reasoning, including mathematical and logical reasoning, and b) the language system, which supports linguistic computations and has been hypothesized to mediate some forms of thought. We replicate prior findings of a network of frontal and parietal areas that are robustly engaged by physical reasoning and identify an additional physical-reasoning area in the left frontal cortex, which also houses components of the MD and language systems. Critically, direct comparisons with tasks that target the MD and the language systems reveal that the physical-reasoning system overlaps with the MD system, but is dissociable from it in fine-grained activation patterns, which replicates prior work. Moreover, the physical-reasoning system does not overlap with the language system. These results suggest that physical reasoning does not rely on linguistic representations, nor exclusively on the domain-general abstract reasoning that the MD system supports.

### 1. Introduction

The ability to reason about the physical world is essential to our everyday lives. According to one proposal, physical reasoning relies on a generative probabilistic model of physical causes and effects, similar to a video game's physics engine (Eberly, 2010; Battaglia et al., 2013). This "Intuitive Physics Engine" model includes a dictionary of representational primitives (e.g., objects, surfaces), knowledge of their physical properties, and of the constraints on object movement and inter-object or object-to-surface interactions, allowing for the representation of diverse physical events, including making predictions about how the world will change over time (Smith and Vul, 2013).

Where in the brain is this Intuitive Physics Engine implemented? Fischer et al. (2016) contrasted brain responses while participants made intuitive physical judgments versus performed a difficulty-matched color-judgment task on the same stimuli (Fig. 2B) and identified a set of frontal and parietal areas that respond more strongly during the physical-reasoning condition. In line with these areas' role in physical reasoning, subsequent work has established that they represent physical mass (Schwettmann et al., 2019) and stability (Pramod et al., 2022), and support forward prediction of physical events (Pramod et al., 2022;

Fischer and Mahon, 2021).

What kind of representations mediate physical reasoning? One possibility is that physical reasoning draws on abstract domain-general representations that support other kinds of reasoning, like mathematical and logical reasoning. Indeed, the topography of the physical-reasoning system bears some resemblance to the domain-general Multiple Demand (MD) system (Duncan, 2010; Fedorenko et al., 2013; Duncan et al., 2020), which is implicated in diverse goal-directed behaviors and supports several kinds of formal reasoning (Duncan and Owen, 2000; Fox et al., 2005; Hampshire et al., 2011; Niendam et al., 2012; Fedorenko et al., 2013; Hearne et al., 2017; Amalric and Dehaene, 2019; Woolgar et al., 2018; Assem et al., 2020a,b; Ivanova et al., 2020; Liu et al., 2020). In line with this possibility, past work has established that the representations in the physical-reasoning system are quite *abstract*: for example, representations of physical stability are invariant to the animacy of the object/entity (Pramod et al., 2022), representations of object mass are invariant to diverse aspects of the scene (Schwettmann et al., 2019), and representations of whether or not two objects make a contact are invariant to object and contact types (Pramod et al.). However, Fischer et al. (2016) found that the physical-reasoning areas are at least partially dissociable from the MD system. In line with

This article is part of a special issue entitled: Honouring John Duncan published in Neuropsychologia.

\* Corresponding author. Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, 02139, United States.

E-mail address: [hopekean@mit.edu](mailto:hopekean@mit.edu) (H.H. Kean).

this neural dissociation, physical reasoning appears to be cognitively separable from spatial cognition and general fluid intelligence, as revealed in a recent individual-differences behavioral investigation (Mitko et al., 2024).

Another possibility is that physical reasoning draws on linguistic representations. Formal approaches to physical systems (such as the Physics Engine approach introduced above; Eberly, 2010) provide structured “languages” for representing dynamic, physical world states (for a related approach—situation calculus—see McCarthy and Hayes, 1969; Kowalski and Sergot, 1986; Pinto and Reiter, 1993). Such approaches emphasize the importance of rule-based changes to represent the dynamic physical world. Linguistic structures, including tense and aspect systems, robustly encode events and state changes (Partee, 1973, 1984; Moens, 1987; Moens and Steedman, 1988; Pulman, 1997; Grønn and Von Stechow, 2016) and appear capable of subtly affecting the encoding of the physical and social structure of visual events (Tversky and Kahneman, 1981; Gleitman, 1990; Skordos et al., 2020; Vurgun et al., 2022, 2024). These properties make linguistic structures well-suited as symbolic, rule-like representations of physical world dynamics (Wong, Grand et al., 2023). The sufficiency of language for representing at least some aspects of the physical world is further evidenced by the ability of large language models to develop an internal representation of the physical world from text alone—including spatial relations, object interactions, and causal structure (Li et al., 2023; Nanda et al., 2023; Gurnee and Tegmark, 2024; Marks and Tegmark, 2024). Finally, the developmental trajectory of physical reasoning abilities provides another potential point of support for the role of language. In particular, although some physical reasoning abilities are already present in infancy (Leslie and Keeble, 1987; Baillargeon et al., 1985, 1992; Spelke et al., 1992; Spelke, 2022), understanding of certain physical concepts, such as solidity and support, exhibit a dip in performance during toddlerhood (e.g., Berthier et al., 2000; Hood et al., 2000, 2003). The causes of this dip are debated (e.g., Keen, 2003; Xu, 2019), but one possibility is that the development of linguistic abilities is transforming the early-emerging (‘core’; Spelke, 2022) physical reasoning abilities (mediated by visual-perceptual representations) into more abstract and structured ones based on the linguistic encoding of information, and this process leads to temporary difficulties. On the other hand, at least some evidence exists suggesting that advanced physical reasoning can take place largely independent of language: Varley (2002) reports two individuals with severe aphasia who were nevertheless able to identify the physical component causing a Rube Goldberg machine to fail—a task that requires sophisticated physical and functional understanding.

To shed further light on the representational format of intuitive physical reasoning, we first replicate Fischer et al.’s (2016) findings using a larger participant sample and identify an additional component of the physical-reasoning system in the left frontal cortex. The location of this new area provides additional motivation for examining overlap with the MD and language systems, both of which have left frontal components. In the critical analyses, we find that the physical-reasoning system overlaps with the MD system but is dissociable from it in the response profiles and fine-grained activation patterns, and it shows no overlap with the language system. Thus, physical reasoning—at least the type of intuitive physical reasoning examined here—does not recruit linguistic nor fully abstract domain-general representations, and plausibly relies on domain-specific knowledge structures.

## 2. Methods

### 2.1. Participants

Forty participants were recruited from MIT and the surrounding community. All participants were native speakers of English, had normal hearing and vision, and no history of language impairment. All but one participant were right-handed; the left-handed participant had a left-lateralized language system (as determined by the language localizer

task described below), and was therefore included in all analyses (Willems et al., 2014). One (right-handed) participant had a right-lateralized language system and was excluded from the analysis of the language system’s responses to physical reasoning, leaving 39 participants for that analysis. All participants provided written informed consent in accordance with the requirements of MIT’s Committee on the Use of Humans as Experimental Subjects (COUHES) and were paid for their time.

### 2.2. Design

All participants completed the intuitive physics localizer task (Fischer et al., 2016) and a language localizer task (Fedorenko et al., 2010). Twenty-nine of the participants additionally completed a spatial working memory task (from Fedorenko et al., 2011), which is commonly used to localize the domain-general Multiple Demand system.

**The intuitive physics localizer.** This localizer, introduced in Fischer et al. (2016), included two conditions in a blocked design. Participants viewed videos of unstable block towers made up of yellow and blue blocks (Fig. 2B) located on a floor surface divided in the middle such that half of the floor is red, and the other half is green. In the critical (Physics) condition, participants judged whether—if the tower tumbles—more blocks would land on the red part of the floor or the green part of the floor. In the control (Color) condition, participants judged whether the tower contained more yellow or more blue blocks (Fig. 2B). The stimuli were visually identical between the two conditions, and the tasks were matched for difficulty (Fischer et al., 2016; see Supp. Table 6 for the behavioral data in the current study, which replicate this difficulty matching). The Physics > Color judgment contrast targets cognitive processes related to intuitive physical reasoning. Each stimulus video presentation was 6 s long and the camera viewpoint moved 360° completely circling the block tower. The towers consisted of between 13 and 39 blocks, and the number of blue vs. yellow blocks differed by one to six in every tower (except for one trial in the Color condition where the number of blue and yellow blocks was the same due to a script error; this trial was excluded from the behavioral accuracy analyses). Each video was preceded by a question which appeared on the screen for 1 s cuing the type of judgment the participant had to perform: either “where will it fall?” for the critical, Physics task or “more blue or yellow?” for the control, Color task. The videos were followed by a 2 s response period with a blank screen, for a total trial duration of 9 s. Trials were grouped into blocks of 2 trials of the same condition (18 s total). Each scanning run consisted of 20 blocks (10 per condition) and 3 blocks of a baseline blank screen (18 s each), for a total run duration of 414 s. Condition order was counterbalanced across runs. Each participant completed 2 runs.

**The Multiple Demand system localizer.** This spatial working memory task, introduced in Fedorenko et al. (2011) and used in many subsequent studies as a localizer for the MD system (Blank et al., 2014; Shashidhara et al., 2019, 2020, 2024; Shashidhara and Erez, 2021; Diacheck et al., 2020; Malik-Moraleda et al., 2022), included two conditions in a blocked design. Participants had to keep track of spatial locations presented in a sequence (8 locations in the Hard condition, 4 locations in the Easy condition) (Fig. 2E). The Hard > Easy contrast targets cognitive processes broadly related to performing demanding tasks—what is often referred to by an umbrella term ‘executive function processes’. Each trial consisted of a brief fixation cross shown for 500 ms followed by 4 sequential flashes of unique locations within the 3 × 4 grid (1 s per flash; two locations at a time in the Hard condition, one location at a time in the Easy condition). Each trial ended with a two-alternative, forced-choice question (two sets of locations were presented for up to 3.25 s, and participants had to choose the set of locations they just saw; if they responded before 3.25 s elapsed, there was a blank screen for the remainder of the 3.25 s period). Finally, participants were given feedback in the form of a green checkmark (correct response) or a red cross (incorrect response or no response) shown for 250 ms. The total trial

duration was 8 s. Trials were grouped into blocks of 4 trials of the same condition (32 s total). Each scanning run consisted of 12 blocks (6 per condition) and 4 blocks of a baseline fixation screen (16 s each), for a total run duration of 448 s. Condition order was counterbalanced across runs. Each participant completed 2 runs.

Importantly, the Hard > Easy spatial working memory contrast generalizes to other contrasts of more vs. less demanding conditions (e.g., Duncan and Owen, 2000; Fedorenko et al., 2013; Hugdahl et al., 2015; Shashidhara et al., 2019; Assem et al., 2020b), and a system that closely corresponds to the one activated by the MD system localizer emerges from task-free (resting state) data (e.g., Assem et al., 2020b; Braga et al., 2020; Du et al., 2024; Rajimehr et al., 2024). Nevertheless, we also report a version of the analyses which used an alternative MD system localizer, to demonstrate generalizability (Supp. Figure 4).

**The language localizer.** This localizer, introduced in Fedorenko et al. (2010) and used in many subsequent studies (e.g., Blank et al., 2016; Fedorenko et al., 2020; Hu et al., 2022; Chen et al., 2023; Tuckute et al., 2024; Shain, Kean et al., 2024; the task is available for download from <https://www.evlab.mit.edu/resources>). Participants silently read sentences and lists of unconnected, pronounceable nonwords in a blocked design (Fig. 3B). The Sentences > Nonwords contrast targets cognitive processes related to high-level language comprehension, including understanding word meanings and combinatorial linguistic processing. Each stimulus (sentence or nonword list) was 6 s long and consisted of 12 words or nonwords presented one word/nonword at a time at the rate of 450 ms per word/nonword. The main task was attentive reading. Each stimulus was followed by a simple button-press task, which was included to maintain alertness. Trials were grouped into blocks of 3 trials of the same condition (18 s total). Each scanning run consisted of 16 blocks (8 per condition) and 5 blocks of a baseline blank screen (14 s each), for a total run duration of 358 s. Condition order was counterbalanced across runs. Each participant completed 2 runs.

Importantly, the Sentences > Nonwords contrast generalizes across presentation modalities (e.g., reading vs. listening), types of tasks, stimuli within a language (e.g., sentences vs. passages), and diverse languages (e.g., Fedorenko et al., 2010; Scott et al., 2017; Ivanova et al., 2020; Chen et al., 2023; Malik-Moraleda et al., 2022; see Fedorenko et al., 2024a,b for a review). Moreover, a system that closely corresponds to the one activated by the language localizer emerges from task-free (resting state) data (Braga et al., 2020; Du et al., 2024; Rajimehr et al., 2024). All brain regions identified by this contrast show sensitivity to lexico-semantic processing (e.g., stronger responses to real words than nonwords), combinatorial syntactic and semantic processing (e.g., stronger responses to sentences than to unstructured word lists, and sensitivity to syntactic complexity) (e.g., Fedorenko et al., 2010, 2016, 2020; Blank et al., 2016; Shain et al., 2022; Shain, Kean et al., 2024), and to sub-lexical regularities (Bozic et al., 2014; Regev et al., 2024).

### 3. Data acquisition, preprocessing, and first-level modeling

#### 3.1. Data acquisition

Whole-brain structural and functional data were collected on a whole-body 3 Tesla Siemens Prisma/Prisma-fit scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. Slightly different imaging sequences were used across two subsets of the participants. For 23 participants, T1-weighted structural images were collected in 176 axial slices with 1 mm isotropic voxels (repetition time (TR) = 2530 ms; echo time (TE) = 3.57 ms; TI = 900 ms, flip = 9°). Functional blood oxygenation level-dependent (BOLD) data were acquired using a simultaneous multi-slice (SMS) sequence with a 90° flip angle and using an acceleration factor of 3, with the following acquisition parameters: 66 2 mm thick slices acquired in the interleaved order (slice gap = 0 mm), 2 mm × 2 mm in-plane resolution, FoV in the phase encoding (A

» P) direction 204 mm and matrix size 102 × 102, TR = 2000 ms and TE = 35 ms. For the remaining 17 participants, T1-weighted structural images were collected in 208 sagittal slices with 0.85 mm isotropic voxels (TR = 1800 ms, TE = 2.37 ms, TI = 900 ms, flip = 8°). Functional, blood oxygenation level-dependent (BOLD) data were acquired using an SMS sequence with a 90° flip angle and using a slice acceleration factor of 3, with the following acquisition parameters: 72 2 mm thick near-axial slices acquired in the interleaved order (with 10 % distance factor), 2 mm × 2 mm in-plane resolution, FoV in the phase encoding (F » H) direction 208 mm and matrix size 104 × 104, TR = 2000 ms, TE = 30 ms, and partial Fourier of 7/8. The first 10 s of each run were excluded to allow for steady state magnetization.

#### 3.2. Preprocessing

fMRI data were preprocessed and analyzed using SPM12 (release 7487), CONN EvLab module (release 19b), and other custom MATLAB scripts. Each participant's functional and structural data were converted from DICOM to NIFTI format. All functional scans were coregistered and resampled using B-spline interpolation to the first scan of the first session (Friston et al., 1995). Potential outlier scans were identified from the resulting subject-motion estimates as well as from BOLD signal indicators using default thresholds in CONN preprocessing pipeline (5 standard deviations above the mean in global BOLD signal change, or framewise displacement values above 0.9 mm; Nieto-Castañón, 2020). Functional and structural data were independently normalized into a common space (the Montreal Neurological Institute [MNI] template; IXI549Space) using SPM12 unified segmentation and normalization procedure (Ashburner and Friston, 2005) with a reference functional image computed as the mean functional data after realignment across all timepoints omitting outlier scans. The output data were resampled to a common bounding box between MNI coordinates (-90, -126, -72) and (90, 90, 108), using 2 mm isotropic voxels and 4th order spline interpolation for the functional data, and 1 mm isotropic voxels and trilinear interpolation for the structural data. Last, the functional data were smoothed spatially using spatial convolution with a 4 mm FWHM Gaussian kernel.

#### 3.3. First-level modeling

For all experiments, effects were estimated using a general linear model (GLM) in which each experimental condition was modeled with a boxcar function convolved with the canonical hemodynamic response function (HRF) (fixation was modeled implicitly, such that all timepoints that did not correspond to one of the conditions were assumed to correspond to a fixation period). Temporal autocorrelations in the BOLD signal timeseries were accounted for by a combination of high-pass filtering with a 128 s cutoff, and whitening using an AR (0.2) model (first-order autoregressive model linearized around the coefficient  $\alpha = 0.2$ ) to approximate the observed covariance of the functional data in the context of restricted maximum likelihood estimation. In addition to experimental condition effects, the GLM design included first-order temporal derivatives for each condition (included to model variability in the HRF delays), as well as nuisance regressors to control for the effect of slow linear drifts, subject-motion parameters, and potential outlier scans on the BOLD signal.

#### 3.4. Second-level fMRI analyses

The analyses were performed using the spm\_ss toolbox ([http://www.nitrc.org/projects/spm\\_ss](http://www.nitrc.org/projects/spm_ss)), which interfaces with SPM and the CONN toolbox (<https://www.nitrc.org/projects/conn>).

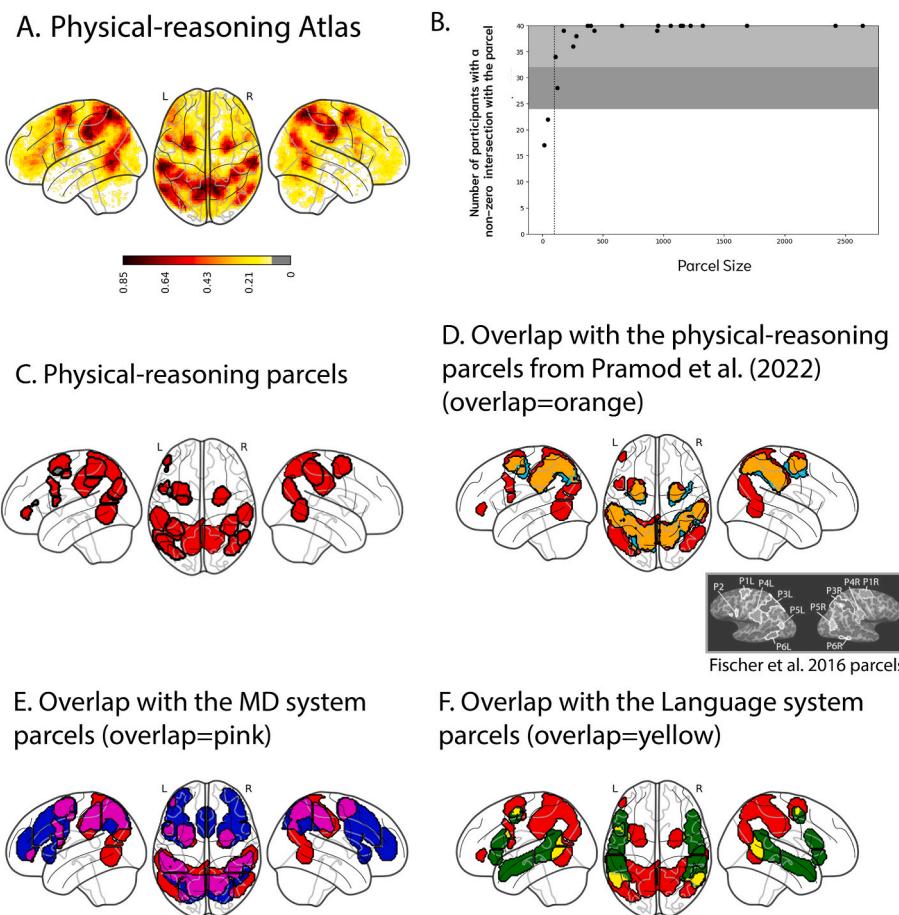
#### 3.5. fROI definition and response estimation

**Definition of the physical-reasoning fROIs:** The initial Fischer et al.

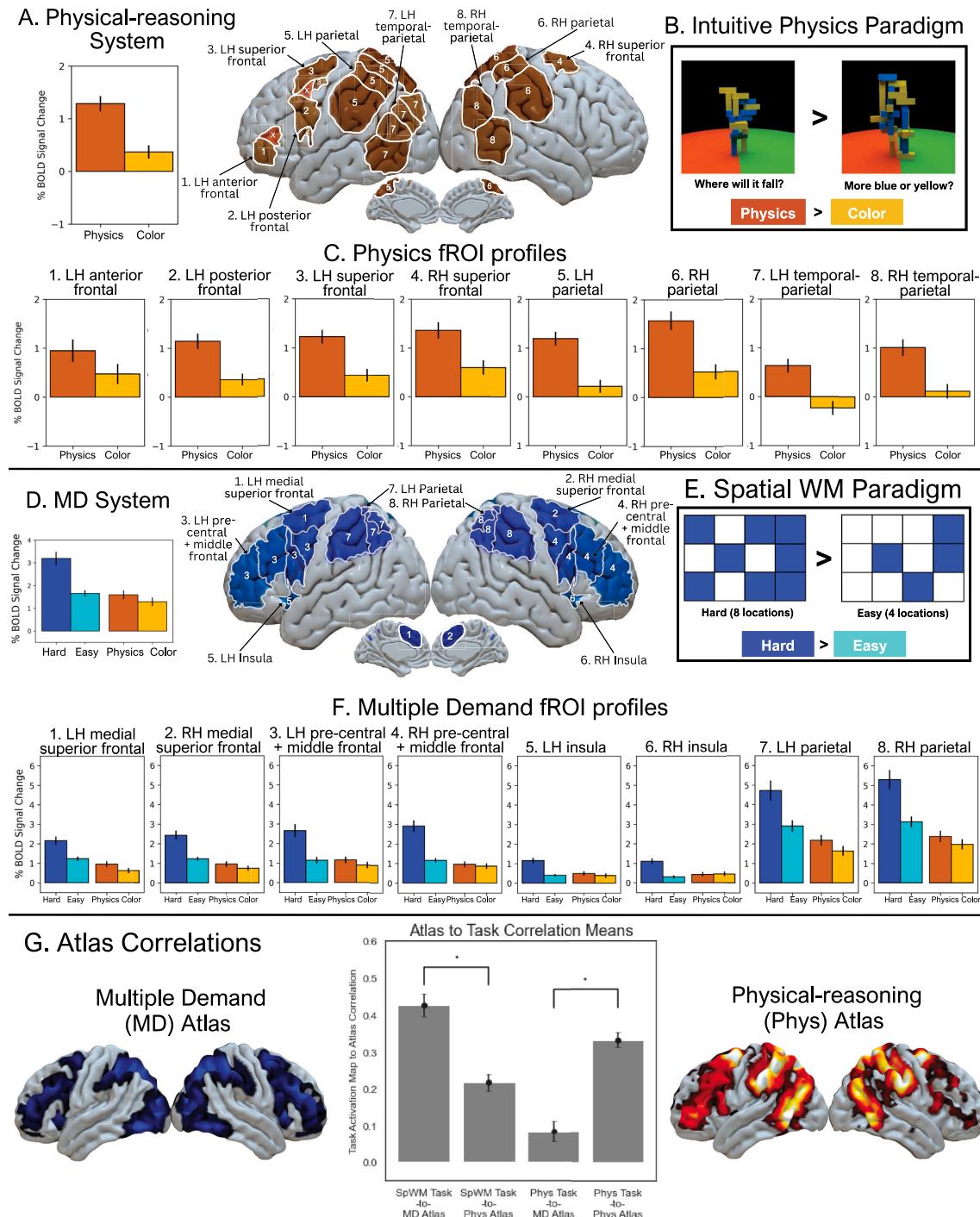
(2016) study included 12 participants; because our set of participants was larger ( $n = 40$ ) and probabilistic overlap maps tend to show greater stability with more participants (e.g., Lipkin et al., 2022), we first performed a group-constrained subject-specific (GSS) analysis (Fedorenko et al., 2012; Julian et al., 2012) on our data in order to create a set of parcels that would be used for defining individual-level fROIs. GSS is a whole-brain analysis that identifies spatially consistent (across participants) areas of activation for some contrast of interest. To do so, we first thresholded the individual t-maps for the Physics > Color contrast by selecting the 10 % of most responsive voxels across the brain. These maps were then binarized (selected voxels were turned into 1s and the remaining voxels into 0s) and overlaid to create a probabilistic overlap map (summing the ones and zeros across participants in each voxel). After dividing the summed value in each voxel by the number of participants (40, in this case), these values can be interpreted as the proportion of the participants for whom that voxel belonged to the top 10 % of most responsive voxels (Fig. 1A). This probabilistic overlap map was

then thresholded, such that voxels with values of 0.1 or lower were removed, and a watershed algorithm was used to segment the map into discrete regions (parcels).

The parcels were evaluated on two criteria. First, each parcel was intersected with the individual binarized activation maps to calculate how many participants have task-responsive voxels within the parcel boundaries. Parcels where 24/40 (60 %) or more of the participants had task-responsive voxels were included (Fig. 1B). And second, we evaluated the replicability of the Physics > Color contrast. To do so, we used an across-runs cross-validation approach to ensure independence between the data used to define the fROIs and to estimate the effects (e.g., Kriegeskorte, 2011; Nieto-Castañón and Fedorenko, 2012). In particular, we used run 1 of the task to define the functional regions of interest (fROIs) (as the top 10 % of most responsive voxels within each parcel, based on the t-values for the Physics > Color contrast; note that this approach ensures that a fROI is defined in every participant, cf. a fixed statistical threshold approach, for which some participants may not have

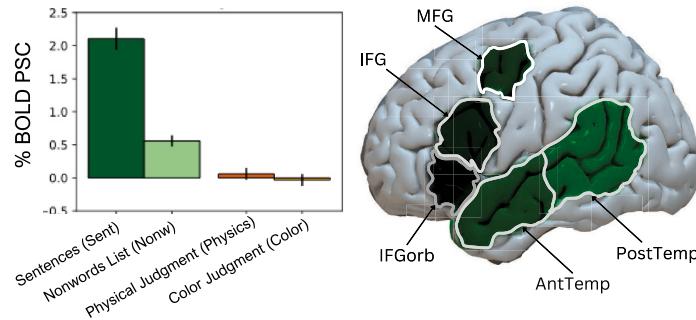


**Fig. 1.** The physical-reasoning system and its broad anatomical overlap with other systems. A. A probabilistic atlas for the physical-reasoning system based on  $n = 40$  participants in the current study. This atlas is derived from individual activation maps for the Physics > Color contrast, where in each individual, we selected the top 10 % of voxels that respond most reliably to the Physics > Color contrast. Areas with darker colors (orange and red) correspond to areas where a higher proportion of participants showed voxel-level overlap. B. The relationship between the size of the parcels (for the full set of 21 parcels that resulted from the GSS analysis) and the number of participants that have a non-zero intersection with the parcel (i.e., at least 1 voxel within the borders of the parcel was selected as the top 10 % of Physics > Color voxels across the brain). The parcels ( $n = 18$ ) that fall within the light gray area have a nonzero intersection with 32 or more of the 40 participants (80 % or more); the parcel that falls within the dark gray area has a nonzero intersection with 24 or more of the participants (60 % or more). C. The 21 parcels that emerged from the GSS analysis. Two parcels that did not meet our selection criteria are shown in gray. D. Overlap between our physical-reasoning parcels (red) and those from Pramod et al. (2022) (light blue). The overlap is shown in orange. Note that Pramod et al.'s set excludes the temporo-parietal parcels bilaterally and the left posterior frontal parcels (see text for details). Otherwise, there is good concordance between the two sets, except that our analysis reveals an additional area in the anterior left frontal lobe. The inset shows the parcels from the original Fischer et al. (2016) study, which also appear similar to our set. E. Overlap between our physical-reasoning parcels (red) and the Multiple Demand parcels (blue; these parcels were derived from a GSS analysis on a dataset of 197 participants; Lipkin et al., 2022). The two sets of parcels show overlap in both frontal and parietal areas. F. Overlap between our physical-reasoning parcels (red) and the language parcels (green; these parcels were derived from a GSS analysis on a dataset of 220 participants). The two sets of parcels show overlap in both frontal and temporal areas. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

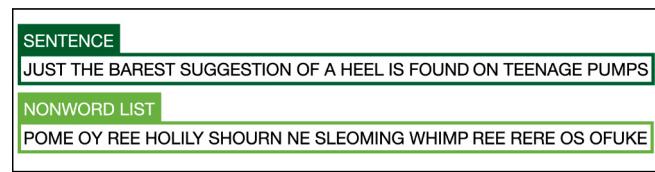


**Fig. 2. The physical-reasoning system and its relationship with the Multiple Demand system** **A.** The response in the physical-reasoning fROIs to the physical-reasoning localizer conditions (Physics, Color), averaged across the fROIs; and the physical-reasoning system parcels (excluding parcels 20, 21, shown in red; see Methods). **B.** The physical-reasoning localizer. During the critical (Physics) condition, participants answered “where will it fall?” by judging whether the block tower would fall towards the green or red side of the floor; during the control (Color) condition, participants answered “more blue or yellow?” by judging whether the block tower consisted of more yellow or blue blocks (see Methods for details). **C.** The response in the physical-reasoning fROIs, broken down by fROI group, to the physical-reasoning localizer conditions (Physics, Color). We observe a strong Physics > Color effect in all fROI groups (Table 1A). **D.** The response in the MD system fROIs to the spatial working memory MD localizer and physical-reasoning localizer conditions (Hard, Easy, Physics, Color), averaged across the fROIs; and the MD system parcels. **E.** The spatial working memory MD localizer. Participants were tasked with remembering four sequential flashes of unique locations with a  $3 \times 4$  grid. During the Hard condition, each flash consisted of a pair of locations, while during the Easy condition, each flash consisted of a single location (see Methods for details). **F.** The response in the MD system fROIs, broken down by fROI group, to the MD localizer and physical-reasoning localizer conditions (Hard, Easy, Physics, Color). We observe a strong Hard > Easy effect in all fROI groups (Table 2A). **G.** Pearson correlation between physical-reasoning and MD system atlases to the physical-reasoning localizer and spatial working memory MD localizer task contrasts (Physics > Color, Hard > Easy). Here and elsewhere, error bars show standard error of the mean across participants. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

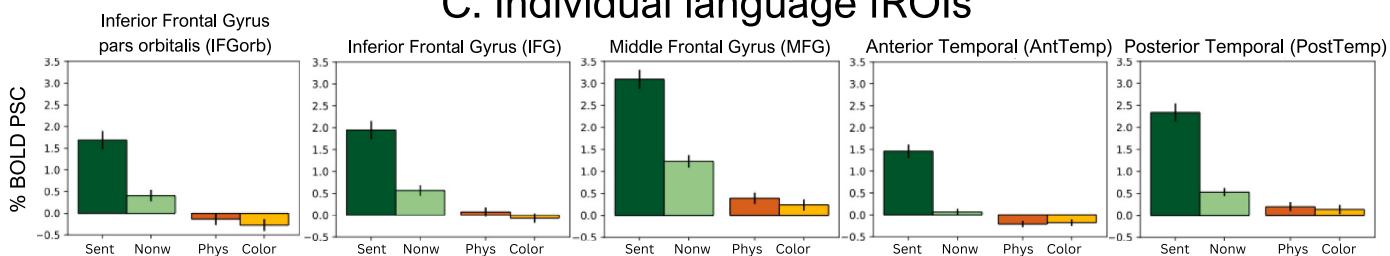
## A. Language Network



## B. Language Localizer Paradigm



## C. Individual language fROIs



**Fig. 3. The language system and its relationship to physical-reasoning.** A. The response in the language system fROIs to the language localizer and to the physical-reasoning localizer conditions (Sentences, Nonwords, Physics, Color), averaged across the fROIs; and the language system parcels. B. The language localizer. During Sentences trials, participants viewed complete sentences one word at a time. During Nonwords trials, participants viewed nonword lists (see [Methods](#) for details). C. The response to the language localizer and physical-reasoning localizer conditions (Sentences, Nonwords, Physics, Color), broken down by individual fROI within each system. We observe a strong Sentences > Nonwords effect in all fROIs (Table 3A). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

any significant voxels in a given parcel) and run 2 to estimate the responses; then we used run 2 to define the fROIs and run 1 to estimate the responses; finally, we averaged these estimates to obtain a single estimate per participant per parcel. Parcels where the Physics > Color contrast reliably differed from zero were included in the critical analyses.

To examine the responses in the physics fROIs to the conditions of other tasks, the fROIs were defined using the data from both runs of the physics localizer.

**Definition of the MD fROIs:** Each individual map for the Hard > Easy spatial working memory contrast from the MD localizer was intersected with a set of 20 parcels (10 in each hemisphere). These parcels (available at <https://www.evlab.mit.edu/resources>) were derived from a probabilistic activation overlap map for the same contrast in a large set of independent participants ( $n = 197$ ) and covered the frontal and parietal components of the MD system bilaterally (Duncan, 2010; Fedorenko et al., 2013). Within each parcel, a participant-specific MD fROI was defined as the top 10 % of voxels with the highest t-values for the localizer contrast. To estimate the response in the MD fROIs to the conditions of the MD localizer, the same cross-validation procedure was used as described above. As expected, the MD fROIs showed a robust Hard > Easy spatial working memory effect ( $ps < 0.001$ ,  $|d|s > 1.84$ ; here and elsewhere, p-values are corrected for the number of fROIs using the false discovery rate (FDR) correction (Benjamini and Yekutieli, 2001)).

**Definition of the language fROIs:** Each individual map for the Sentences > Nonwords contrast from the language localizer was intersected with a set of 5 parcels. These parcels (available at <https://www.evlab.mit.edu/resources>) were derived from a probabilistic activation overlap map for the same contrast in a large set of independent participants ( $n = 220$ ) and covered the fronto-temporal language system in the left hemisphere (Fedorenko et al., 2024a,b). Within each parcel, a participant-specific language fROI was defined as the top 10 % of voxels with the highest t-values for the localizer contrast. To estimate the

response in the language fROIs to the conditions of the language localizer, the same cross-validation procedure was used as described above, to ensure independence. As expected, the language fROIs showed a robust Sentences > Nonwords effect ( $ps < 0.001$ ,  $|d|s > 2.05$ ).

### 3.6. Statistical analyses of fROI response profiles

All analyses were performed with linear mixed-effects models using the “lme4” package in R (version 1.1.26; Bates et al., 2015) with p value approximation performed by the “lmerTest” package (version 3.1.3; Kuznetsova et al., 2017) and effect sizes (Cohen’s d) estimated by the “EMATools” package (version 0.1.3; Kleiman et al., 2017).

Past work on the physical-reasoning, language, and MD systems has established that different regions within each system show functionally similar responses. However, to allow for potential differences in the degree of inter-system overlap in different parts of the brain, we chose to differentiate among the regions in each system. For the physical-reasoning and MD systems, we grouped regions into a few sets based on anatomy. In particular, for the physical-reasoning system, we grouped fROIs into eight sets: left hemisphere (LH) anterior frontal, LH posterior frontal, LH and right hemisphere (RH) superior frontal, LH and RH parietal, and LH and RH temporal-parietal (Fig. 2C). For the MD system, we also grouped fROIs into eight sets: LH and RH medial-superior frontal, LH and RH precentral + middle frontal, LH and RH insular, and LH and RH parietal (Fig. 2F; see also Supp. Figures 2–3 for the results on the individual fROIs for the physical-reasoning and the MD systems). For the language system, we examined the five fROIs separately.

First, to examine responses to each contrast in each system, we fit a linear mixed-effect regression model for each fROI (for the language system) or fROI group (for the physical-reasoning and MD systems) predicting the level of BOLD response from condition, with random intercepts for fROIs (when groups consisted of multiple fROIs) and participants:

$$\text{BOLD} \sim \text{Condition} + (1 | \text{fROI}) + (1 | \text{Participant})$$

For comparisons between systems for a given contrast (e.g., asking whether the Physics > Color contrast is significantly larger in the physical-reasoning system compared to the MD system), we fit a linear mixed-effect regression model predicting the level of BOLD response from condition, system, and their interaction, with random intercepts for fROIs and participants:

$$\text{BOLD} \sim \text{Condition} * \text{System} + (1 | \text{fROI}) + (1 | \text{Participant})$$

In a similar fashion, for comparisons between tasks within a system (e.g., asking whether in the physical-reasoning system the Physics > Color contrast is significantly larger than the Hard > Easy contrast from the spatial working memory task), we fit a linear mixed-effect regression model predicting the level of BOLD response from condition (critical vs. control), task (e.g., physics vs. spatial WM), and their interaction, with random intercepts for fROIs and participants:

$$\text{BOLD} \sim \text{Condition} * \text{Task} + (1 | \text{fROI}) + (1 | \text{Participant})$$

### 3.7. Whole-brain multivariate correlation and Dice overlap analyses

Given that, as will be discussed below, the inter-system overlap analyses via univariate fROI response profiles revealed some overlap between the physical-reasoning and the MD systems, we asked whether the activation patterns for the Physics > Color and Hard > Easy Spatial WM contrasts may be dissociable in their fine-grained spatial topographies, similar to an analysis reported in Fischer et al. (2016). To do so, we first created probabilistic activation overlap atlases for each contrast. This procedure is described for the Physics > Color contrast in “fROI definition and response estimation”; for the Hard > Easy contrast, we used the probabilistic atlas created from  $n = 691$  participants (Lipkin et al., 2022; doi.org/10.6084/m9.figshare.22306348). We then computed a correlation between each individual map for the Physics > Color contrast ( $n = 40$  participants) and each of the two atlases, and between each individual map for the Hard > Easy contrast ( $n = 29$  participants) and each atlas. The distributions of these correlations were compared via independent-samples t-tests, to see whether the maps for the Physics > Color contrast are more similar to the physical-reasoning system atlas versus the MD system atlas, and whether the maps for the Hard > Easy spatial working memory contrast are more similar to the MD system atlas versus the physical-reasoning system atlas (Fig. 2G). For completeness, we additionally computed for each participant a Dice overlap measure (Dice, 1945; Sørensen, 1948) between the activation maps for the Physics > Color contrast and the Hard > Easy contrast (and other contrasts, for comparison). We used several fixed whole-brain statistical thresholds:  $p < 0.001$ ,  $p < 0.01$ , and  $p < 0.05$  uncorrected thresholds; we included quite liberal thresholds to increase our chances of observing overlap if it exists.

## 4. Results

### 1. Replication and extension of prior findings on the physical-reasoning system (Fischer et al., 2016): A set of bilateral frontal, temporal, and parietal areas support physical reasoning.

Using an fMRI localizer paradigm introduced in Fischer et al. (2016), we identified a set of brain areas engaged during physical reasoning. The paradigm is based on a contrast of judgments about physical stability of rotating block towers (the critical condition) vs. about the color composition of the same towers (Fig. 2B). The critical condition requires participants to rely on their intuitions about the tower’s center of gravity to decide whether more blocks would fall on one or the other half of the floor surface. A whole-brain GSS analysis (see Methods) identified 21

parcels corresponding to areas of spatially consistent (across participants) activation for the Physics > Color contrast (Fig. 1A–C). Based on the combination of two criteria—presence in 60 % or more of the participants and replicability of the Physics > Color effect across runs—19 of the 21 parcels were selected for the critical analyses (Fig. 1C).

The topography of these parcels—spanning frontal and parietal cortices bilaterally—closely mirrors Fischer et al.’s (2016) findings (Fig. 2B in the 2016 paper, included as an inset in Fig. 1D) and a subsequent study using the same localizer paradigm (Pramod et al., 2022), and the behavioral data again show similar performance between the critical and control conditions (Supp. Table 6). However, in our data, an additional area in the left anterior frontal lobe emerged, which passed our inclusion criteria. This area was likely missed in the earlier studies because of its small size; smaller areas become easier to detect with larger samples of participants (e.g., Lipkin et al., 2022). The fROIs defined within these parcels all show a robust Physics > Color effect, as estimated using an across-runs cross-validation approach ( $ps$  for all fROI groups  $< 0.001$ ; Fig. 2A–C; Table 1A).

Note also that both Fischer et al. (2016) and Pramod et al. (2022) exclude a subset of the parcels from consideration—the bilateral temporo-parietal and the LH posterior frontal ones—because they failed to pass an additional test (see Fig. 1D for the comparisons of the parcels). In particular, in addition to the physical stability localizer, Fischer et al. (2016) had participants perform a task where they were asked to make physical vs. social judgments about simple moving geometric stimuli (two colored dots). Only the bilateral superior frontal and parietal regions showed a physical > social effect in that task. In our analyses, we examine the full set of regions for completeness, but we also report a version of the critical analyses where we only include the bilateral superior frontal and parietal regions for ease of comparison with prior work (see the rows referring to the Fischer subset in Table 1A–B, 4A–B, and Supp. Tables 1A–B).

### 2. The physical-reasoning system is at least partially dissociable from the Multiple Demand system.

Next, we examined the relationship between the physical-reasoning system and the Multiple Demand (MD) system, whose topography broadly resembles the physical-reasoning system (Fig. 1E). For this analysis, we used the subset of 29 participants, who performed an MD system localizer (Fig. 2E). We performed two types of analyses to assess inter-system overlap.

First, we examined the response profiles in the two sets of fROIs, starting with the MD fROIs. Replicating much prior work (e.g., Fedorenko et al., 2013; Assem et al., 2020a,b), the MD fROIs show a robust Hard > Easy effect for the spatial working memory task (used as the localizer), as estimated using an across-runs cross-validation approach ( $ps$  for all fROI groups  $< 0.001$ ; Fig. 2D–F; Table 2A). Critically, several of the MD fROI groups also show a positive (and a few—significant) Physics > Color contrast (Table 2B; see Supp. Table 3A for the responses of the MD fROIs to all contrasts, and Supp. Table 3B for comparisons across tasks). However, the effect is overall small (Fig. 2D–F) and significantly smaller than in the physical-reasoning system, as supported by a reliable condition-by-system interaction ( $p < 0.001$ ; Table 4A; see Supp. Fig. 4 for a side-by-side comparison of the MD system’s responses when the MD system is localized by the main, spatial working memory localizer versus an alternative localizer based on an arithmetic addition task; the results are similar across the two ways of defining the MD fROIs). That said, in the flip-side analysis, where we examined the responses of the physical-reasoning fROIs to the spatial working memory task, we found strong responses and significant Hard > Easy effects in six of the eight fROI groups (Supp. Fig. 3; see Supp. Table 1A for the responses of the physical-reasoning fROIs to all contrasts, and Supp. Table 1B for comparisons across tasks), although the size of the Hard > Easy effect is significantly smaller than in the MD system, as supported by a reliable condition-by-system interaction ( $p < 0.001$ ; Table 4B). It is

**Table 1a**

The responses of the physical-reasoning system to the Physics > Color contrast. This table reports the beta estimates, standard errors of the mean, t-values, and p-values from a linear mixed-effects model (see [Methods](#)). We first report the effects for the system as a whole, for the core regions that [Fischer et al. \(2016\)](#) and subsequent studies focused on (for ease of comparisons with those earlier studies), and then for each fROI group separately. For the fROI groups, we report uncorrected p-values, but we mark the values that survive the Bonferroni correction for the number of fROI groups (n = 9: 8 groups and the system as a whole) in **bold font**.

Task	Type	Beta	SEM	t	p
Phys	System	0.886829961	0.042281108	20.97461518	< 0.001
Phys	System_CoreRegions	1.01288101	0.05558128	18.2234187	< 0.001
Phys	Group Phys_LH_ant_front	0.476769375	0.128333818	3.71507201	< 0.001
Phys	Group Phys_LH_post_front	0.784105	0.106551568	7.358925023	< 0.001
Phys	Group Phys_LH_sup_front	0.78933915	0.097375336	8.10615074	< 0.001
Phys	Group Phys_RH_sup_front	0.76371005	0.097112027	7.864216993	< 0.001
Phys	Group Phys_LH_parietal	1.085572613	0.096197281	11.28485751	< 0.001
Phys	Group Phys_RH_parietal	1.148043775	0.110693093	10.37141292	< 0.001
Phys	Group Phys_LH_temp_par	0.865920931	0.086946198	9.959273049	< 0.001
Phys	Group Phys_RH_temp_par	0.894639563	0.116689265	7.666854055	< 0.001

**Table 1b**

The responses of the physical-reasoning system to the Sentences > Nonwords contrast. This table reports the beta estimates, standard errors of the mean, t-values, and p-values from a linear mixed-effects model (see [Methods](#)). We first report the effects for the system as a whole, and then for each fROI group separately. For the fROI groups, we report uncorrected p-values, but we mark the values that survive the Bonferroni correction for the number of fROI groups (n = 9) in **bold font**. (Two fROI groups here show a significant effect for the opposite, Nonwords > Sentences contrast.)

Task	Type	Beta	SEM	t	p
Lang	System	-0.075711934	0.039322957	-1.925387621	0.0544
Lang	System_CoreRegions	-0.1465097	0.03765454	-3.8908903	< 0.001
Lang	Group Phys_LH_ant_front	-0.560139487	0.137039845	-4.0874206	< 0.001
Lang	Group Phys_LH_post_front	-0.189940833	0.117520655	-1.616233612	0.1090
Lang	Group Phys_LH_sup_front	-0.092001038	0.072940223	-1.261321044	0.2100
Lang	Group Phys_RH_sup_front	-0.017213538	0.044340009	-0.388216845	0.7000
Lang	Group Phys_LH_parietal	-0.279767731	0.075482864	-3.706374084	< 0.001
Lang	Group Phys_RH_parietal	-0.048270043	0.069019309	-0.699370127	0.4850
Lang	Group Phys_LH_temp_par	0.133383455	0.107004067	1.246526968	0.2140
Lang	Group Phys_RH_temp_par	0.048880705	0.111723531	0.437514861	0.6630

**Table 2a**

The responses of the MD system to the Hard > Easy contrast. This table reports the beta estimates, standard errors of the mean, t-values, and p-values from a linear mixed-effects model (see [Methods](#)). We first report the effects for the system as a whole, and then for each fROI group separately. For the fROI groups, we report uncorrected p-values, but we mark the values that survive the Bonferroni correction for the number of fROI groups (n = 9) in **bold font**.

Task	Type	Beta	SEM	t	p
SpWM	System	1.537551	0.087604	17.551151	< 0.001
SpWM	Group MD_LH_med_sup_front	0.92688174	0.21645264	4.28214564	< 0.001
SpWM	Group MD_RH_med_sup_front	1.20964828	0.20522222	5.89433387	< 0.001
SpWM	Group MD_LH_prec_mid_front	1.5104128	0.19686781	7.6722182	< 0.001
SpWM	Group MD_RH_prec_mid_front	1.7424743	0.21176718	8.22825468	< 0.001
SpWM	Group MD_LH_insula	0.75599296	0.1167514	6.47523653	< 0.001
SpWM	Group MD_RH_insula	0.803655	0.13206182	6.08544538	< 0.001
SpWM	Group MD_LH_parietal	1.80720287	0.28656473	6.30643849	< 0.001
SpWM	Group MD_RH_parietal	2.16171833	0.32748154	6.60103869	< 0.001

**Table 2b**

The responses of the MD system to the Physics > Color contrast. This table reports the beta estimates, standard errors of the mean, t-values, and p-values from a linear mixed-effects model (see [Methods](#)). We first report the effects for the system as a whole, and then for each fROI group separately. For the fROI groups, we report uncorrected p-values, but we mark the values that survive the Bonferroni correction for the number of fROI groups (n = 9) in **bold font**.

Task	Type	Beta	SEM	t	p
Phys	System	0.3082425	0.05517682	5.58644875	< 0.001
Phys	Group MD_LH_med_sup_front	0.36396395	0.08676529	4.19481053	< 0.001
Phys	Group MD_RH_med_sup_front	0.22140019	0.10975274	2.0172634	0.0468
Phys	Group MD_LH_prec_mid_front	0.3170827	0.10217979	3.10318422	0.0022
Phys	Group MD_RH_prec_mid_front	0.11426385	0.11337393	1.00784943	0.3150
Phys	Group MD_LH_insula	0.07258097	0.04499141	1.61321812	0.1180
Phys	Group MD_RH_insula	-0.04065452	0.04435981	-0.9164648	0.3670
Phys	Group MD_LH_parietal	0.60637771	0.15429581	3.92996881	< 0.001
Phys	Group MD_RH_parietal	0.47255856	0.17861755	2.64564464	0.0091

**Table 3a**

The responses of the language system to the Sentences > Nonwords contrast. This table reports the beta estimates, standard errors of the mean, t-values, and p-values from a linear mixed-effects model (see [Methods](#)). We first report the effects for the system as a whole, and then for each fROI separately. For the fROIs, we report uncorrected p-values, but we mark the values that survive the Bonferroni correction for the number of fROIs ( $n = 6$ ) in **bold font**.

Task	Type	Beta	SEM	t	p
Lang	System	0.992064288	0.053534777	18.5312119	< 0.001
Lang	LH Inferior Frontal Gyrus (orbital)	1.284653795	0.152222582	8.439311569	< 0.001
Lang	LH Inferior Frontal Gyrus	1.379889923	0.155064095	8.898835843	< 0.001
Lang	LH Middle Frontal Gyrus	1.86698841	0.178961683	10.43233604	< 0.001
Lang	LH Anterior Temporal	1.389281538	0.118291459	11.74456341	< 0.001
Lang	LH Posterior Temporal	1.811737795	0.155516339	11.64982284	< 0.001

**Table 3b**

The responses of the language system to the Physics > Color contrast. This table reports the beta estimates, standard errors of the mean, t-values, and p-values from a linear mixed-effects model (see [Methods](#)). We first report the effects for the system as a whole, and then for each fROI separately. For the fROIs, we report uncorrected p-values, but we mark the values that survive the Bonferroni correction for the number of fROIs ( $n = 6$ ) in **bold font**.

Task	Type	Beta	SEM	t	p
Phys	System	0.105882318	0.038001317	2.786280264	0.0507
Phys	LH Inferior Frontal Gyrus (orbital)	0.137547462	0.058360468	2.356860173	0.0237
Phys	LH Inferior Frontal Gyrus	0.141801846	0.049455125	2.867283125	<b>0.0067</b>
Phys	LH Middle Frontal Gyrus	0.155651897	0.057364939	2.713362895	<b>0.0010</b>
Phys	LH Anterior Temporal	-0.031725077	0.034565603	-0.917822187	0.3650
Phys	LH Posterior Temporal	0.062227513	0.042287678	1.471528238	0.1490

**Table 4a**

The comparison of responses to the Physics > Color contrast between the MD system and the physical-reasoning system. This table reports the estimates, standard errors of the mean, and p-values from a linear mixed-effects model (see [Methods](#)). The critical interaction between system and contrast is shown in the last row.

	Predictor	Beta	SEM	p-value
All physical-reasoning regions	P vs. C (in ref = MD Sys)	0.31	0.05	<0.001
	P vs. C in <b>Phys</b> (vs. ref = MD Sys)	<b>0.58</b>	0.07	<0.001
Only the subset of core regions as defined by <a href="#">Fischer et al. (2016)</a>	P vs. C (in ref = MD Sys)	0.31	0.05	<0.001
	P vs. C in <b>Core</b> (vs. ref = MD Sys)	<b>0.70</b>	0.08	<0.001

**Table 4b**

The comparison of responses to the Hard > Easy contrast between the MD system and the physical-reasoning system. This table reports the estimates, standard errors of the mean, and p-values from a linear mixed-effects model (see [Methods](#)). The critical interaction between system and contrast is shown in the last row.

	Predictor	Beta	SEM	p-value
All physical-reasoning regions	H vs. E (in ref = MD Sys)	1.54	0.09	<0.001
	H vs. E in <b>Phys</b> (vs. ref = MD Sys)	<b>-0.64</b>	0.12	<0.001
Only the subset of core regions as defined by <a href="#">Fischer et al. (2016)</a>	H vs. E (in ref = MD Sys)	1.54	0.09	<0.001
	H vs. E in <b>Core</b> (vs. ref = MD Sys)	<b>-0.36</b>	0.14	0.011

interesting to note that the left anterior frontal physical-reasoning fROI, which did not emerge in the [Fischer et al. \(2016\)](#) study, shows the most selective profile relative to the spatial working memory task, with the response to the hard spatial memory condition being no higher than the response to the control, Color condition of the physical-reasoning localizer ([Supp. Fig. 3](#)); in all other physical reasoning fROIs, the response to the hard spatial working memory condition is as high or higher than the critical, Physics condition.

Given this partial overlap, in the second analysis, we examined fine-grained spatial topographies to see if they show a dissociation (see [Fischer et al., 2016](#); for a similar analysis). We found that the individual

activation maps for the physical-reasoning task show a stronger correlation with the probabilistic atlas for the physical-reasoning system (mean  $r = 0.33$ , SEM = 0.02) compared to the MD system atlas ( $r = 0.08$ , SEM = 0.03;  $t = 6.928$ ;  $p < 0.001$ ). In contrast, the individual activation maps for the spatial working memory task show a stronger correlation with the probabilistic atlas for the MD system (mean  $r = 0.43$ , SEM = 0.03) compared to the physical-reasoning system atlas ( $r = 0.08$ , SEM = 0.03;  $t = 5.544$ ;  $p < 0.001$ ). The complementary Dice overlap analysis mirrored these results: even at the most liberal ( $p < 0.05$  uncorrected) threshold, only 0.26 of the voxels overlapped between the two contrasts ([Supp. Table 5](#)), which is considered to be in the low-moderate range ([Wilson et al., 2017](#)). Thus, the two systems show some spatial overlap, but the fine-grained activation patterns are robustly distinct ([Fig. 2G](#)).

3. The physical-reasoning system does not overlap with the language system.

Finally and critically, we examined the relationship between the physical-reasoning system and the language-selective system—a relationship that has not been previously explored. We first examined the responses in the language fROIs to the physical-reasoning task. Repeating much prior work (see [Fedorenko et al., 2024a,b](#) for a review), the

**Table 5**

**Behavioral performance in the Physics and Color conditions of the physical-reasoning task.** This table reports the accuracies for the two conditions and the statistical comparison between them. As reported in [Fischer et al. \(2016\)](#), the two conditions are matched for accuracy. (Note that behavioral responses from 8 participants did not get recorded due to a technical error, so we report the data from the subset of 32 participants.)

Physics condition accuracy	Color condition accuracy	Student's t-test results
85.5 %	86.2 %	0.60

language regions show a robust Sentences > Nonwords effect, as estimated using an across-runs cross-validation approach ( $p_s$  for all fROIs  $<0.001$ ; Fig. 3A–C; Table 3A; see Supp. Table 2A for the responses of the language fROIs to all contrasts, and Supp. Table 2B for comparisons across tasks). Critically, however, these regions do not respond to the physical-reasoning task (Fig. 3C; Table 3B). The effect does not reach significance at the system level; it does reach significance in two fROIs, but a) the effect is small, and b) the responses to both the Physics and the Color conditions are close to the fixation baseline, and substantially below the control, Nonwords, condition of the language localizer. Similarly, in the flip-side analysis, where we examined the responses of the physical-reasoning fROIs to the language task, we found that they do not respond to the Sentences > Nonwords contrast; in fact, similar to what has been previously reported for the MD network (e.g., Fedorenko et al., 2011; 2013), most of the physical-reasoning fROI groups respond more strongly to the Nonwords compared to the Sentences condition, and some reliably so (Fig. 4A–C; Table 1B; see Supp. Table 5 for the concordant results from the Dice overlap analysis).

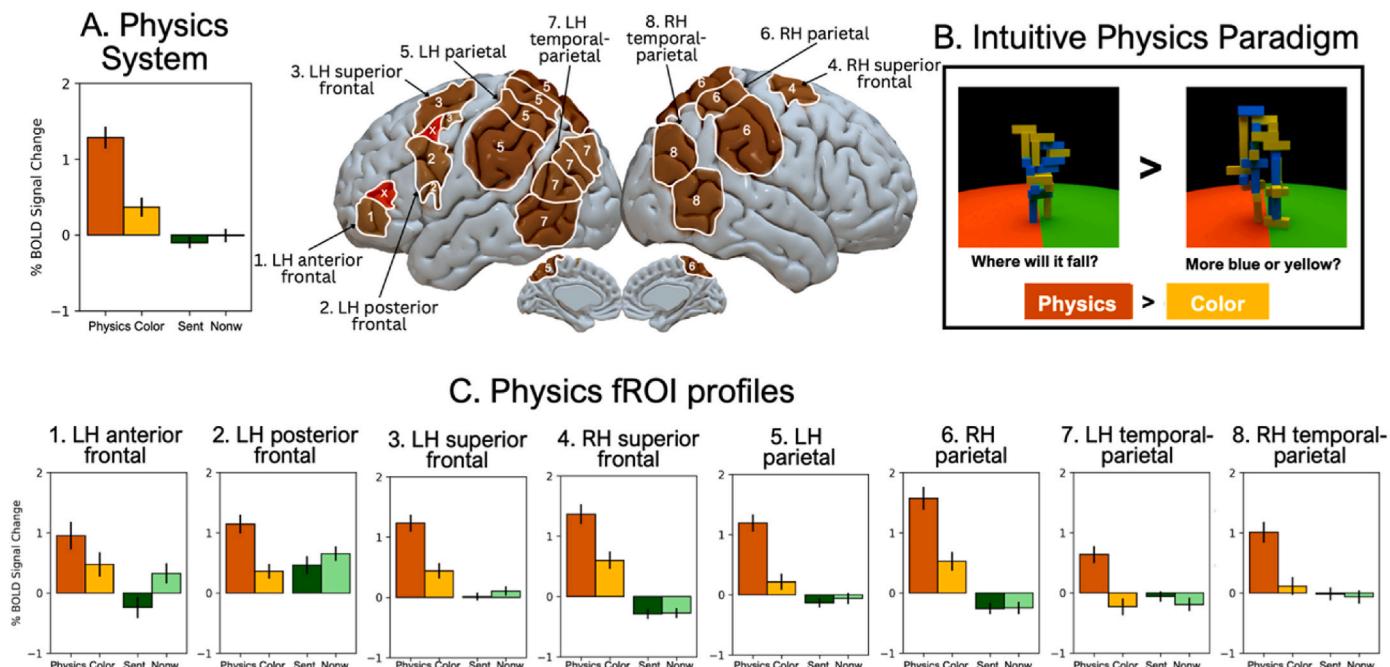
In addition to examining the relationship of the physical-reasoning system to the MD system and the language system, we examined its relationship to another high-level reasoning system: the social reasoning or ‘Theory of Mind’ network (e.g., Saxe and Kanwisher, 2003); some components of this system fall in broadly similar areas as the physical-reasoning brain areas (Supp. Fig. 1). For the 21 participants who completed a Theory of Mind localizer task (Saxe and Kanwisher, 2003; Dodell-Feder et al., 2011), we performed similar overlap analyses as for the other systems and found almost no overlap (Supp. Fig. 1; Supp. Tables 4A and B; see Supp. Table 5 for the concordant results from the Dice overlap analysis).

## 5. Discussion

We here examined the physical-reasoning system, which was

originally identified by Fischer et al. (2016; see also Schwettmann et al., 2019; Pramod et al., 2022), and its relationship with other known cognitive systems in an effort to illuminate the representations that may mediate intuitive physical reasoning. We replicated Fischer et al.’s original findings of a set of frontal and parietal brain areas that respond more strongly during a physical reasoning task (making physical stability judgments about block towers) than a difficulty-matched color-judgment task on the same stimuli, although we found an additional area in the anterior left frontal lobe. We found that the physical-reasoning system overlaps with the domain-general Multiple Demand (MD) system (Duncan, 2010), which has been implicated in executive control and in some forms of reasoning (e.g., Duncan and Owen, 2000; Fox et al., 2005; Hampshire et al., 2011; Niendam et al., 2012; Fedorenko et al., 2013; Hearne et al., 2017; Amalric and Dehaene, 2019; Woolgar et al., 2018; Assem et al., 2020a,b; Ivanova et al., 2020; Liu et al., 2020). However, in line with an analysis reported in Fischer et al. (2016), we found a dissociation in the fine-grained patterns of activation (see Pramod et al., in prep. for a more in-depth exploration of the relationship between the physical-reasoning system and the MD system). Moreover, the newly discovered left anterior frontal area shows clear selectivity for physical reasoning relative to the spatial working memory task in its univariate response profile. Critically, we found that the physical-reasoning system does not overlap with the language-selective system: the response in the language areas during physical reasoning is close to the low-level baseline, and the physical-reasoning areas show a stronger response to the nonword-list condition compared to the sentence condition—the opposite of the response in the language system. Below, we discuss the implications of our findings for understanding physical reasoning and the general structure of human cognition.

First, this work sheds light on the mental representations that undergird our processing of the physical world around us. Here, we replicate a past finding that the physical-reasoning system is at least



**Fig. 4. The physical-reasoning system and its relationship to language processing.** A. The response in the physical-reasoning system fROIs to the physical-reasoning localizer and language localizer conditions (Physics, Color, Sentences, Nonwords), averaged across the fROIs; and a display of the physical-reasoning system parcels (excluded parcels 20 and 21 shown in red; see Methods). B. The language localizer. During Sentences trials, participants viewed complete sentences one word at a time. During Nonwords trials, participants viewed nonword lists (see Methods for details). C. The response to the physical-reasoning localizer and language localizer conditions (Physics, Color, Sentences, Nonwords), broken down by individual fROI group. We observe a strong Physics > Color effect in all fROI groups (Table 1A) and no positive Sentences > Nonwords effect in any fROI group (Table 1B). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

partially dissociable from the domain-general MD system, which rules out the possibility that we represent the physical world using the kind of abstract representations that enable diverse goal-directed behaviors, novel problem solving, and mathematical reasoning (Duncan, 2010; Duncan et al., 2020; Amalric and Dehaene, 2019; Woolgar et al., 2018; Ivanova et al., 2020; Liu et al., 2020). The lack of overlap with the language network further suggests that our representations of the physical world are not linguistic in nature, in spite of the fact that they may be symbolic (e.g., Battaglia et al., 2013; Smith and Vul, 2013). The physical-reasoning system may therefore be a unique system that shares perceptual grounding with e.g., high-level visual areas (such as the object perception area in the lateral occipital cortex; e.g., Grill-Spector et al., 2001) but, at the same time, shows selectivity for particular types of abstract content (e.g., representations of an object's mass or stability; Schwettmann et al., 2019; Pramod et al., 2022), which is characteristic of high-level systems of reasoning, such as the Theory of Mind system (Castelli et al., 2000; Ruby and Decety, 2003; Saxe and Kanwisher, 2003) or the MD system (Duncan, 2010). An interesting question to explore in future work is whether the components of the physical-reasoning system that are located in closer proximity to the MD system encode more abstract features of physical events. Based on the spatial proximity to—and partial overlap with—the MD system, one could also speculate that the two systems were one and the same earlier in the mammalian evolutionary history, as has been hypothesized for some spatially adjacent large-scale networks (e.g., DiNicola et al., 2022; Deen and Freiwald, 2022); in particular, perhaps the MD system split off from the physical-reasoning system with the expansion of the association cortex (Buckner and Krienen, 2013), which allowed for a greater degree of abstraction beyond the representations of the local physical environment.

Second, this study adds to the growing body of evidence suggesting that the language system is highly specialized for linguistic computations and does not support non-linguistic cognition. The language areas are not engaged when individuals perform diverse forms of reasoning (e.g., Monti et al., 2009, 2012; Fedorenko et al., 2011; Ivanova et al., 2020), and some individuals with severe linguistic deficits (aphasia) retain their ability to think abstractly (Varley and Siegal, 2000; Varley et al., 2005; for reviews see Fedorenko and Varley, 2016; Fedorenko et al., 2024a,b). Our study adds intuitive physical reasoning to the list of diverse types of reasoning that do not engage linguistic-processing mechanisms. The findings also reinforce the conclusions drawn from prior research on individuals with aphasia, who appear to retain the ability to reason about the physical world (Varley and Siegal, 2000).

Third, our study contributes to the understanding of the ontology of cognition/thought. Prior research has identified a specialized system for thinking about others' mental states—sometimes referred to as theory of mind (ToM) or mentalizing (e.g., Castelli et al., 2000; Ruby and Decety, 2003; Saxe and Kanwisher, 2003). Fischer et al. (2016) reported a system for intuitive physical reasoning—a finding that we replicate. By demonstrating that the physical-reasoning system is distinct from the ToM system and at least partially, from the domain-general MD system (see also Pramod et al., in prep.), we provide additional evidence that human cognition relies on multiple specialized systems rather than a single, general-purpose reasoning system. This distinction is further supported by behavioral individual-differences studies. For example, Mitko and Fischer (2020) observed a dissociation between performance on intuitive physical reasoning tasks and tasks tapping spatial cognition. These dissociations beg the question of what other kinds of reasoning may be supported by specialized systems. Developmental work on core knowledge systems (Spelke and Kinzler, 2007) can provide inspiration for additional domain-specific reasoning systems built out of early-emerging conceptual primitives, although at least some of those systems have already been shown to not draw on specialized neural resources (e.g., the numerical reasoning system, which appears to rely on the MD system; Shuman and Kanwisher, 2004; Fedorenko et al., 2013; Niendam et al., 2012).

Finally, our findings highlight the functional heterogeneity of the left frontal lobe. Contra unified accounts of frontal lobe function (e.g., Cohen et al., 1996; Miller and Cohen, 2001; Duncan et al., 1996, 2020; Dosenbach et al., 2006; Cole and Schneider, 2007), evidence exists of both structural (e.g., Amunts et al., 2010) and functional (e.g., Stuss and Knight, 2013) distinctions among nearby areas in the frontal cortex. With respect to functional dissociations, the left frontal lobe in humans has been shown to house components of the language and the MD systems (Fedorenko et al., 2011, 2013; Braga et al., 2020; see Fedorenko and Blank, 2020 for a review), the articulatory motor-planning area (Hillis et al., 2004; Flinker et al., 2015; Long et al., 2016; Basilakos et al., 2018; Wolna et al., 2024)—the area originally discovered by Broca; Broca (1861), components of the Theory of Mind network and the episodic default network (e.g., Saxe and Kanwisher, 2003; Braga et al., 2020; DiNicola et al., 2022; Du et al., 2024), and perhaps areas specialized for aspects of logical reasoning (Coetze and Monti, 2018; Kean et al., 2024) and causal reasoning (Pramod et al., 2024). Fischer et al. (2016) further established the existence of physical-reasoning areas in the frontal cortex that are at least partially dissociable from the MD system. We replicate these findings and identify an additional component of the physical-reasoning system in the anterior left frontal lobe. As new selectivities continue to emerge, understanding the organizing principles of the frontal cortex becomes increasingly important. Although it remains possible that some computations are shared among all of these distinct areas, any account that spans these established functional boundaries would need to explain this heterogeneity of response and functional connectivity profiles (see Xu et al., 2022 for evidence that even in non-human primates, the frontal lobes are highly functionally heterogeneous, with different areas exhibiting distinct patterns of connections to posterior brain areas; and see Mansouri et al., 2006, 2015, 2017, 2020, 2022 for evidence that lesions to different frontal areas in non-human primates lead to distinct kinds of behavioral deficits).

Our study is limited in several ways. We highlight three key limitations. *First*, our conclusions pertain to the kind of intuitive reasoning required for making a judgment about a block tower's center of gravity. Whether other kinds of physical reasoning—such as more complex intuitive reasoning necessary for predicting fluid motion (Bates et al., 2019), evaluating the strength of knots and tangles (Croom and Firestone, 2024), or more abstract physical reasoning, which requires drawing on knowledge of laws of physics—rely on the same brain regions as the type of reasoning examined here remains to be discovered. We cannot rule out the possibility that these other types of physical reasoning would more heavily engage domain-general reasoning resources supported by the MD network or require linguistic representations. *Second*, in our key analyses, we treated the physical-reasoning network as an integrated functional system: although we report the results for different components of the system separately, more work is needed to understand whether different brain areas comprising this network support different computations. For example, the physical-reasoning brain region that we identified in the anterior portion of the left frontal lobe appears to differ from the other physical-reasoning regions in that it responds very little to the demanding spatial working memory task. Does this more selective response to the physical-reasoning contrast imply a distinct contribution to physical reasoning (for example, encoding more abstract properties of physical events)? Additional experiments examining responses to diverse physical reasoning tasks will be needed to understand the computations implemented in each of the physical-reasoning system's components. And *third*, with respect to the overlap with language-processing mechanisms, we have here focused on the core language areas in the left frontal and temporal lobes. However, a number of other brain areas respond reliably to language across modalities, including the homotopic areas in the right hemisphere, medial frontal and parietal areas, subcortical areas, and areas in the cerebellum (e.g., Lindell, 2006; Hasson et al., 2007; Covington and Duff, 2016;

Klostermann and Ehlen, 2013; LeBel and D'Mello, 2023). At least some of these areas show less selective responses to language compared to the core language areas (Wolna et al., in prep.; Casto et al., in prep.). Whether any part(s) of the extended language network contribute to physical reasoning, and how, remains to be established.

Overall, our study contributes to the understanding of the neural architecture underlying high-level cognitive functions. We provide evidence that intuitive physical reasoning—as tapped by a paradigm introduced in Fischer et al. (2016)—is mediated by specialized, domain-specific representations that are distinct from both linguistic and domain-general abstract representations. The existence of specialized systems for different types of reasoning suggests that the modular nature of the human mind and brain extends beyond the perceptual domain (e.g., Kanwisher, 2010) and highlights the need for further research into the ontology of human thought. Once the key components of the mind have been identified, we can begin tackling exciting questions pertaining to how these different systems work together to enable complex thought and behavior, including in more naturalistic settings.

### CRediT authorship contribution statement

**Hope H. Kean:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alexander Fung:** Writing – review & editing, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation. **R.T. Pramod:** Writing – review & editing, Resources, Investigation, Data curation. **Jessica Chomik-Morales:** Writing – review & editing, Resources, Investigation, Data curation. **Nancy Kanwisher:** Writing – review & editing, Supervision, Methodology. **Evelina Fedorenko:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization.

### Acknowledgments

We would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, including the technical team—Steve Shannon and Atsushi Takahashi. We would also like to thank Kevin Smith, Anya Ivanova, Aryan Zoroufi, Josh Tenenbaum, Steve Piantadosi, and Dorothy Kean for invaluable comments and discussions on the manuscript, and Moshe Poliak for help with the statistical analyses. HK was supported by a graduate fellowship from the K. Lisa Yang Integrative Computational Neuroscience (ICoN) Center. EF was supported by research funds from the McGovern Institute for Brain Research, the Department of Brain and Cognitive Sciences, the Quest Initiative, and a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuropsychologia.2025.109125>.

### Data availability

Data will be made available on request.

### References

- Amalric, M., Dehaene, S., 2019. A distinct cortical network for mathematical knowledge in the human brain. *Neuroimage* 189, 19–31. <https://doi.org/10.1016/j.neuroimage.2019.01.001>.
- Amunts, K., Lenzen, M., Friederici, A.D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., Zilles, K., 2010. Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8, e1000489. <https://doi.org/10.1371/journal.pbio.1000489>.
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26, 839–851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>.
- Assem, M., Blank, I.A., Mineroff, Z., Ademoglu, A., Fedorenko, E., 2020a. Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex; a journal devoted to the study of the nervous system and Behav.* 131, 1. <https://doi.org/10.1016/j.cortex.2020.06.013>.
- Assem, M., Glasser, M.F., Van Essen, D.C., Duncan, J., 2020b. A domain-general cognitive core defined in multimodally parcellated human cortex. *Cerebr. Cortex* 30, 4361–4380. <https://doi.org/10.1093/cercor/bhaa023>.
- Baillargeon, R., Needham, A., Devos, J., 1992. The development of young infants' intuitions about support. *Early Dev. Parent.* 1, 69–78. <https://doi.org/10.1002/edp.2430010203>.
- Baillargeon, R., Spelke, E.S., Wasserman, S., 1985. Object permanence in five-month-old infants. *Cognition* 20, 191–208. [https://doi.org/10.1016/0010-0277\(85\)90008-3](https://doi.org/10.1016/0010-0277(85)90008-3).
- Basilakos, A., Smith, K.G., Fillmore, P., Fridriksson, J., Fedorenko, E., 2018. Functional characterization of the human speech articulation network. *Cerebr. Cortex* 28, 1816–1830. <https://doi.org/10.1093/cercor/bhx100>.
- Bates, C.J., Yıldırım, I., Tenenbaum, J.B., Battaglia, P., 2019. Modeling human intuitions about liquid flow with particle-based simulation. *PLoS Comput. Biol.* 15, e1007210. <https://doi.org/10.1371/journal.pcbi.1007210>.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using *lme4*. *J. Stat. Software* 67. <https://doi.org/10.18637/jss.v067.i01>.
- Battaglia, P.W., Hamrick, J.B., Tenenbaum, J.B., 2013. Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci. USA* 110, 18327–18332. <https://doi.org/10.1073/pnas.1306572110>.
- Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. <https://doi.org/10.1214/aos/1013699998>.
- Berthier, N.E., DeBlois, S., Poirier, C.R., Novak, M.A., Clifton, R.K., 2000. Where's the ball? Two- and three-year-olds reason about unseen events. *Dev. Psychol.* 36, 394–401. <https://doi.org/10.1037/0012-1649.36.3.394>.
- Blank, I., Balewski, Z., Mahowald, K., Fedorenko, E., 2016. Syntactic processing is distributed across the language system. *Neuroimage* 127, 307–323. <https://doi.org/10.1016/j.neuroimage.2015.11.069>.
- Blank, I., Kanwisher, N., Fedorenko, E., 2014. A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J. Neurophysiol.* 112, 1105–1118. <https://doi.org/10.1152/jn.00884.2013>.
- Bozic, M., Fonteneau, E., Su, L., Marslen-Wilson, W.D., 2014. Grammatical analysis as a distributed neurobiological function. *Hum. Brain Mapp.* 36, 1190. <https://doi.org/10.1002/hbm.22696>.
- Braga, R.M., DiNicola, L.M., Becker, H.C., Buckner, R.L., 2020. Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *J. Neurophysiol.* 124, 1415–1448. <https://doi.org/10.1152/jn.00753.2019>.
- Broca, M.P., 1861. Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin et mémoires de la Société Anatomique de Paris* 6, 330–357.
- Buckner, R.L., Krienen, F.M., 2013. The evolution of distributed association networks in the human brain. *Trends Cognit. Sci.* 17, 648–665. <https://doi.org/10.1016/j.tics.2013.09.017>.
- Castelli, F., Happé, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325. <https://doi.org/10.1006/nimg.2000.0612>.
- Casto, C., Small, H., Poliak, M., Tuckte, G., Lipkin, B., D'Mello, A., Fedorenko, E. (in preparation). The Cerebellar Components of the Human Language Network.
- Chen, X., Affourtit, J., Ryskin, R., Regev, T.I., Norman-Haignere, S., Jouravlev, O., Malik-Moraleda, S., Kean, H., Varley, R., Fedorenko, E., 2023. The human language system, including its inferior frontal component in "Broca's area," does not support music perception. *Cerebr. Cortex* 33, 7904–7929. <https://doi.org/10.1093/cercor/bhd087>.
- Coetze, J.P., Monti, M.M., 2018. At the core of reasoning: dissociating deductive and non-deductive load. *Hum. Brain Mapp.* 39, 1850–1861. <https://doi.org/10.1002/hbm.23979>.
- Cohen, J.D., Braver, T.S., O'Reilly, R.C., 1996. A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 351, 1515–1527. <https://doi.org/10.1098/rstb.1996.0138>.
- Cole, M.W., Schneider, W., 2007. The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage* 37, 343–360. <https://doi.org/10.1016/j.neuroimage.2007.03.071>.
- Covington, N.V., Duff, M.C., 2016. Expanding the language network: direct contributions from the hippocampus. *Trends Cognit. Sci.* 20, 869–870. <https://doi.org/10.1016/j.tics.2016.10.006>.
- Croome, S., Firestone, C., 2024. Tangled physics: knots strain intuitive physical reasoning. *Open Mind* 8, 1170–1190. [https://doi.org/10.1162/opmi\\_a\\_00159](https://doi.org/10.1162/opmi_a_00159).
- Deen, B., Freiwald, W.A., 2022. Parallel systems for social and spatial reasoning within the cortical apex. <https://doi.org/10.1101/2021.09.23.461550>.
- Diachek, E., Blank, I., Siegelman, M., Affourtit, J., Fedorenko, E., 2020. The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.2036-19.2020>.
- Dice, L.R., 1945. Measures of the amount of ecological association between species. *Ecology* 26, 297–302. <https://doi.org/10.2307/1932409>.
- DiNicola, L.M., Sun, W., Buckner, R.L., 2022. Side-by-side regions in dorsolateral prefrontal cortex estimated within the individual respond differentially to domain-

- specific and domain-flexible processes. *J. Neurophysiol.* 130, 1602–1615. <https://doi.org/10.1152/jn.00277.2023>.
- Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. *Neuroimage* 55, 705–712. <https://doi.org/10.1016/j.neuroimage.2010.12.040>.
- Dosenbach, N.U.F., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., Petersen, S.E., 2006. A core system for the implementation of task sets. *Neuron* 50, 799–812. <https://doi.org/10.1016/j.neuron.2006.04.031>.
- Du, J., DiNicola, L.M., Angeli, P.A., Saadon-Grosman, N., Sun, W., Kaiser, S., Ladopoulou, J., Xue, A., Yeo, B.T.T., Eldaief, M.C., Buckner, R.L., 2024. Organization of the human cerebral cortex estimated within individuals: networks, global topography, and function. *J. Neurophysiol.* 131, 1014–1082. <https://doi.org/10.1152/jn.00308.2023>.
- Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cognit. Sci.* 14, 172–179. <https://doi.org/10.1016/j.tics.2010.01.004>.
- Duncan, J., Assem, M., Shashidhara, S., 2020. Integrated intelligence from distributed brain activity. *Trends Cognit. Sci.* 24, 838–852. <https://doi.org/10.1016/j.tics.2020.06.012>.
- Duncan, J., Emslie, H., Williams, P., Johnson, R., Freer, C., 1996. Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cogn. Psychol.* 30, 257–303. <https://doi.org/10.1006/cogp.1996.0008>.
- Duncan, J., Owen, A.M., 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483. [https://doi.org/10.1016/S0166-2236\(00\)01633-7](https://doi.org/10.1016/S0166-2236(00)01633-7).
- Eberly, D.H., 2010. *Game Physics, second ed.* Chapman and Hall/CRC, Saint Louis.
- Fedorenko, E., Piantadosi, S.T., Gibson, E.A.F., 2024b. Language is primarily a tool for communication rather than thought. *Nature* 630, 575–586. <https://doi.org/10.1038/s41586-024-07522-w>.
- Fedorenko, E., Ivanova, A.A., Regev, T.I., 2024a. The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* 25, 289–312. <https://doi.org/10.1038/s41583-024-00802-4>.
- Fedorenko, E., Blank, I., Siegelman, M., Mineroff, Z., 2020. Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition* 203, 104348. <https://doi.org/10.1016/j.cognition.2020.104348>.
- Fedorenko, E., Blank, I.A., 2020. Broca's area is not a natural kind. *Trends Cognit. Sci.* 24, 270–284. <https://doi.org/10.1016/j.tics.2020.01.001>.
- Fedorenko, E., Varley, R., 2016. Language and thought are not the same thing: evidence from neuroimaging and neurological patients: language versus thought. *Ann. N. Y. Acad. Sci.* 1369, 132–153. <https://doi.org/10.1111/nyas.13046>.
- Fedorenko, E., Scott, T.L., Brunner, P., Coon, W.G., Pritchett, B., Schalk, G., Kanwisher, N., 2016. Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci.* 113, E6256–E6262. <https://doi.org/10.1073/pnas.1612132113>.
- Fedorenko, E., Duncan, J., Kanwisher, N., 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16616–16621. <https://doi.org/10.1073/pnas.1315235110>.
- Fedorenko, E., Nieto-Castañón, A., Kanwisher, N., 2012. Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, Multivoxel pattern analysis and cognitive theories 50 (4), 499–513. <https://doi.org/10.1016/j.neuropsychologia.2011.09.014>. March 1.
- Fedorenko, E., Behr, M.K., Kanwisher, N., 2011. Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16428–16433. <https://doi.org/10.1073/pnas.1112937108>.
- Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104, 1177–1194. <https://doi.org/10.1152/jn.00032.2010>.
- Fischer, J., Mahon, B.Z., 2021. What tool representation, intuitive physics, and action have in common: the brain's first-person physics engine. *Cogn. Neuropsychol.* 38, 455–467. <https://doi.org/10.1080/02643294.2022.2106126>.
- Fischer, J., Mikhail, J.G., Tenenbaum, J.B., Kanwisher, N., 2016. Functional neuroanatomy of intuitive physical inference. *Proc. Natl. Acad. Sci. U.S.A.* 113. <https://doi.org/10.1073/pnas.1610344113>.
- Flinker, A., Korzeniewska, A., Shestyuk, A.Y., Franaszczuk, P.J., Dronkers, N.F., Knight, R.T., Crone, N.E., 2015. Redefining the role of Broca's area in speech. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2871–2875. <https://doi.org/10.1073/pnas.1414491112>.
- Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. <https://doi.org/10.1073/pnas.0504136102>.
- Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995. Analysis of fMRI time-series revisited. *Neuroimage* 2, 45–53. <https://doi.org/10.1006/nimg.1995.1007>.
- Gleitman, L., 1990. The structural sources of verb meanings. *Language acquisition*. A. J. Develop. Linguistics. 1, 3–55. [https://doi.org/10.1207/s15327817la01\\_2](https://doi.org/10.1207/s15327817la01_2).
- Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vis. Res.* 41, 1409–1422. [https://doi.org/10.1016/s0042-6989\(01\)00073-6](https://doi.org/10.1016/s0042-6989(01)00073-6).
- Grønn, A., Stechow, A.V., 2016. Tense. In: Aloni, M., Dekker, P. (Eds.), *The Cambridge Handbook of Formal Semantics*. Cambridge University Press, pp. 313–341. <https://doi.org/10.1017/CBO9781139236157.012>.
- Gurnee, W., Tegmark, M., 2024. Language models represent space and time. <https://doi.org/10.48550/arXiv.2310.02207>.
- Hampshire, A., Highfield, R.R., Parkin, B.L., Owen, A.M., 2011. Fractionating human intelligence. *Neuron* 76, 1225–1237. <https://doi.org/10.1016/j.neuron.2012.06.022>.
- Hasson, U., Nusbaum, H.C., Small, S.L., 2007. Brain networks subserving the extraction of sentence information and its encoding to memory. *Cerebr. Cortex* 17, 2899–2913. <https://doi.org/10.1093/cercor/bhm016>.
- Hearne, L.J., Cocchi, L., Zalesky, A., Mattingley, J.B., 2017. Reconfiguration of brain network architectures between resting-state and complexity-dependent cognitive reasoning. *J. Neurosci.* 37, 8399–8411. <https://doi.org/10.1523/JNEUROSCI.0485-17.2017>.
- Hillis, A.E., Work, M., Barker, P.B., Jacobs, M.A., Breese, E.L., Maurer, K., 2004. Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* 127, 1479–1487. <https://doi.org/10.1093/brain/awh172>.
- Hood, B., Carey, S., Prasada, S., 2000. Predicting the outcomes of physical events: two-year-olds fail to reveal knowledge of solidity and support. *Child Dev.* 71, 1540–1554. <https://doi.org/10.1111/1467-8624.00247>.
- Hood, B., Cole-Davies, V., Dias, M., 2003. Looking and search measures of object knowledge in preschool children. *Dev. Psychol.* 39, 61–70.
- Hu, J., Small, H., Kean, H., Takahashi, A., Zekelman, L., Kleinman, D., Ryan, E., Nieto-Castañón, A., Ferreira, V., Fedorenko, E., 2022. Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebr. Cortex* 33, 4384–4404. <https://doi.org/10.1093/cercor/bhac350>.
- Hugdahl, K., Raichle, M.E., Mitra, A., Specht, K., 2015. On the existence of a generalized non-specific task-dependent network. *Front. Hum. Neurosci.* 9. <https://doi.org/10.3389/fnhum.2015.00430>.
- Ivanova, A.A., Srikant, S., Sueoka, Y., Kean, H.H., Dhamala, R., O'Reilly, U.-M., Bers, M. U., Fedorenko, E., 2020. Comprehension of computer code relies primarily on domain-general executive brain regions. *Elife* 9, e58906. <https://doi.org/10.7554/eLife.58906>.
- Julian, J.B., Fedorenko, E., Webster, J., Kanwisher, N., 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60, 2357–2364. <https://doi.org/10.1016/j.neuroimage.2012.02.055>.
- Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11163–11170. <https://doi.org/10.1073/pnas.1005062107>.
- Kean, H., Fung, A., Rule, J., Tenenbaum, J., Piantadosi, S., Fedorenko, E., 2024. Deduction and induction dissociate in the human brain. *Comp. Cogn. Neurosci. CCN*.
- Keen, R., 2003. Representation of objects and events: why do infants look So smart and toddlers look So dumb? *Curr. Dir. Psychol. Sci.* 12, 79–83. <https://doi.org/10.1111/1467-8721.01234>.
- Kleiman, E.M., Turner, B.J., Fedor, S., Beale, E.E., Huffman, J.C., Nock, M.K., 2017. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. *J. Abnorm. Psychol.* 126, 726–738. <https://doi.org/10.1037/abn0000273>.
- Klostermann, F., Ehlen, F., 2013. Functional roles of the thalamus for language capacities. *Front. Syst. Neurosci.* 7. <https://doi.org/10.3389/fnsys.2013.00032>.
- Kowalski, R., Sergot, M., 1986. A logic-based calculus of events. *New Generat. Comput.* 4, 67–95. <https://doi.org/10.1007/BF03037383>.
- Kriegeskorte, N., 2011. Pattern-information analysis: from stimulus decoding to computational-model testing. *NeuroImage*, Multivariate Decoding and Brain Reading 56, 411–421. <https://doi.org/10.1016/j.neuroimage.2011.01.061>.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Software* 82. <https://doi.org/10.18637/jss.v082.i13>.
- LeBel, A., D'Mello, A.M., 2023. A seat at the (language) table: incorporating the cerebellum into frameworks for language processing. *Curr. Opinion. Behav. Sci.* 53, 101310. <https://doi.org/10.1016/j.cobeha.2023.101310>.
- Leslie, A.M., Keeble, S., 1987. Do six-month-old Infants Perceive Causality?, vol. 63, pp. 1646–1669. <https://doi.org/10.1016/j.neuroimage.2012.06.065>, 3.
- Li, L., Xu, J., Dong, Q., Zheng, C., Liu, Q., Kong, L., Sun, X., 2023. Can language models understand physical concepts? <https://doi.org/10.48550/arXiv.2305.14057>.
- Lindell, A.K., 2006. In your right mind: right hemisphere contributions to language processing and production. *Neuropsychol. Rev.* 16, 131–148. <https://doi.org/10.1007/s11065-006-9011-9>.
- Lipkin, B., Tuckett, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocewicz, L., Pritchett, B., Siegelman, M., Hoeflin, C., Pongos, A., Blank, I.A., Struhl, M.K., Ivanova, A., Shannon, S., Sathe, A., Hoffmann, M., Nieto-Castañón, A., Fedorenko, E., 2022. Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Sci. Data* 9, 529. <https://doi.org/10.1038/s41597-022-01645-3>.
- Liu, Y.-F., Kim, J., Wilson, C., Bedny, M., 2020. Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. *Elife* 9, e59340. <https://doi.org/10.7554/eLife.59340>.
- Long, M.A., Katlowitz, K.A., Svirsky, M.A., Clary, R.C., Byun, T.M., Majaj, N., Oya, H., Matthew A Howard, I.I.I., Greenlee, J.D., 2016. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* 89, 1187. <https://doi.org/10.1016/j.neuron.2016.01.032>.
- Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., Fedorenko, E., 2022. An investigation across 45 languages and 12 language families reveals a universal language network. *Nat. Neurosci.* 25, 1014–1019. <https://doi.org/10.1038/s41593-022-01114-5>.
- Mansouri, F.A., Buckley, M.J., Fehring, D.J., Tanaka, K., 2020. The role of primate prefrontal cortex in bias and shift between visual dimensions. *Cerebr. Cortex* 30, 85–99. <https://doi.org/10.1093/cercor/bhz072>.

- Mansouri, F.A., Buckley, M.J., Mahboubi, M., Tanaka, K., 2015. Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices. *Proc. Natl. Acad. Sci.* 112, E3940–E3949. <https://doi.org/10.1073/pnas.1422629112>.
- Mansouri, F.A., Buckley, M.J., Tanaka, K., 2022. The neural substrate and underlying mechanisms of executive control fluctuations in Primates. *Prog. Neurobiol.* 209, 102216. <https://doi.org/10.1016/j.pneurobio.2022.102216>.
- Mansouri, F.A., Koehlin, E., Rosa, M.G.P., Buckley, M.J., 2017. Managing competing goals - a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* 18, 645–657. <https://doi.org/10.1038/nrn.2017.111>.
- Mansouri, F.A., Matsumoto, K., Tanaka, K., 2006. Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin card sorting test analog. *J. Neurosci.* 26, 2745–2756. <https://doi.org/10.1523/JNEUROSCI.5238-05.2006>.
- Marks, S., Tegmark, M., 2024. The geometry of truth: emergent linear structure in large language model representations of true/false datasets. <https://doi.org/10.48550/0/arXiv.2310.06824>.
- McCarthy, J., Hayes, P.J., 1969. Some philosophical problems from the standpoint of artificial intelligence. *Mach. Intell.* 4, 463–502.
- Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>.
- Mitko, A., Fischer, J., 2020. When it all falls down: the relationship between intuitive physics and spatial cognition. *Cognitive Research: Principles and Implications* 5, 24. <https://doi.org/10.1186/s41235-020-00224-7>.
- Mitko, A., Navarro-Cebrián, A., Cormiea, S., Fischer, J., 2024. A dedicated mental resource for intuitive physics. *iScience* 27, 108607. <https://doi.org/10.1016/j.isci.2023.108607>.
- Moens, M., 1987. *Tense, Aspect and Temporal Reference*.
- Moens, M., Steedman, M., 1988. Temporal ontology and temporal reference. *Comput. Linguist.* 14, 15–28.
- Monti, M.M., Parsons, L.M., Osherson, D.N., 2012. Thought beyond language: neural dissociation of algebra and natural language. *Psychol. Sci.* 23, 914–922. <https://doi.org/10.1177/0956797612437427>.
- Monti, M.M., Parsons, L.M., Osherson, D.N., 2009. The boundaries of language and thought through deductive inference. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12554–12559. <https://doi.org/10.1073/pnas.0902422106>.
- Nanda, N., Lee, A., Wattenberg, M., 2023. Emergent linear representations in world models of self-supervised sequence models. <https://doi.org/10.48550/arXiv.2309.09041>.
- Niendam, T.A., Laird, A.R., Ray, K.L., Dean, Y.M., Glahn, D.C., Carter, C.S., 2012. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognit. Affect. Behav. Neurosci.* 12, 241–268. <https://doi.org/10.3758/s13415-011-0083-5>.
- Nieto-Castañón, A., 2020. *Handbook of Functional Connectivity Magnetic Resonance Imaging Methods in CONN*. Hilbert Press. <https://doi.org/10.56441/hilbertpress.2207.6598>.
- Nieto-Castañón, A., Fedorenko, E., 2012. Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63, 1646–1669. <https://doi.org/10.1016/j.neuroimage.2012.06.065>.
- Partee, B.H., 1984. Nominal and temporal anaphora. *Ling. Philos.* 7, 243–286. <https://doi.org/10.1007/bf00627707>.
- Partee, B.H., 1973. Some structural analogies between tenses and pronouns in English. *J. Philos.* 70, 601–609. <https://doi.org/10.2307/2025024>.
- Pinto, J., Reiter, R., 1993. Temporal reasoning in logic programming: a case for the situation calculus. <https://doi.org/10.7551/mitpress/4305.003.0023>.
- Pramod, R., Cohen, M.A., Tenenbaum, J.B., Kanwisher, N., 2022. Invariant representation of physical stability in the human brain. *Elife* 11, e71736. <https://doi.org/10.7554/elife.71736>.
- Pramod, R.T., Chomik, Jessica, Schulz, Laura, Kanwisher, Nancy, 2024. A region in human left prefrontal cortex selectively engaged in causal reasoning. Proceedings of the Annual Meeting of the Cognitive Science Society 46, 0. <https://escholarship.org/uc/item/6ms537c4>.
- Pramod, R., Hutchinson, Kanwisher, N., in prep. Intuitive Physical Reasoning and Multiple Demand Systems Comprise Dissociable Cortical Networks.
- Pulman, S.G., 1997. Aspectual shift as type coercion. *Trans. Philol. Soc.* 95, 279–317. <https://doi.org/10.1111/1467-968X.00020>.
- Rajimehr, R., Xu, H., Farahani, A., Kornblith, S., Duncan, J., Desimone, R., 2024. Functional architecture of cerebral cortex during naturalistic movie watching. *Neuron* 112, 4130–4146.e3. <https://doi.org/10.1016/j.neuron.2024.10.005>.
- Regev, T.I., Casto, C., Hosseini, E.A., Adamek, M., Ritaccio, A.L., Willie, J.T., Brunner, P., Fedorenko, E., 2024. Neural populations in the language network differ in the size of their temporal receptive windows. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-024-01944-2>.
- Ruby, P., Decety, J., 2003. What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci.* 17, 2475–2480. <https://doi.org/10.1046/j.1460-9568.2003.02673.x>.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind.". *Neuroimage* 19, 1835–1842. [https://doi.org/10.1016/s1053-8119\(03\)00230-1](https://doi.org/10.1016/s1053-8119(03)00230-1).
- Schwettmann, S., Tenenbaum, J.B., Kanwisher, N., 2019. Invariant representations of mass in the human brain. *Elife* 8, e46619. <https://doi.org/10.7554/elife.46619>.
- Scott, T.L., Gallée, J., Fedorenko, E., 2017. A new fun and robust version of an fMRI locator for the frontotemporal language system. *Cognit. Neurosci.* 8, 167–176. <https://doi.org/10.1080/17588928.2016.1201466>.
- Shain, C., Blank, I.A., Fedorenko, E., Gibson, E., Schuler, W., 2022. Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *J. Neurosci.* 42, 7412–7430. <https://doi.org/10.1523/JNEUROSCI.1894-21.2022>.
- Shain, C., Kean, H., Casto, C., Lipkin, B., Affourtit, J., Siegelman, M., Mollica, F., Fedorenko, E., 2024. Distributed sensitivity to syntax and semantics throughout the language network. *J. Cognit. Neurosci.* 36, 1427–1471. [https://doi.org/10.1162/jocn\\_a\\_02164](https://doi.org/10.1162/jocn_a_02164).
- Shashidhara, S., Assem, M., Glasser, M.F., Duncan, J., 2024. Task and stimulus coding in the multiple-demand network. *Cerebr. Cortex* 34, bhae278. <https://doi.org/10.1093/cercor/bhae278>.
- Shashidhara, S., Erez, Y., 2021. Reward motivation increases univariate activity but has limited effect on coding of task-relevant information across the frontoparietal cortex. *Neuropsychologia* 160, 107981. <https://doi.org/10.1016/j.neuropsychologia.2021.107981>.
- Shashidhara, S., Mitchell, D.J., Erez, Y., Duncan, J., 2019. Progressive recruitment of the frontoparietal multiple-demand system with increased task complexity, time pressure, and reward. *J. Cognit. Neurosci.* 31, 1617–1630. [https://doi.org/10.1162/jocn\\_a\\_01440](https://doi.org/10.1162/jocn_a_01440).
- Shashidhara, S., Spronkers, F.S., Erez, Y., 2020. Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the "Multiple-demand" frontoparietal network. *J. Cognit. Neurosci.* 32, 1348–1368. [https://doi.org/10.1162/jocn\\_a\\_01554](https://doi.org/10.1162/jocn_a_01554).
- 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>.
- Skordos, D., Bunger, A., Richards, C., Selimis, S., Trueswell, J., Papafragou, A., 2020. Motion verbs and memory for motion events. *Cogn. Neuropsychol.* 37, 254–270. <https://doi.org/10.1080/02643294.2019.1685480>.
- Smith, K.A., Vul, E., 2013. Sources of uncertainty in intuitive physics. *Topics in Cognitive Science* 5, 185–199. <https://doi.org/10.1111/tops.12009>.
- Sørensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 5, 1–34.
- Spelke, E., 2022. *What Babies Know: Core Knowledge and Composition: Volume 1, Oxford Cognitive Development Series*. Oxford University Press, New York.
- Spelke, E.S., Breinlinger, K., Macomber, J., Jacobson, K., 1992. Origins of knowledge. *Psychol. Rev.* 99, 605–632. <https://doi.org/10.1037/0033-295X.99.4.605>.
- Spelke, E.S., Kinzler, K.D., 2007. Core knowledge. *Dev. Sci.* 10, 89–96. <https://doi.org/10.1111/j.1467-7687.2007.00569.x>.
- Stuss, D.T., Knight, R.T., 2013. *Principles of Frontal Lobe Function*. OUP, USA.
- Tuckute, G., Sathe, A., Srikanth, S., Taliaferro, M., Wang, M., Schrimpf, M., Kay, K., Fedorenko, E., 2024. Driving and suppressing the human language network using large language models. *Nat. Hum. Behav.* 8, 544–561. <https://doi.org/10.1038/s41562-023-01783-7>.
- Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science* 211, 453–458. <https://doi.org/10.1126/science.7455683>.
- Varley, R., 2002. Science without grammar: scientific reasoning in severe agrammatic aphasia. In: Carruthers, P., Stich, S., Siegal, M. (Eds.), *The Cognitive Basis of Science*. Cambridge University Press, pp. 99–116. <https://doi.org/10.1017/CBO9780511613517.006>.
- Varley, R., Siegal, M., 2000. Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Curr. Biol.* 10, 723–726. [https://doi.org/10.1016/S0960-9822\(00\)00538-8](https://doi.org/10.1016/S0960-9822(00)00538-8).
- Varley, R.A., Klessinger, N.J.C., Romanowski, C.A.J., Siegal, M., 2005. Agrammatic but numerate. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3519–3524. <https://doi.org/10.1073/pnas.0404747102>.
- Vurgun, U., Ji, Y., Papafragou, A., 2024. Aspectual processing shifts visual event apprehension. *Cogn. Sci.* 48, e13476. <https://doi.org/10.1111/cogs.13476>.
- Vurgun, U., Ji, Y., Papafragou, A., 2022. Linguistic aspect constrains event apprehension. *Proceedings of the Annual Meeting of the Cognitive Science Society* 44.
- Willems, R.M., Der Haegen, L.V., Fisher, S.E., Francks, C., 2014. On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat. Rev. Neurosci.* 15, 193–201. <https://doi.org/10.1038/nrn3679>.
- Wilson, S.M., Bautista, A., Yen, M., Lauderdale, S., Eriksson, D.K., 2017. Validity and reliability of four language mapping paradigms. *Neuroimage: Clinic* 16, 399–408. <https://doi.org/10.1016/j.nicl.2016.03.015>.
- Wolna, Agata, Wright Aaron, Fedorenko Evelina. (in preparation) The Extended Language Network.
- Wolna, A., Szewczyk, J., Diaz, M., Domagalik, A., Szwed, M., Wodniecka, Z., 2024. Tracking components of bilingual language control in speech production: an fMRI study using functional localizers. *Neurobiology of Language* 5, 315–340. [https://doi.org/10.1162/nol\\_a\\_00128](https://doi.org/10.1162/nol_a_00128).
- Wong, L., Grand, G., Lew, A.K., Goodman, N.D., Mansinghka, V.K., Andreas, J., Tenenbaum, J.B., 2023. From Word Models to World Models: Translating from Natural Language to the Probabilistic Language of Thought. <https://doi.org/10.48550/arXiv.2306.12672>.
- Woolgar, A., Duncan, J., Manes, F., Fedorenko, E., 2018. The multiple-demand system but not the language system supports fluid intelligence. *Nat. Hum. Behav.* 2, 200–204. <https://doi.org/10.1038/s41562-017-0282-3>.
- Xu, F., 2019. Towards a rational constructivist theory of cognitive development. *Psychol. Rev.* 126, 841–864. <https://doi.org/10.1037/rev0000153>.
- Xu, R., Bichot, N.P., Takahashi, A., Desimone, R., 2022. The cortical connectome of primate lateral prefrontal cortex. *Neuron* 110, 312–327.e7. <https://doi.org/10.1016/j.neuron.2021.10.018>.