Cognition 232 (2023) 105331

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Graded human sensitivity to geometric and topological concepts

Vijay Marupudi *, Sashank Varma

Georgia Institute of Technology, United States of America

ARTICLE INFO

Keywords:
Geometric reasoning
Mental rotation
Fluid intelligence
Mathematical achievement
Graded sensitivity
Individual differences

ABSTRACT

In a seminal study, Dehaene et al. (2006) found evidence that adults and children are sensitive to geometric and topological (GT) concepts using a novel odd-one-out task. However, performance on this task could reflect more general cognitive abilities than intuitive knowledge of GT concepts. Here, we developed a new 2-alternative forced choice (2-AFC) version of the original task where chance represents a higher bar to clear (50% vs. 16.67%) and where the role of general cognitive abilities is minimized. Replicating the original finding, American adult participants showed above-chance sensitivity to 41 of the 43 GT concepts tested. Moreover, their performance was not strongly driven by two general cognitive abilities, fluid intelligence and mental rotation, nor was it strongly associated with mathematical achievement as measured by ACT/SAT scores. The performance profile across the 43 concepts as measured by the new 2-AFC task was found to be highly correlated with the profiles as measured using the original odd-one-out task, as an analysis of data sets spanning populations and ages revealed. Most significantly, an aggregation of the 43 concepts into seven classes of GT concepts found evidence for *graded sensitivity*. Some classes, such as Euclidean geometry and Topology, were found to be more domain-specific: they "popped out" for participants and were judged very quickly and highly accurately. Others, notably Symmetry and Geometric transformations, were found to be more domain-general: better predicted by participants' general cognitive abilities and mathematical achievement. These results shed light on the graded nature of GT concepts in humans and challenge computational models that emphasize the role of induction.

1. Introduction

Understanding the spatial structure of the environment has survival value for animals (Vallortigara, 2017), and may be part of the core knowledge of human infants (Spelke & Kinzler, 2007). This understanding may be underpinned by sensitivity to geometric and topological (GT) concepts. This sensitivity may in turn be the foundation for the formal and abstract development of these topics by mathematicians, and may be the basis for how students come to understand these topics through formal schooling. In a pioneering study, Dehaene et al. (2006) found evidence for sensitivity to GT concepts in adults and children whether or not they had attended traditional schools. Izard and Spelke (2009) extended these findings to children as young as 3 to 6 years old located in New York and Boston who had not yet received substantial school-based instruction on GT concepts. Other researchers have used variants of the Dehaene et al. (2006) task to investigate sensitivity to different mathematical concepts in Senegal (van der Ham et al., 2017).

The results of Dehaene et al. (2006) have not gone unchallenged. A methodological concern is that chance on their task is quite low (16.7%), and thus above-chance performance is only weak evidence for sensitivity to these concepts. Another challenge to the claim that these concepts are intuitive comes from a computational model of the Dehaene et al. (2006) data by Lovett and Forbus (2011). This model utilizes representations that directly encode geometric and topological attributes such as straight line, curve, closure, and inside (i.e., containment), consistent with the claim that people are directly sensitive to these concepts. However, it also employs a domain-general, analogical generalization mechanism to extract other attributes such as rotation. In addition, it uses this domain-general mechanism to explain how representations are reasoned over when performing the Dehaene et al. (2006) task. Thus, the question of whether sensitivity to GT concepts is *sufficient* for explaining performance on this task, or whether domain-general abilities such as analogical generalization are also required, remains open.

The current study has first the goal of evaluating the Dehaene et al. (2006) proposal that people are sensitive to GT concepts. A sample of American adults performed a new 2-alternative forced choice (2-AFC) version of the original task that has several desirable properties. One is that chance on the new version is 50 percent, representing a much higher bar for participants to clear to demonstrate sensitivity to GT concepts. Another property is that it minimizes the potential role

* Corresponding author.
E-mail address: vijaymarupudi@gatech.edu (V. Marupudi).

https://doi.org/10.1016/j.cognition.2022.105331
Received 5 October 2021; Received in revised form 16 October 2022; Accepted 17 November 2022

of domain-general abilities such as induction and visuospatial reasoning in choosing the correct answer. This is important for evaluating the Lovett and Forbus (2011) model, which posits a central role for domain-general abilities in performing the Dehaene et al. (2006) task. To further address this point, we collected independent measures of two domain-general abilities, fluid intelligence and mental rotation, and directly evaluated the association between performance on these measures and sensitivity to GT concepts.

The current study also addressed other research goals. Given the importance of geometry in the secondary school curriculum, we evaluated whether sensitivity to GT concepts is associated with mathematical achievement more generally. In addition, because the new version of the task is speeded, we analyzed whether people are more sensitive to some GT concepts compared to others by checking whether these concepts "pop out" for participants as indicated by very fast reaction times and very high accuracies. Finally, we compared the performance profiles of the participants in the Dehaene et al. (2006) study, the Izard and Spelke (2009) study, and the current study to assess whether participants from different populations and of different ages find the same concepts relatively easy or difficult across the two versions of the task. Together, these analyses expand our understanding of human sensitivity to GT concepts.

1.1. Evidence for sensitivity to geometric and topological concepts

Discussion of intuitions regarding geometry and sensitivity to geometric concepts goes back to Plato (Hohol, 2019). In the Socratic dialog *Meno*, Plato describes a conversation between Socrates and an uneducated slave boy on how to double the area of a square. The boy initially suggests doubling the length of the sides, which is incorrect. Through a series of questions, Socrates guides the boy to the correct answer, which involves constructing a larger square using the diagonal of the smaller square. Plato uses this conversation to illustrate that the uneducated boy did not learn this geometric knowledge, but had (in some sense) possessed it all along. To Socrates, the fact that a simple conversation allowed the boy to "recollect" this knowledge suggested an intuition for geometric and mathematical truths more generally. Although Descartes and Kant did not agree with the view that people possessed geometric knowledge from birth, they emphasized the necessity of intuition for geometric concepts in the practice and epistemology of mathematics (Descartes, 1701; Kant, 1783).

Possessing an intuitive and sophisticated geometric reasoning system may confer an evolutionary advantage because understanding the layout of the environment and the shapes and configurations of objects can potentially facilitate actions that can save resources and time. There is substantial evidence that non-human animals are sensitive to information about both layout and object geometry and use it to navigate their spatial environments (Tommasi et al., 2012; Vallortigara, 2017; Regolin & Vallortigara, 1995). For example, newly hatched chicks raised in either circular or rectangular cages are equally adept at navigating rectangular enclosures, suggesting a stronger role for predisposed conceptions versus early learning of layout geometry (Chiandetti & Vallortigara, 2008, 2010). Using a similar task, Hermer and Spelke (1994) found that young children are also sensitive to layout geometry information, and fail to use non-geometric information even when it is relevant for task performance.

Experiments on macaque monkeys have found neural evidence for sensitivity to object geometry. Pasupathy and Connor (1999) identified cells in the V4 region of macaque visual cortex that responded specifically to the geometric properties of 2D objects. These cells were more sensitive to the underlying forms of objects rather than their edges. Other cells showed sensitivity to curvature and angles, and some of these cells showed a general response bias towards convexity, an important topological property. Additionally, neurons have been found in macaque inferotemporal (IT) cortex that are tuned to 3D geometric properties such as shape, orientation, and the relative positions of shape fragments (Yamane et al., 2008). Finally, neurons in anterior IT cortex have been shown to help code for object categories invariant to transformations involving scale, rotation, and translation (Hung et al., 2005).

This evidence is consistent with the core knowledge framework put forth by Spelke and colleagues, which differentiates sensitivity to GT concepts into two relatively independent systems that have been identified in non-human animals and infants (Spelke & Kinzler, 2007). The layout geometric system is responsible for encoding the geometry of the environment, and is sensitive to distance and sense, i.e., left–right direction or symmetry. The object geometry system is responsible for the representation of objects, and is sensitive to distance and angle. These systems have been shown to be sensitive to geometric properties depending on the context. The layout geometric system is more sensitive to GT concepts in large-scale navigation (Lee & Spelke, 2008), and the object geometry system is more sensitive for smaller pictures and objects (Gibson, 1969). Neither system alone is sufficient for representing Euclidean geometry concepts. Representing complex GT concepts requires integration between the two systems, a process that might in turn depend on development and learning (Spelke et al., 2010). That said, we note that core knowledge of geometry may not be a sufficient basis for later, formal understanding of GT concepts. Such understanding may require experience with cultural artifacts and representations (Ferreirós & García-Pérez, 2020).

In a seminal study, Dehaene et al. (2006) found evidence of remarkable sensitivity to even abstract GT concepts. They asked Mundurucu adults and children, and also American adults and children, to complete an odd-one-out task where they view a 3×2 matrix of images and judge which one of them is "ugly" (Fig. 1a). The images differ randomly on a number of incidental visual dimensions. Critically, five of the images exemplify the same GT concept whereas the sixth does not. Sensitivity to the concept is demonstrated by above-chance performance in selecting the conceptual deviant as the "ugly" one. All four groups of participants performed above chance on most of the 43 tested concepts. This performance did not simply reflect explicit instruction in geometry or topology because the Mundurucu have little access to formal education. Note that Mundurucu adults and children performed similarly to each other and similarly also to American children, whereas Americans adults performed better than the other groups. This opens the door to a potential role for formal schooling in becoming "sensitive" to certain GT concepts, a point to which we return below. Regardless of any absolute performance differences, there was congruence between the Mundurucu and American participants. Generally speaking, those of the 43 concepts that were difficult for the Mundurucu adults were also difficult for the American adults, and similarly for the children of both groups. Taken together, these findings are striking evidence that people are sensitive to GT concepts.

The question of whether sensitivity to GT concepts is present early in development or whether it is learned over time through experience, including formal schooling, has also been investigated. Experiment 1 of Izard and Spelke (2009) tested children ranging from 3 to 6 years old on the same odd-one-out task used by Dehaene et al. (2006). They found above-chance sensitivity to 27 of the 43 concepts. They also found interesting variation across classes of related concepts. For example, the children were sensitive to all of the eight Euclidean geometry concepts, but to none of the eight Geometrical transformation concepts. Interestingly, these classes were also among the easiest and most difficult, respectively, for the Mundurucu participants of the original Dehaene et al. (2006) study. (We return to the consistency of findings across studies below.) These developmental findings suggest that some classes of GT concepts may be "more intuitive" than others. They also raise the possibility that the "less intuitive" concepts might rely more heavily on domain-general abilities such as fluid and visuospatial reasoning, and might also require additional environmental support to learn (Greenough et al., 1987).

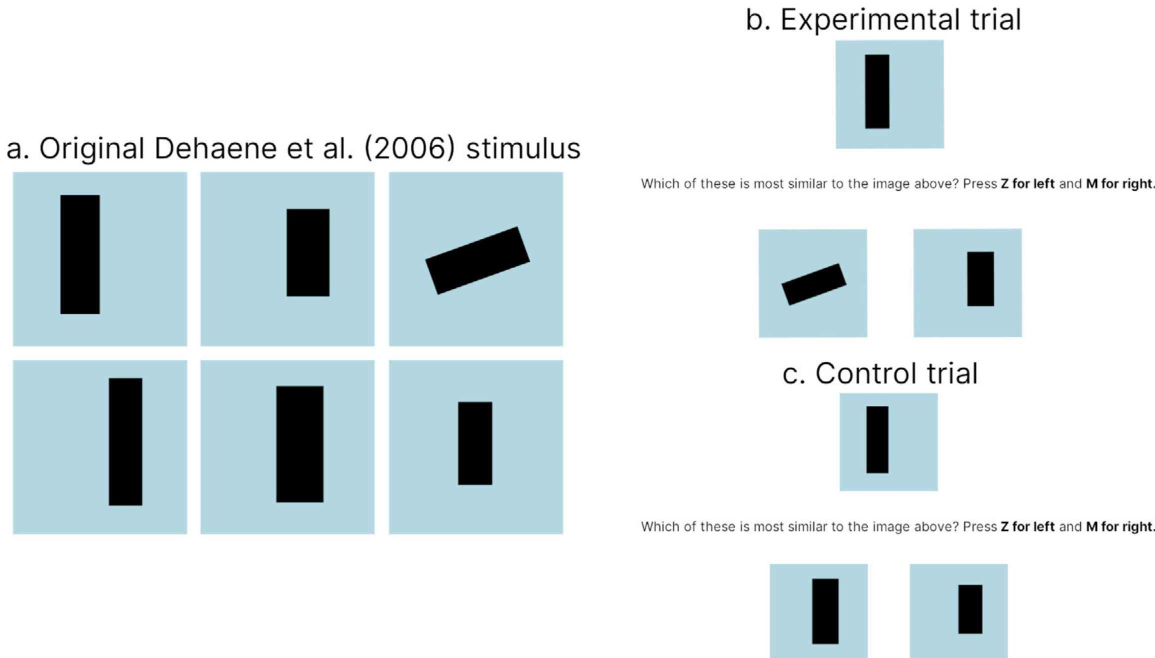

Fig. 1. (a) A stimulus from the original odd-one-out version of the Dehaene et al. (2006) task. (b) The six images were randomly used to generate the stimuli of the new 2-AFC task. The experimental stimulus includes the conceptual deviant. By contrast, the control stimulus only includes visual–perceptual distractors.

1.2. A visuospatial, inductive model

Lovett and Forbus (2011) offered a computational model of performance on the Dehaene et al. (2006) task. Their model combines existing models of visual perception, visuospatial reasoning, and analogical reasoning for the purpose of inductive generalization. It uses CogSketch (Forbus et al., 2011) to encode the 3×2 images of a stimulus. The representation scheme encodes both generic visual–perceptual attributes (e.g., length, edge) and those of a more geometric or topological nature (e.g., parallel, centered, collinear). These latter features can be thought of as embodying the Dehaene et al. (2006) claim that people are sensitive to (at least some) GT concepts. The model additionally uses the Structure Mapping Engine (SME; Falkenhainer et al. (1989)), a model of analogical generalization, to *compute* other GT concepts, such as rotation. After the 6 images have been encoded, the model uses the SME for a second purpose: to identify the odd-one-out. It first applies the SME to the three (training) images in the top row to induce a generalization. This generalization is then fit against each of the three (test) images in the bottom row. Next, this process is repeated but with the bottom row supplying the training images and the top row the test images. The image that offers the worst fit to the generalization derived from the other row is chosen as the odd-one-out.

The Lovett and Forbus (2011) model, in including features of a geometric or topological nature in its representation of images, is consistent with the Dehaene et al. (2006) claim that people are sensitive to GT concepts. However, this knowledge is not *sufficient* for performing the task. The model must also use analogical generalization to compute some GT concepts (e.g., rotation) and to perform the odd-one-out task itself. The success of the model therefore raises the question of the respective contributions of domain-specific abilities (i.e., intuitive GT concepts) and domain-general abilities (i.e., analogical generalization) to performance on the Dehaene et al. (2006) task. Perhaps the canonical example of a domain-general ability is fluid intelligence. It is instructive that the Lovett and Forbus (2011) approach has also been applied to account for performance on a measure of fluid intelligence, Raven's Standard Progressive Matrices (SPM) test (Lovett & Forbus, 2017). Thus, it is fair to say that a domain-general ability, analogical generalization, is a *necessary* component of the Lovett and Forbus (2011) model of human sensitivity to GT concepts.

The central role played by a domain-general ability in the Lovett and Forbus (2011) model further motivates the new version of the Dehaene et al. (2006) task developed for the current study. These modifications render it less of a reasoning task and more of a one-shot decision task, where the decision is more directly driven by sensitivity to GT concepts and is less dependent on domain-general reasoning.

1.3. The current study

The experimental findings reviewed above have been used to argue that human adults and children possess intuitive GT concepts, and that this knowledge is largely independent of formal schooling (Dehaene et al., 2006; Izard & Spelke, 2009). Lovett and Forbus (2011) offered a computational model of some of these findings that utilizes a representational encoding sensitive to some geometric and topological features, consistent with the intuitiveness proposal. However, it also accords a central role to a domain-general ability, inductive generalization via analogy, in explaining human performance. Here, we seek stronger evidence for the proposal that people are sensitive to GT concepts than has thus far been offered in the literature.

Our study of American adults utilized a version of the Dehaene et al. (2006) task modified so that chance represents a much higher bar to clear: 50%. Specifically, in the 2-AFC version of the task (Fig. 1), participants view a target image and must determine to which of the two alternative images it is more similar. One alternative image exemplifies the same geometric or topological concept as the standard, whereas the other does not. If people are sensitive to this concept, then they should notice the overlapping concept and privilege it in making their choice over mere visual–perceptual similarity.

The 2-AFC task also enables us to test a core proposal of the Lovett and Forbus (2011) model. This new version is less dependent on inductive generalization for successful performance. There is a single standard, and therefore much less of a basis for inducing a generalization than in the 3×2 stimuli of the original Dehaene et al. (2006) task, where five of the images exemplify the concept of interest.

To further isolate the contributions of domain-general abilities, we also collected independent measures of two that are important in individual differences research, and are potentially relevant to geometric and topological reasoning: inductive ability (i.e., fluid intelligence) and visuospatial ability (i.e., mental rotation). The former is directly relevant for evaluating the Lovett and Forbus (2011) model, specifically the proposal that performance on the Dehaene et al. (2006) task partially reflects the domain-general ability to form inductive generalizations. If this is the case, then a person's performance on the 2-AFC task should be associated with their performance on an independent measure of fluid intelligence. The visuospatial measure is also of interest because of the potential relevance of this ability to geometric and topological reasoning. If performance on the Dehaene et al. (2006) task partially reflects this domain-general ability, then there should be an association between performance on the 2-AFC task and performance on an independent measure of mental rotation. By contrast, if the 2-AFC task is a relatively pure measure of sensitivity to GT concepts, then relatively little of the variation on this task should be explained by these domain-general measures.

We also examined the role of formal schooling in sensitivity to GT concepts. This has previously been investigated at the group level, either by comparing groups that do or do not have access to Western schooling, such as American versus Mundurucu participants (Dehaene et al., 2006), or by comparing children before they entering Western schools to older children and adults from the same population (Izard & Spelke, 2009). Here, we take an individual differences approach. For our sample of young adults, we obtained their mathematical achievement (i.e., ACT-Math or SAT-Quantitative) scores and evaluated whether they were associated with their sensitivity to GT concepts.

We also evaluated the relationship between performance on the original Dehaene et al. (2006) task and performance on our 2-AFC version to determine whether they measure the same underlying concepts. We analyzed whether the profile of sensitivities to the 43 GT concepts was the same in the new 2-AFC version of the task and the original odd-one-out version across the various populations (i.e., Mundurucu vs. Americans) and ages (i.e., children vs. adults) that have been reported by Dehaene et al. (2006) and Izard and Spelke (2009).

Finally, the naturally speeded nature of the 2-AFC task offered a new avenue for finding evidence for sensitivity to GT concepts, in participants' reaction times. Prior studies using the original Dehaene et al. (2006) task have given participants as much time as they need to identify the odd-one-out, and analyses have focused on the accuracy data. We capitalized on the speeded nature of the 2-AFC version of the task to investigate relative sensitivity to different GT concepts by whether the correct choice "pops out" out (i.e., is identified very quickly and highly accurately) or not.

2. Methods

2.1. Participants

Eighty-eight undergraduates (60 female, 23 male, 3 non-binary or gender non-conforming, 2 preferred not to say) from a large public university in the Midwestern US enrolled in the study. They were recruited via email from a pool of participants who had participated in prior studies in our lab and via social media posts to student groups associated with the university (e.g., "Class of 2023"). The average age of the participants was 20.61 (SD = 2.40) years. Participants were compensated with a $15 electronic gift card. The protocol for the study was approved by the university's IRB.

2.2. Materials

2.2.1. Two-alternative forced choice task

The stimuli for the 2-AFC task were generated from the stimulus images of the original Dehaene et al. (2006) task, which were generously provided to us by Dr. Stanislas Dehaene. In the original task, participants picked which of the six images in a stimulus was the "odd one out". Five exemplified the geometric or topological concept of interest; only the deviant image did not. A response was coded as correct when the deviant was correctly identified as the odd-one-out. Dehaene et al. (2006) presented participants with 43 experimental items spanning a variety of GT concepts. The full list is provided in the supplementary materials.

We adapted the images of the original Dehaene et al. (2006) stimuli as follows. Each odd-one-out stimulus of that study was used to generate two 2-AFC stimuli, an experimental stimulus and a control stimulus. The experimental stimulus showed a target image that exemplified the concept of interest and two alternative images below it, one that exemplified the concept and one that did not, i.e., that was the deviant image in the original stimulus. The corresponding control stimulus was composed of the remaining three images. Notably, all three exemplified the concept of interest, and thus there was no "correct" choice from a mathematical perspective. This process generated 43 experimental and 43 control stimuli. These pairs of 2-AFC stimuli were randomly generated for each block for each participant. Thus, the concept-exemplifying images that appeared in the target position and the correct alternative position in the experimental stimulus, and in the three positions of the corresponding control stimulus, were shuffled randomly across participants and, for each participant, across blocks. The dependent measures were accuracy on the experimental stimuli and reaction time on the experimental and control stimuli.

The inclusion of the control stimuli is important. Performance on these trials is necessarily driven by purely visual–perceptual similarities and differences. Thus, their presence may have biased participants to make their choices on *all* trials, whether control of experimental, based more on visual–perceptual considerations than mathematical considerations. If this is the case, then their presence potentially worked *against* the hypothesis that people are sensitive to GT concepts. Their presence also potentially dampened the expectation that there is a "correct" answer on each trial, and may have caused participants to be guided more by their implicit sensitivities than by explicit reasoning. These are all desirable qualities from the perspective of raising the bar for demonstrating sensitivity to GT concepts.

The 43 GT concepts of Dehaene et al. (2006) were grouped by those authors into seven classes: *Topology*, *Euclidean geometry*, *Geometrical figures*, *Symmetrical figures*, *Chiral figures*, *Metric properties*, and *Geometrical transformations*. Other researchers have followed this grouping (Izard & Spelke, 2009; Lovett & Forbus, 2017; van der Ham et al., 2017), and we do the same to maintain continuity in the literature. However, it should be noted that there is some overlap between the classes. For example, Euclidean geometry is formally defined by Euclid's five postulates. These postulates play a role in almost all the GT concepts tested in this study, such as Symmetry and Geometric transformations. For the purpose of the current study, Euclidean geometry concepts refer more specifically to properties of lines and relationships between them such as parallelism, intersection, and angles that are invariant to transformations such as rotation and translation; they specifically exclude properties of position, orientation, and symmetry.

2.2.2. Mental rotation task

We used the Vandenberg and Kuse (1978) measure of mental rotation ability, which is based on the pioneering study of Shepard and Metzler (1971). For the specific images, we used the redrawn versions by Peters et al. (1995). The original measure is administered on paper; the current study used an online adaptation. On each trial, participants were presented with a target three-dimensional figure composed of blocks forming "arms", as in the original Shepard and Metzler (1971) stimuli. They were also presented with four other such figures as choices. Two depicted the original figure rotated by some number of degrees around one of the three principal axes. The other two depicted mirror images of the original figure that were similarly rotated. Participants had to select the two choices that were rotations of the original figure. An example trial is shown in Fig. 2. To be correct

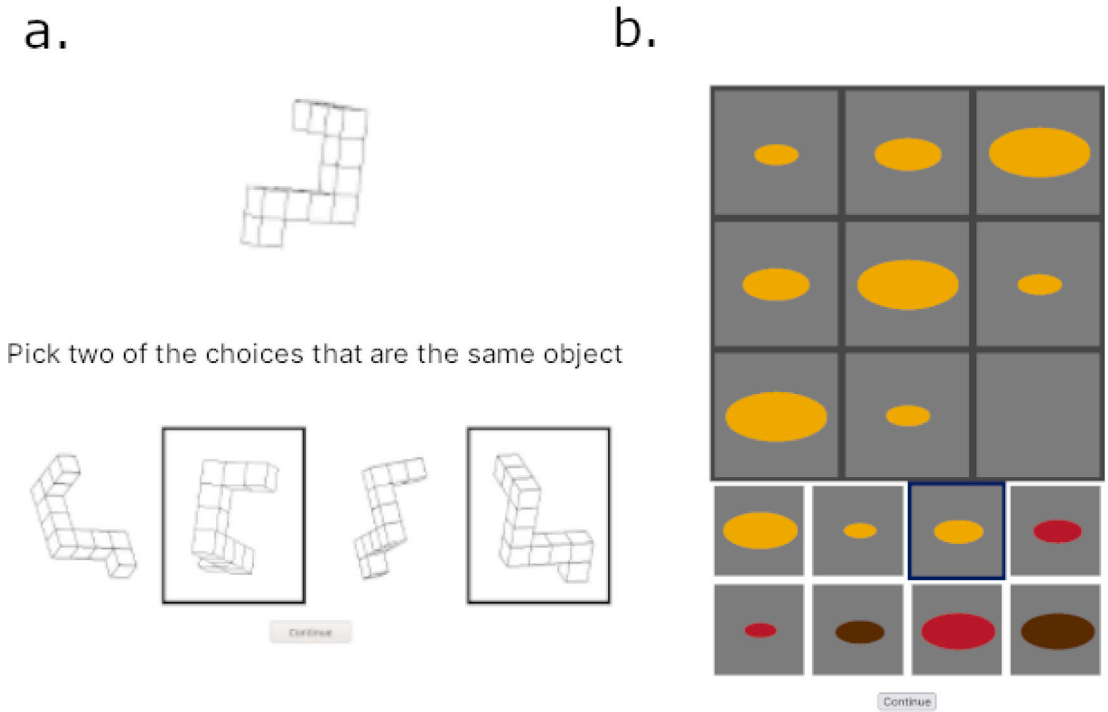

Fig. 2. Example trials of (a) the mental rotation task and (b) the UCMRT measure of fluid intelligence. For the mental rotation task, participants had to identify which two of the four choices were rotations of the figure shown in the target image. For the UCMRT, they had to identify which of the eight candidate images best fits in the empty cell given the patterns governing the matrix.

on a trial, participants had to select both images depicting rotations of the original figure. They had unlimited time to make their choices on the 24 items of the measure. The dependent variable was average accuracy across the items.

2.2.3. *Matrix reasoning task*

We administered the University of California Matrix Reasoning Task (UCMRT) developed by Pahor et al. (2019). This task measures fluid intelligence and inductive ability. We used the UCMRT instead of Raven's Advanced Progressive Matrices (RAPM) test (Raven, 1998) in part to reduce the time taken to complete the measure of fluid intelligence, from 50 min to 15 min, while still obtaining an accurate measure of inductive ability. Performance on UCMRT is comparable to performance on the RAPM: the correlation between the two measures is 0.6 among undergraduate students (Pahor et al., 2019).

We used the practice stimuli and the Set A stimuli of the UCMRT. We modified the task for online administration. Each item consists of a 3 × 3 matrix of cells. All but the bottom-right cell contains an image composed of simple geometric forms (e.g., colored squares, textured bars). Participants had to induce the rules governing the organization of the geometric forms across the rows and columns of the matrix, and had to select the one of the 8 candidate images that "best fit" in the empty cell, i.e., that was most consistent with the rules governing the matrix. An example item is shown in Fig. 2. Participants had unlimited time to complete the 23 items of Set A. The dependent variable was average accuracy across the items.

2.2.4. *Mathematical achievement measure*

As part of the informed consent process, participants gave permission for the experimenters to access their educational records. Of the 88 participants, 83 had ACT or SAT scores on file. The ACT is more prevalent than the SAT at the university where the data were collected. Therefore, we used percentiles on the ACT-Math and ACT-English sections as measures of mathematical and verbal achievement, respectively, when they were available. When they were not available, if the participant had taken the SAT, then percentiles on the SAT-Quantitative and SAT-Verbal sections were used instead. The mathematics portion of the ACT test involves understanding and solving equations; arithmetic operations on vectors and matrices; understanding functions; geometric principles involving trigonometry, the Pythagorean theorem, angles, areas, and triangle similarity; and basic statistics and probability (ACT, Inc., 0000). The mathematics portion of the SAT test involves linear equations, systems of linear equations, functions, ratios, percentages, proportional reasoning, graphs, geometry, trigonometry, and complex numbers (CollegeBoard, 0000). Notably, both tests include items that measure explicit geometric knowledge.

2.3. *Procedure*

Participants completed the entire study online using their web browser. Data were collected using a custom built Javascript based interface (code available at https://osf.io/szm3b/). Reaction times were measured using the 'performance.now()' API (Weiss, 2022), which provides times with millisecond precision on all major supported browsers (Firefox, Chrome, Edge, & Safari/WebKit). Testing conditions were variable, as participants could complete the study at any time, location, and using any computer during the data collection period. The median time taken to complete the study was 43.73 min.

After providing consent, participants first completed the 2-AFC version of the Dehaene et al. (2006) task. They were instructed to, on each trial, pick the alternative image that was most similar to the target image at the top. Participants first completed three practice trials without feedback. They then completed one block of 86 trials consisting of all 43 experimental trials and all 43 control trials presented in a random order. After an optional break, participants then completed a second block of 86 trials. Note that for the second block, the experimental and control stimuli were dynamically re-generated and their order again randomized. Participants were then provided an optional break before completing the mental rotation task. They first completed two practice items on which performance feedback was provided, and then completed the 24 experimental items. The order of the items was randomized. After another optional break, participants completed the UCMRT task. They first completed five practice items and were provided feedback. They had to answer each practice item correctly before they could move on to the next one, to ensure they understood the demands of the task. Participants then completed the 23 experimental items, this time without feedback. After completing the study, they were debriefed and emailed electronic gift cards as compensation.

3. Results

3.1. Software packages for analyses

Almost all the statistical analyses were conducted using the Python programming language, relying on the *numpy*, *pandas*, and *statsmodels* packages (Harris et al., 2020; McKinney, 2010; The pandas development team, 2020; Seabold & Perktold, 2010). Graphs were generated using the matplotlib library (Hunter, 2007). Statistical analyses that involved multiple comparisons among the GT concepts and among the classes were conducted using the R programming language and the *multcomp* package (R. Core Team, 2022; Hothorn et al., 2008). The data and the code for the analyses are available at https://osf.io/szm3b/.

3.2. Sensitivity to geometric and topological concepts

Participants showed remarkable sensitivity to GT concepts in the 2-AFC task. Fig. 3 shows the average accuracy on the experimental stimuli for each of the 43 concepts tested. Participants performed above chance (binomial test, $p < .05$) for almost all of the concepts. Chance here is 50%, and above-chance performance is indicated by the black vertical line in the figure. For only two concepts, Rotation and Center of quadrilateral, was their performance indistinguishable from chance. Exact p values for the 43 binomial tests are listed in the supplementary materials.

The 2-AFC experimental paradigm differs from the Dehaene et al. (2006) task in having less complex stimuli (i.e., three images vs. six images), asking participants to make a “similarity” judgment vs. an “odd-one-out” judgment, and being speeded. For these reasons, we did not conduct an *a priori* power analysis, and instead recruited a sample at least as large as those of prior studies (Dehaene et al., 2006; Izard & Spelke, 2009). However, we did conduct a *post hoc* binomial power analysis to evaluate whether our sample size had been sufficient (Rosner, 2011). Because of the high accuracy on almost all of the items, the estimated power of our study was greater than 99%. Individual *post hoc* power analyses on each of the 41 items for which performance was significantly above chance revealed only one GT concept for which the observed power was less that 80%: Horizontal symmetry (65.5%).

We next considered performance by aggregating the 43 GT concepts into seven classes, following prior studies: *Topology*, *Symmetrical figures*, *Metric properties*, *Geometrical transformations*, *Geometrical figures*, *Euclidean geometry*, and *Chiral figures*. We conducted binomial tests comparing accuracies at the class level against a chance level of 50%. Participants performed above chance for all seven classes (ps < 0.0001, Fig. 4). We conducted Tukey pairwise comparisons between the different classes to determine participants' relative sensitivity to them. We used a single-step p value correction using the `multcomp` R package. Tukey's range test is conservative to groups with unequal sample sizes, which was the case here (i.e., because the seven classes included different numbers of the original 43 concepts). The results of the $\frac{7 \cdot 6}{2} = 21$ comparisons are listed in the supplementary materials. To summarize them, against a p value threshold of 0.05, four clusters of similar performance emerged. Participants performed worst on the *Metric properties* and *Geometrical transformations* classes. They performed better on the *Topology*, *Symmetrical figures*, and *Chiral figures* classes. They performed better still on the *Geometrical figures* class, and they performed best on the *Euclidean geometry* class.

3.3. Matrix reasoning

We next evaluated the degree to which performance was driven not by domain-specific knowledge of GT concepts but rather by more domain-general abilities. The first such ability we considered was inductive generalization. Recall that one of the motivations for modifying the Dehaene et al. (2006) task was to minimize the potential contribution of this ability to task performance. The 2-AFC modification makes it more difficult to induce the relevant concept from an experimental stimulus because only two (of the three) images exemplify the concept versus five (of the six) images in the original odd-one-out task. The analyses reported here are particularly relevant for the Lovett and Forbus (2011) model, which uses analogical generalization (i.e., the SME) to induce some geometric and topological features of images, such as rotation, and also to compute abstractions over subsets of images.

We adopted an individual differences approach: For each participant, we computed their overall accuracy on the 2-AFC task averaged across all 43 concepts. We indexed their inductive ability by their fluid intelligence as measured by the UCMRT (Pahor et al., 2019). We fit a linear model predicting accuracy on the 2-AFC task from accuracy on the matrix reasoning task. The model was significant ($p < 0.005$), explaining 12.52% of the variance. The results are shown in Fig. 5. Thus, it appears that fluid intelligence plays a role in sensitivity to GT concepts, but it is only a small one. More than 85% of the variation in this sensitivity was left unexplained after accounting for this ability. Thus, other factors must be playing significant roles.

3.4. Mental rotation

The second domain-general ability we considered that might be relevant for geometric and topological reasoning was mental rotation. We again used an individual differences approach, fitting a linear model predicting accuracy on the 2-AFC task from accuracy on the mental rotation task. The model was significant ($p < 0.001$), explaining 18.73% of the variance. Fig. 5 shows the relationship between performance on the two tasks. The modest R^2 value indicates that more than 80% of the variation in geometric and topological sensitivity is left unexplained by mental rotation ability. Again, other factors must be playing significant roles.

3.5. Combined analysis

The prior analyses showed that neither fluid intelligence nor mental rotation alone contribute substantially to explaining sensitivity to GT concepts. However, it is possible that their contributions are largely independent, and that together they explain considerable variation. To evaluate this possibility, we fit a linear model predicting accuracy on the 2-AFC task from accuracy on each of the matrix reasoning and mental rotation measures. Although the model was significant ($F(2, 83) = 11.73$, $p < 0.0001$), it did not explain much more variance ($R^2 = 0.2163$) than the individual models considered above. In the combined model, only mental rotation ($\beta = 0.086$, $p = 0.005$) was a significant predictor. Thus, fluid intelligence and mental rotation ability appear to explain only a small portion of performance on the 2-AFC task, leaving substantial variation unexplained. This is consistent with the proposal that sensitivity to GT concepts is more of a domain-specific ability.

That said, it is possible that sensitivity to *some* of the seven classes of concepts are substantially driven by the domain-general abilities considered here. To test this possibility, we repeated the combined analysis separately for each class. That is, we computed the mean accuracy across the concepts that each class spans and then fit a linear model predicting this variable using performance on the UCMRT and the mental rotation tasks. The models were significant for the Geometrical transformations ($R^2 = 0.12, p = 0.004$), Metric properties

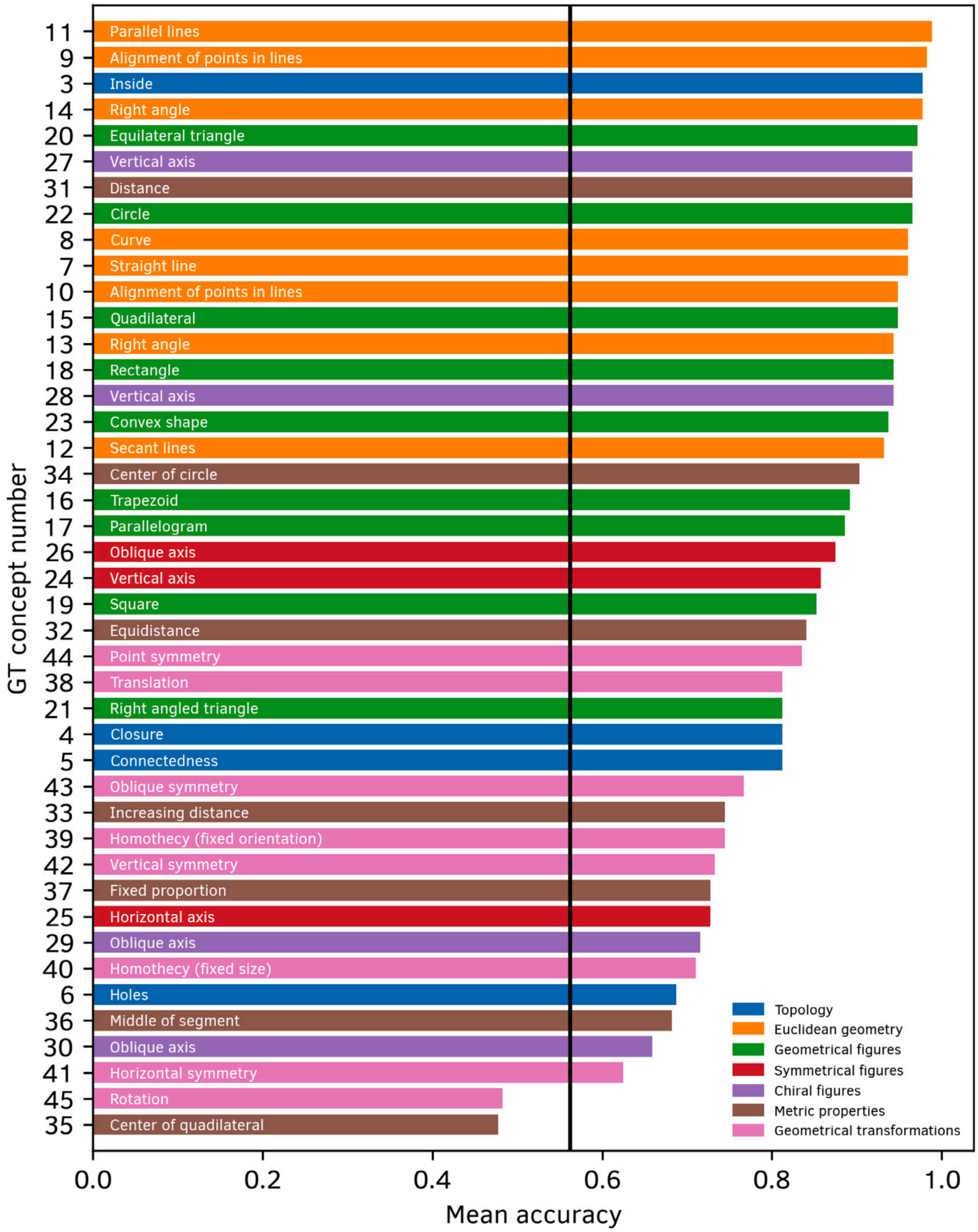

Fig. 3. Average accuracy on each of the 43 GT concepts. The color indicates the class to which each concept belongs. The black line represents above-chance (binomial test, $p < 0.05$) performance. Participants were above chance for all but two concepts, Rotation and Center of quadrilateral.

($R^2 = 0.13, p = 0.003$), and Symmetrical figures ($R^2 = 0.21, p = 0.00005$) classes. They were not significant for the Chiral Figures, Euclidean geometry, Geometrical figures, and Topology (Chiral figures, $p = 0.053$; rest, $p > 0.08$) classes. These findings suggest the some classes of GT concepts may rely more on domain-general abilities than others. We return to this point below, in the Discussion.

The best fit of the combined model was for the Symmetrical figures class (although that model leaves 79% of the variance unexplained). Why might this be the case? One possible explanation is that there appear to be similarities between the stimulus images for the Symmetrical figures concepts and the items of fluid intelligence tests such as the UCMRT and RAPM. Both require mental transformation of images to align corresponding elements. These mental operations might also be considered Geometrical transformations, and we note that performance on this class was also significantly associated with the domain-general abilities.

3.6. Reaction time

An advantage of the 2-AFC task over the original Dehaene et al. (2006) task is that it is naturally speeded, making participants' reaction times potentially informative. Participants were fast in an absolute sense: half completed trials in an average time of less than 2000 ms (Median = 1755ms), with an interquartile range of 1266ms. This provides an opportunity to investigate their relative sensitivity to different GT concepts. In particular, are there some concepts for which the correct answers (i.e, the alternative image that shares the same concept as the target image) "pop out" for participants, as indicated behaviorally by both fast reaction times and highly accurate responses?

To address this research question, we used a linear model to estimate the association between reaction time and accuracy across the 43 experimental stimuli. There was a significant, negative association,

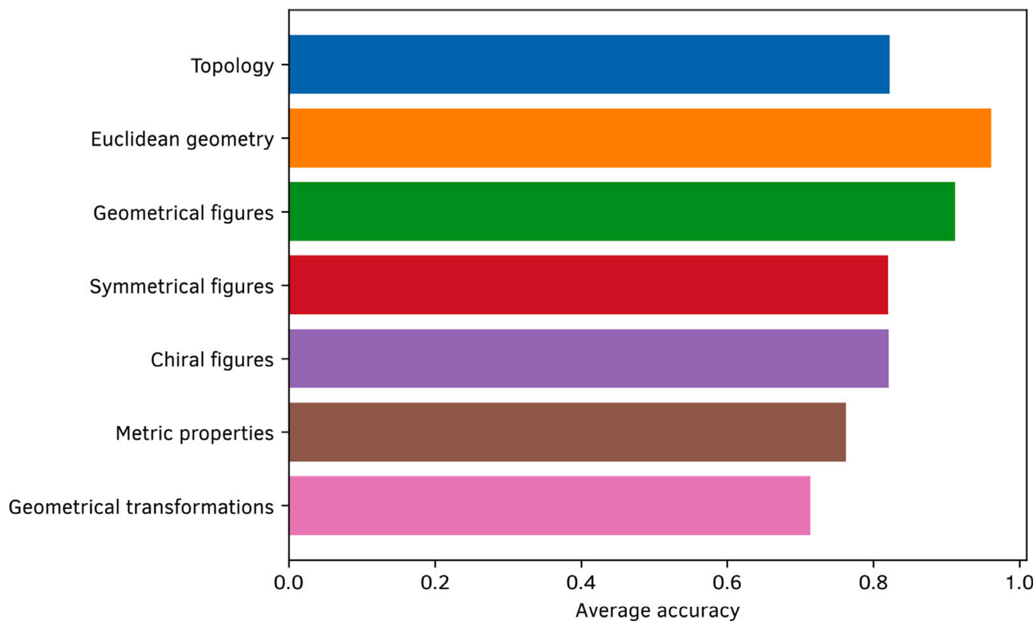

Fig. 4. Average accuracy for the seven classes of GT concepts for the experimental trials. Participants were above chance for all classes ($p < 0.0001$).

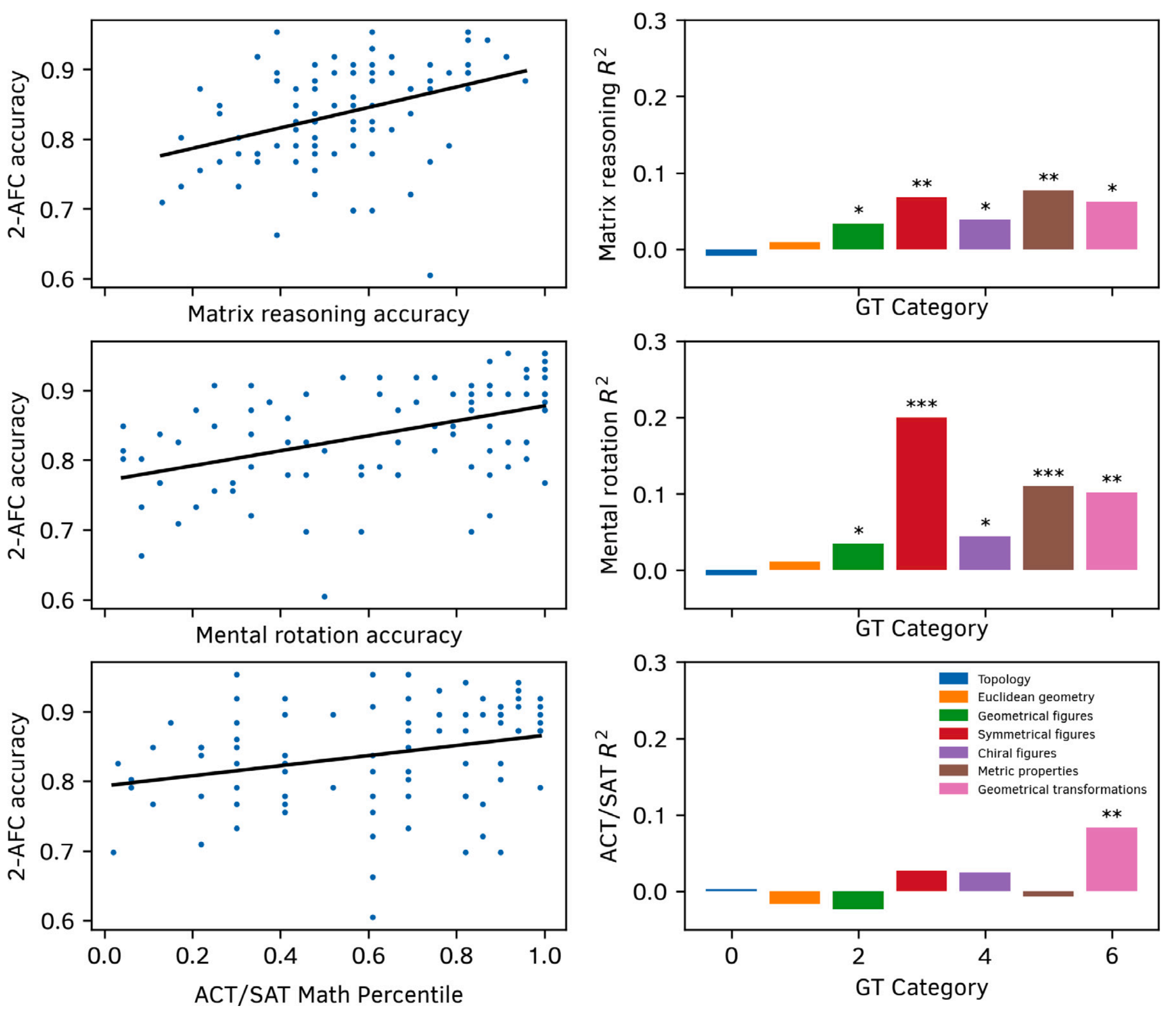

Fig. 5. Left: Relationship between average accuracy on the 2-AFC task and performance on the matrix reasoning task, mental rotation task, and ACT/SAT math test, for each participant. Right: R^2 for models predicting accuracy on the 2-AFC separately for each of the seven classes of GT concepts from performance on the matrix reasoning task, mental rotation task, and ACT/SAT math test. Model term significance is indicated by asterisks: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

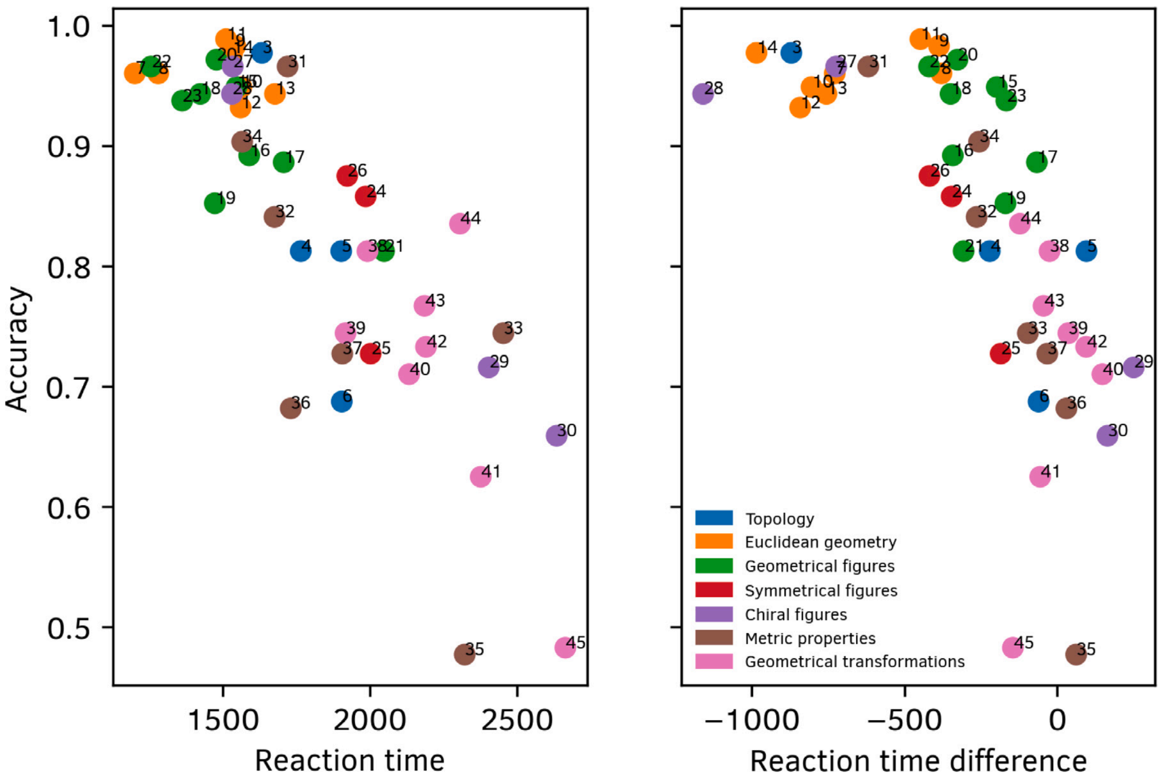

Fig. 6. Association between accuracy on the 43 GT concepts and (left) reaction time and (right) reaction time difference (experimental trial minus control trial). Colors indicate the classes to which the concepts belong.

with faster reaction times associated with higher accuracies ($R^2 = 0.70$, $p < 0.00001$, Fig. 6). More precisely, an increase in one second of reaction time was associated with a decrease in average accuracy of 30%. Thus, the geometric and topological concepts that were judged most quickly were also the ones judged most accurately. Methodologically, this is reassuring: there was no speed–accuracy trade-off. More substantively, this finding suggests that for some concepts more than others, the alternate image that shares the same concept as the target image "pops out" as more similar. The concepts that "popped out" are located in the upper left of the figure, and are dominated by the Euclidean geometry and Geometrical figures classes.

The inclusion of the control trials enables a more refined analysis of the reaction time data. The time to complete a control trial should be driven by visual–perceptual processing alone, since both alternatives exemplify the same concept as the standard. By contrast, the time to complete an experimental trial should additionally be driven by – or even dominated by – the salience of the concept of interest, which is exemplified in both the correct alternative image and the standard. Thus, a contrastive analysis can reveal the concepts to which participants are most sensitive to above and beyond the visual–perceptual qualities of the stimuli.

We conducted this analysis at two grain sizes. At a coarser grain, we evaluated whether participants were faster on average on experimental versus control trials; this would be the case if participants are sensitive to GT concepts and their presence in the correct alternative on experimental trials. This prediction found support. Participants were faster on average on experimental versus control trials ($t(42) = 5.2303$, $p < 0.0001$, 95% CI $= [168, 379]$ ms), with a median RT difference of 273 ms. Again, this is consistent with the proposal that people are generally sensitive to GT concepts.

We then evaluated whether participants differed in their reaction times for control trials for the 43 GT concepts and the 7 classes. We should not see such differences if our assumption that the control trials were completed based on the visual–perceptual features of the stimuli alone, which should be the unsystematic. We fit two linear models with the GT concept and the GT class, respectively, predicting reaction time on the control trials. The assumption was that these models should explain very little variation. This was indeed the case. Although the model using GT concept to predict reaction time on control trials was significant ($F(42, 3741) = 2.675, p < 0.001$), it explained little variation ($R^2 = 0.029$). The model using GT class was not significant ($F(6, 609) = 1.985, p = 0.066$) and again explained very little variation ($R^2 = 0.019$). This indicates that the control trials generally took participants the same amount of time across the 43 different GT concepts and across the 7 different classes. It licenses their serving as baselines for comparison with experimental trials.

Building on this, we conducted a finer-grained analysis using the pairs of experimental and control trials. For each concept, we computed the difference between RT on the experimental trial and RT on the control trial, averaged across participants. This difference indexes people's sensitivity to the concept: The more negative this difference, the faster the conceptually-driven experimental trial is relative to the perceptually-driven control trial, and thus the more "pop-out" participants might have experienced. The data are shown in Fig. 6. We fit a linear model predicting accuracy for each concept using this RT difference measure. This model explained 44% of the variance ($F(1, 41) = 34.5$). There was a significant negative association, with each second of increase in reaction time difference associated with a 16% decrease in accuracy. At the class level, Euclidean geometry concepts showed the most "pop out" – they had the most negative RT difference (−922 ms) and the highest accuracy (0.96). By comparison, Geometric transformation concepts had the least negative RT difference (−24 ms) and the lowest accuracy (0.71), indicating that participants were least sensitive to them.

3.7. Relationship to mathematical achievement

We next investigated whether there is an association between sensitivity to GT concepts and general mathematical achievement. There might be a positive association if children with greater sensitivity are better able to learn from formal mathematics instruction. Consistent with this possibility is the finding that sensitivity to these concepts increases from ages 3 to 11, as American children receive formal mathematics instruction in preschool and elementary school (Izard & Spelke, 2009). However, there might be no such association if sensitivity to GT concepts is less important for learning from formal mathematics instruction than other abilities, for example facility with symbol systems. Consistent with this possibility, van der Ham et al. (2017) found that sensitivity to GT concepts as measured by the original Dehaene et al. (2006) task is associated with the *number of years* of formal education an individual has had. However, time in school is not the same as mathematical achievement in school, the variable of interest here.

We evaluated whether performance on the 2-AFC task predicts mathematical achievement. We first examined the association between a participant's overall accuracy on the 2-AFC task and their math percentile (Fig. 5). There was a significant relationship between the two variables ($F(1,84) = 7.220, R^2 = 0.079, p = 0.009$). Note that this relationship may overestimate the predictive power of performance on the 2-AFC task for mathematical achievement, for a number of reasons. One is that both the ACT-Math and SAT-Quantitative tests include items that measure knowledge of geometry, and so there may be an element of "geometry predicting geometry" here.

The significant correlation between 2-AFC accuracy and math percentile may also reflect a number of incidental factors that we can control for, such as general intellectual ability and comfort with timed testing. To control for these factors, we fit a linear model predicting accuracy on the 2-AFC task using both math percentile and verbal percentile as predictors. If math percentile still predicts accuracy on the 2-AFC task after controlling for verbal percentile, then this would be stronger evidence for an association between mathematical achievement and sensitivity to GT concepts. In fact, this was the case, though only weakly. The overall model was significant ($F(2,80) = 3.427$, $p =$ 0.03), but the variance accounted for was rather modest ($R^2 = 0.079$). That said, math percentile was a significant predictor in the full model ($p = 0.04$), whereas verbal percentile was not ($p = 0.93$).

We repeated this analysis separately for each of the seven classes of GT concepts. Math percentile and verbal percentile were the predictors in all of the models. The results are shown in Fig. 5. The only class for which 2-AFC accuracy was predicted by the model was Geometric transformations ($R^2 = 0.09, p = 0.009$). This was not the case for the other classes: Euclidean geometry ($p = 0.7$), Topology ($p = 0.33$), Chiral figures ($p = 0.06$), Geometrical figures ($p = 0.95$), Metric properties ($p = 0.38$), and Symmetrical figures ($p = 0.10$). We return to this pattern of differential associations below, in the Discussion.

3.8. Comparisons with prior studies

We modified the original Dehaene et al. (2006) task to increase the performance level participants had to achieve to demonstrate above-chance sensitivity to GT concepts, and also to minimize the role of domain-general abilities such as inductive generalization (i.e., fluid intelligence) and visuospatial reasoning (i.e., mental rotation). Thus, the 2-AFC task should be a purer measure of people's sensitivity to these concepts. This raises the question of how the modified task compares with the original six-panel, odd-one-out task of Dehaene et al. (2006). We evaluated the correspondence between the two tasks in a correlational analysis of multiple data sets.

Dehaene et al. (2006) provides the accuracy data for the 43 GT concepts for the combined group of Mundurucu adults and children. (Collapsing the data in this way makes sense because the overall performance of the two groups was comparable.) Izard and Spelke (2009) used the same stimuli with largely middle-class participants recruited from Boston and New York City. Their Experiment 1 presents the accuracy data for children ages 3–6 years old. Their Experiment 2 tested children and adults covering a wide range of ages. For this

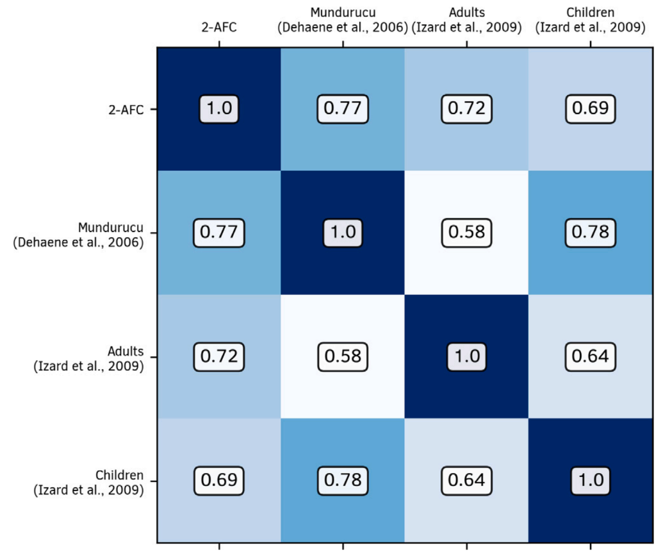

Fig. 7. Matrix of Pearson *r* correlations comparing accuracies on the 43 GT concepts pairwise among four studies using different populations (Western and Mundurucu, adults and children) and different tasks (2-AFC, odd-one-out).

experiment, we requested the accuracy data for the age band (adults ages 18–25 years old) that most closely matches the inclusion criteria of the current study (adults ages 18–24 years), which Dr. Izard generously provided.

We evaluated the correspondence between the four data sets – the American young adults of the current study, the Mundurucu adults and children of the Dehaene et al. (2006) study, the American children of Experiment 1 of the Izard and Spelke (2009) study, and the American young adults of Experiment 2 of the same study. For each pair of data sets, we calculated the Pearson *r* coefficient across the accuracies on the 43 concepts. The resulting correlation matrix is shown in Fig. 7. The adults in the current study showed an accuracy profile similar to that of the adults ($r = 0.72$) and the children ($r = 0.69$) of the Izard and Spelke (2009) study, which also sampled from a Western population. Their profile was also similar to that of the Mundurucu adults and children of the Dehaene et al. (2006) study, with $r = 0.76$. The correlations between our sample, collected using the 2-AFC task, and the other samples, collected using the odd-one-out task, were at least as high as the correlations among the data sets that used the same odd-one-out task. Thus, despite their different formats, our 2-AFC task and the original Dehaene et al. (2006) task appear to be tapping the same sensitivity to GT concepts.

Following Dehaene et al. (2006), we take the definitions of the 43 GT concepts and their assignment to 7 classes to reflect how mathematicians organize the domains of geometry and topology. That is, the 7 classes are *mathematically* meaningful. This raises the question of whether they are also *psychologically* meaningful. We addressed this question in an exploratory factor analysis. See the supplementary materials for the results.

4. Discussion

The primary purpose of the current study was to investigate whether young adults are sensitive to geometric and topological (GT) concepts while minimizing or controlling for other factors. Using a modified version of the Dehaene et al. (2006) odd-one-out task, we found evidence that people possess such sensitivity. Strikingly, this was the case for 41 of the 43 concepts tested.

In more detail, the new version of the task uses a 2-AFC format where participants view a standard image and have to choose which of two alternative images is most similar to it. The task is more difficult than the original Dehaene et al. (2006) task for three reasons. First, each experimental stimulus only shows two examples of the target concept, not the five examples shown in the original task. This reduces the potential role for inductive generalization in making the correct choice. This in turn reduces the potential contributions of domain-general abilities and makes performance more reliant on pre-existing, domain-specific understanding of these concepts. Second, the chance criterion is higher than in the original study (50% versus 16.67%). Thus, participants have a higher bar to clear to demonstrate above-chance sensitivity to these concepts. Third, the experimental trials, where only one alternative image exemplifies the same concept as the target image, were interspersed with an equal number of control trials, where both alternative images exemplify the same concept, and thus the choice between them must be driven solely by visual–perceptual properties. The presence of the control trials removes the expectation of a "mathematically" correct answer on each trial and arguably biases participants to make their choices for *all* trials based more on visual–perceptual properties, working against the sensitivity hypothesis. For these reasons, the finding of above-chance sensitivity to GT concepts using the 2-AFC task is stronger evidence than has been provided in prior studies.

The speeded nature of the 2-AFC task and the inclusion of control trials enabled an expanded analysis of sensitivity to GT concepts. This is because accuracy information is less informative when performance is at ceiling. For example, consider the Euclidean geometry items shown in the left panel of Fig. 6. Participants were very accurate on all of them, making it difficult to determine if they were more sensitive to some Euclidean geometry concepts compared to others. However, when reaction time information was also considered, some separation began to appear: People were faster for items 7 and 8 than for the other items of this class. A further sharpening of the story was possible by also considering the control items. In the right panel, the reaction time for the control trial, which presumably reflects the inherent visual–perceptual complexity of the images, is subtracted off from the reaction time for the experimental trial, which is additionally driven by the GT concept of interest. With this "purer" measure of speed, we see the items of the Euclidean geometry class separate into two distinct clusters, suggesting higher sensitivity to items 7, 10, and 12–14 compared to items 8, 9, and 11. Thus, utilizing a speeded task and including control trials enables detection of fine-grained differences in sensitivity to GT concepts.

We also investigated whether participants' performance on the 2-AFC task could be explained by two domain-general abilities, inductive generalization and visuospatial reasoning, as indexed by measures of fluid intelligence and mental rotation, respectively. Here, we took an individual differences approach. As predicted, fluid intelligence and mental rotation only explained a small portion of the variance on the 2-AFC task — around 20 percent. Thus, these domain-general abilities play a relatively small role in task performance, and other factors must explain why some participants are more sensitive to some GT concepts than others. The explanation favored here is that this variation is driven by differences in people's domain-specific understanding of GT concepts.

A natural question is whether sensitivity to GT concepts is a useful foundation for the formal study of mathematics in school. If so, then we would expect to see an association between the sensitivity a participant shows towards these concepts and his or her scores on standardized tests of mathematical achievement. This was indeed the case: math percentile was significantly associated with sensitivity to GT concepts, even when controlling for verbal percentile, which was used as a proxy for general academic achievement and comfort with timed testing. It should be noted that the association between math percentile and sensitivity to GT concepts was relatively weak ($R^2 = 0.079$). Nevertheless, it suggests that it may be productive to explore grounding formal instruction on GT concepts on students' intuitive understandings of them. An interesting developmental question is whether sensitivity to GT concepts early in life predicts formal knowledge of GT concepts later in life. A related educational question is whether students' sensitivity to these concepts can be sharpened through training, and whether such improvements carry over and support better learning from formal instruction.

Finally, we consolidated the findings of several studies that have investigated people's sensitivity to GT concepts. Dehaene et al. (2006) was the first such study. It used the six-panel, odd-one-out task with a sample of Mundurucu children and adults. Izard and Spelke (2009) used the same task, both in a sample of Western children ages 3–6 years old and in a sample of Western adults including those ages 18–25 years old. The current study used a modified 2-AFC version of the original task and recruited a sample of Western adult ages 18–24 years old. Despite these differences in population, age, and task, correlations of the accuracy profiles across the 43 GT concepts revealed remarkable agreement across the studies. This indicates a stability in sensitivity to GT concepts regardless of how it is measured, and establishes the 2-AFC task as a useful paradigm for future studies. This version offers several advantages over the original in addition to those mentioned above. It is a speeded task, making participants' reaction time data more readily interpretable. This enabled us to rule out a potential speed–accuracy tradeoff in people's performance. In addition, the reduced stimulus complexity and response demands of the 2-AFC task may make it more useful for future studies of the geometric and topological sensitivities of very young children.

4.1. Models of human geometric and topological sensitivity

The results of the current study using the new 2-AFC task have implications for computational models of human performance on the original six-panel, odd-one-out task (Dehaene et al., 2006). The Lovett and Forbus (2011) model employs a representational scheme that directly encodes geometric and topological attributes such as straight line, curve, closure, and inside (i.e., containment). This is consistent with the Dehaene et al. (2006) proposal that people possess intuitive GT concepts. However, the model derives other attributes, such as rotation, using a domain-general mechanism, analogical generalization as implemented by the SME (Falkenhainer et al., 1989). It also uses the SME to reason over the six images to identify the odd-one-out.

The current findings offer a theoretical challenge to the Lovett and Forbus (2011) model. Analogical generalization and fluid intelligence are potentially related domain-general abilities. In fact, their relation is implied in the successful application of the same, SME-centered modeling approach to account for human performance on Raven's SPM, a standard measure of fluid intelligence (Lovett & Forbus, 2017). However, the current study found that fluid intelligence explained only 12.52% of the variance in people's sensitivity to GT concepts. This is potentially at odds with the large role the Lovett and Forbus (2011) model accords to this domain-general mechanism.

Another challenge the current findings offer is empirical. It is an open question whether the Lovett and Forbus (2011) model can attain the high accuracies achieved by our participants on the 2-AFC task. Recall that the model uses the SME to induce one generalization from the top three images of a six-panel stimulus, and a second generalization from the bottom three images. It then uses each generalization to identify the anomalous image in the held-out row. However, the 2-AFC task provides fewer examples of the concept: only two of the three images exemplify it, versus five of the six images in the original task. This may not be a sufficient basis for the Lovett and Forbus (2011) model to form its generalizations. That said, it is possible that the model can be modified to use various subsets of images in the 2-AFC task stimuli to make a decision. Whether it can be extended to match the exceptional performance of humans on this task, who showed above-chance sensitivity to 41 of the 43 concepts, is a question for future research.

Two other models have attempted to capture human performance on the original Dehaene et al. (2006) task. The McGreggor and Goel

(2013) model adopts a low-level, image processing approach. It derives a mutual fractal representation for each pair of the six images. It then uses these representations to compute the similarity of each image to the other five images, and selects the least similar one as the odd-one-out. The Sheghava and Goel (2020) model uses a *symmetry* approach inspired by Gestalt principles. It uses principal components analysis (PCA) to estimate the principal axis of each of the six images, computes the image's pixel-wise symmetry about this axis, and uses this information to derive further geometric and topological features. It then selects the image whose features are most discrepant with those of the other five images as the odd-one-out. Both of these models share with the Lovett and Forbus (2011) model the ability to approximate human performance on the original Dehaene et al. (2006) task. They also share with that model the general strategy of identifying the least similar (or most discrepant) image as the odd-one-out. This general strategy appears to work well for stimuli where five of six images exemplify the target concept and only one image does not. It may work less well for the experimental stimuli of the 2-AFC task, where only two of the images exemplify the same target concept.

4.2. Graded sensitivity to geometric and topological concepts

The results of the current study coupled with those of prior studies (Dehaene et al., 2006; Izard & Spelke, 2009) together demonstrate that people are sensitive to GT concepts. How should we understand this sensitivity?

One possibility is that sensitivity to GT concepts is a fundamental property of the human mind. For example, Spelke and Kinzler (2007) propose that people have *core knowledge* of geometry. They argue that humans possess cognitive mechanisms for navigating the spatial layout of the environment and parsing its geometric relationships, mechanisms that are ultimately grounded in the evolutionary importance of vision and navigation. The same mechanisms, they argue, explain human sensitivity to GT concepts. From a more formal point of view, Shepard (2001) argues that the universe is governed by universal, invariant principles such as symmetry, and it is therefore not surprising that humans are sensitive to such GT concepts. Or more pithily: "Geometry is more deeply internalized than physics" (p. 585).

The current findings suggest a different, though related, answer to the question of how we should understand human sensitivity to GT concepts. It may be that some GT concepts are more *domain-specific* and others are more *domain-general*. Thus, rather than speak of a "geometric module", we see evidence for a continuum of GT concepts. More domain-specific concepts are those that are processed more quickly and accurately; phenomenologically speaking, they "pop out". They are less grounded on general cognitive abilities such as fluid and visuospatial reasoning, and less associated with success in mathematics classrooms as indexed by mathematical achievement scores. By contrast, more domain-general concepts are those that are processed more slowly and less accurately. Their computation requires the support of general cognitive abilities such as fluid and visuospatial reasoning. In part for these reasons, fluency with more domain-general concepts is more associated with mathematical achievement scores.

To illustrate the proposed continuum, we computed aggregate *domain-centricity* scores for each of the seven classes of GT concepts. These scores were derived from five of the dependent measures reported above: mean accuracy, median RT difference (between experimental and control items for the same concept), and the R^2s for predicting mean accuracy from mental rotation, fluid intelligence, and mathematical achievement scores. We normalized each dependent measure so that 0 corresponds to the greatest domain-specificity: the highest mean accuracy, the most negative median RT difference, and the lowest R^2s for each of the three individual differences measures. Conversely, 1 indicates the greatest domain-generality on the measure. We added together these five normalized dependent measures to form the aggregate domain-centricity score for each of the 43 GT concepts, and averaged these for each of the 7 classes. The scores range from 0 to 5, with 0 being the most domain-specific class and 5 being the most domain-general. They are shown in Fig. 8.

The figure strikingly depicts people's graded sensitivity to GT concepts. Participants were most sensitive to concepts from the Euclidean geometry, Topology, and Geometric figures classes. In our proposal, these classes can be considered the most domain-specific, i.e., the ones most directly driven by a purported geometric module. Participants were least sensitive to the Geometric transformations, Symmetrical figures, and Metric properties classes. These can be considered the most domain-general, i.e., the ones most dependent on general cognitive abilities and learning through formal instruction.

This analysis finds support in recent studies of congenitally blind participants and blindfolded but otherwise sighted participants by Marlair et al. (2021) and Heimler et al. (2021). The most difficult items for these participants were those from the Geometrical transformations and Symmetrical figures classes, which we evaluate to be the most domain-general. To the degree that sensitivity to these concepts depends more strongly on general cognitive abilities such as mental rotation and visuospatial reasoning and to experiences gained in formal mathematics instruction, it makes sense that these participants were least sensitive to them.

The graded sensitivity proposal predicts that formal schooling is less important for sensitivity to more domain-specific concepts. The literature provides some support for this prediction. The depiction in Fig. 8 indicates that Euclidean geometry is the most domain-specific class. Consistent with the prediction, the Mundurucu, who do not generally have access to formal schooling, show strong sensitivity to concepts from this class, such as parallel lines in the plane (Izard et al., 2011). Young children on the cusp of formal schooling also show sensitivity to Euclidean geometry concepts such as right angle (Izard et al., 2014). Strikingly, Izard and Spelke (2009) found that children ages 3–6 years old show above-chance sensitivity to *all* eight concepts of the Euclidean geometry class. (Conversely, they showed sensitivity to *none* of the eight Geometrical transformation concepts, which we propose to be the most domain-general class.) Finally, uneducated Himba adults from rural Namibia are sensitive to the parallel lines concept of the Euclidean geometry class (Sablé-Meyer et al., 2021). They are also sensitive to the square, rectangle, and parallelogram concepts of the Geometric figures class, which is also located nearer to the domain-specific pole of Fig. 8.

Graded sensitivity also predicts that domain-specific concepts should be more evident in non-human animals. Consistent with this prediction, newly hatched chicks can distinguish rectangular cages from circular cages, exhibiting sensitivity to concepts of the relatively domain-specific Geometric figures class (Chiandetti & Vallortigara, 2008, 2010).

That said, the literature also contains apparent exceptions to the predictions of graded sensitivity. For example, the Symmetrical figures class is located near the domain-general pole of the continuum in Fig. 8. Young children have yet to experience formal mathematics instruction and also possess relatively immature general cognitive abilities. Thus, graded sensitivity predicts that they should *not* be sensitive to symmetry in visual stimuli. However, Huang et al. (2018) found that preschool children look longer at symmetrical patterns than asymmetrical patterns, suggesting sensitivity to Symmetrical figures concepts. That said, the children in this study were ages 3.5–5.5 years old and may have received instruction on symmetry concepts in preschool. It is also important to be careful in making generalizations across studies: The "symmetry" stimuli of the Huang et al. (2018) study were quite different from the 6-panel stimuli of the original Dehaene et al. (2006) study and the 2-AFC version used here.

More generally, further empirical evaluation of the graded sensitivity proposal is a goal for future research. Although a GT concept 'popping out" can be interpreted as evidence of intuitiveness, all intuitive concepts might not necessary show this behavioral pattern. It may turn out that, *contra* to our predictions, symmetry concepts are

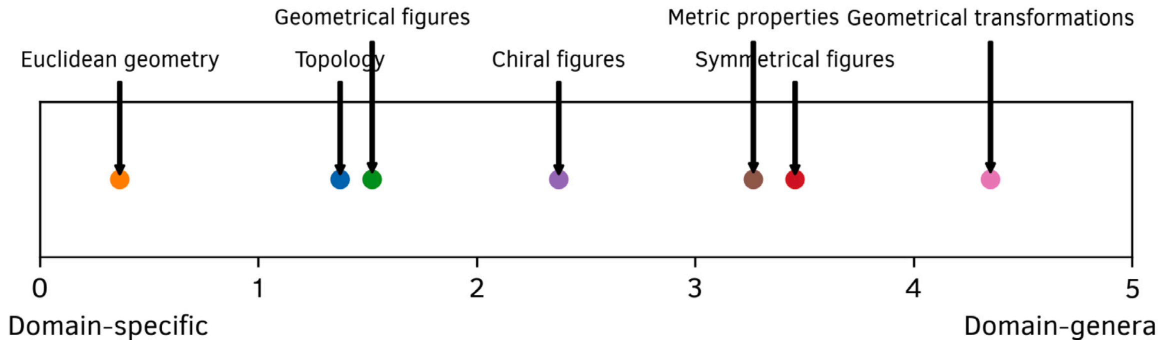

Fig. 8. Domain-centricity score for each geometric and topological class.

relatively domain-specific. This would be consistent with the proposal of Shepard (2001) that symmetry is an organizing principle of the universe, and that human cognition is tuned to it.

More generally, the proposal that some classes of GT concepts are more domain-specific is consistent with the claim that they are also more *experience-expectant* (Greenough et al., 1987), i.e., they emerge as a consequence of the normal maturational processes of organisms undergoing species-typical experiences. As such, they may require relatively little support from the environment. Another explanation is that to the degree humans share common experiences and artifacts such as tools across cultures, they may have similar experiences and acquire similar knowledge (Ferreirós & García-Pérez, 2020). By contrast, the classes of GT concepts that are more domain-general align with the claim that they are more *experience-dependent*, i.e., can develop later and are more structured by unique supports provided by the environment. An important such support is mathematics instruction through formal schooling. A natural prediction, then, is that domain-general GT concepts might be more malleable than domain-specific GT concepts, and thus might benefit more from focused training experiences. Testing this prediction is a goal for future research.

CRediT authorship contribution statement

Vijay Marupudi: Conceptualization, Methodology, Software, Investigation, Formal analysis, Visualization, Writing – original draft. **Sashank Varma:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Data availability

Data is available at the repository mentioned in the manuscript.

Acknowledgments

We thank Rebecca Schmitz for contributing to the early design of the 2-AFC task. We thank Kori Terkelson for assistance in scheduling participants for the online study. We thank Ashok Goel, Keith McGreggor, Snejana Shegheva, and Wangcheng Xu for discussions about the ideas in this manuscript. We thank Dr. Stanislas Dehaene for sharing the stimuli from Dehaene et al. (2006) and Dr. Veronique Izard for sharing the Experiment 2 adult (ages 18–25 years) data from Izard and Spelke (2009).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105331.

References

ACT, Inc. (0000). Description of Math Test.

Chiandetti, C., & Vallortigara, G. (2008). Is there an innate geometric module? Effects of experience with angular geometric cues on spatial re-orientation based on the shape of the environment. *Animal Cognition*, *11*(1), 139–146. http://dx.doi.org/10.1007/s10071-007-0099-y.

Chiandetti, C., & Vallortigara, G. (2010). Animals' representation of enclosed spaces: Evidence for use of a similar frame of reference following different disorientation procedures in the domestic chick (Gallus Gallus). *Journal of Comparative Psychology*, *124*(2), 139–146. http://dx.doi.org/10.1037/a0017013.

CollegeBoard (0000). SAT Math Test Overview – SAT Suite | College Board.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. S. (2006). Core Knowledge of Geometry in an Amazonian Indigene Group, vol. 311. (p. 4).

Descartes, R. (1701). *Regulae ad directionem ingenii*.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*(1), 1–63. http://dx.doi.org/10.1016/0004-3702(89)90077-5.

Ferreirós, J., & García-Pérez, M. J. (2020). Beyond natural geometry: On the nature of proto-geometry. *Philosophical Psychology*, *33*(2), 181–205. http://dx.doi.org/10.1080/09515089.2019.1683726.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: sketch understanding for cognitive science research and for education. *Topics in Cognitive Science*, *3*(4), 648–666. http://dx.doi.org/10.1111/j.1756-8765.2011.01149.x.

Gibson, E. (1969). *Principles of perceptual learning and development*. East Norwalk, CT, US: Appleton-Century-Crofts.

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). *Brain development and cognition: a reader, Experience and brain development* (2nd Ed). (p. 216). Malden: Blackwell Publishing.

Harris, C. R. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. http://dx.doi.org/10.1038/s41586-020-2649-2.

Heimler, B., Behor, T., Dehaene, S., Izard, V., & Amedi, A. (2021). Core knowledge of geometry can develop independently of visual experience. *Cognition*, *212*, Article 104716. http://dx.doi.org/10.1016/j.cognition.2021.104716.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in Young children. *Nature*, *370*(6484), 57–59. http://dx.doi.org/10.1038/370057a0.

Hohol, M. (2019). *Foundations of geometric cognition*. Routledge.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. http://dx.doi.org/10.1002/bimj.200810425.

Huang, Y., Xue, X., Spelke, E. S., Huang, L., Zheng, W., & Peng, K. (2018). The aesthetic preference for symmetry dissociates from early-emerging attention to symmetry. *Scientific Reports*, *8*(1), 6263. http://dx.doi.org/10.1038/s41598-018-24558-x.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. *Science*, *310*(5749), 863–866. http://dx.doi.org/10.1126/science.1117593.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. http://dx.doi.org/10.1109/MCSE.2007.55.

Izard, V., O'Donnell, E., & Spelke, E. S. (2014). Reading angles in maps. *Child Development*, *85*(1), 237–249.

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2011). Flexible intuitions of Euclidean geometry in an Amazonian indigene group. *Proceedings of the National Academy of Sciences*, *108*(24), 9782–9787. http://dx.doi.org/10.1073/pnas.1016686108.

Izard, V., & Spelke, E. S. (2009). Development of Sensitivity to Geometry in Visual Forms. *Human Evolution*, *23*(3), 213–248.

Kant, I. (1783). *Prolegomena to any future metaphysics that will be able to present itself as a science: with two early reviews of the critique of pure reason*.

Lee, S. A., & Spelke, E. S. (2008). Children's use of geometry for reorientation. *Developmental Science*, *11*(5), 743–749. http://dx.doi.org/10.1111/j.1467-7687.2008.00724.x.

Lovett, A., & Forbus, K. (2011). Cultural commonalities and differences in spatial problem-solving: A computational analysis. *Cognition*, *121*(2), 281–287. http://dx.doi.org/10.1016/j.cognition.2011.06.012.

Lovett, A., & Forbus, K. (2017). Modeling visual problem solving as analogical reasoning. *Psychological Review*, *124*(1), 60–90. http://dx.doi.org/10.1037/rev0000039.

Marlair, C., Pierret, E., & Crollen, V. (2021). Geometry intuitions without vision? A study in blind children and adults. *Cognition*, *216*, Article 104861. http://dx.doi.org/10.1016/j.cognition.2021.104861.

McGreggor, K., & Goel, A. (2013). Fractal representations and core geometry. In *Proceedings of the second annual conference on advances in cognitive systems acs, Vol. 3* (p. 19). Citeseer.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt, & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 56–61). http://dx.doi.org/10.25080/Majora-92bf1922-00a.

Pahor, A., Stavropoulos, T., Jaeggi, S. M., & Seitz, A. R. (2019). Validation of a matrix reasoning task for mobile devices. *Behavior Research Methods*, *51*(5), 2256–2267.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in Macaque area V4. *Journal of Neurophysiology*, *82*(5), 2490–2502. http://dx.doi.org/10.1152/jn.1999.82.5.2490.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance. *Brain and cognition*, *28*(1), 39–58.

R. Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raven, J. C. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford pyschologists Press.

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by Young chicks. *Perception & Psychophysics*, *57*(7), 971–976. http://dx.doi.org/10.3758/BF03205456.

Rosner, B. (2011). *Fundamentals of biostatistics (the 7th Edition)*. Boston, MA: Brooks/Cole.

Sablé-Meyer, M., Fagot, J., Caparos, S., van Kerkoerle, T., Amalric, M., & Dehaene, S. (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *Proceedings of the National Academy of Sciences*, *118*(16), http://dx.doi.org/10.1073/pnas.2023123118.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *9th python in science conference*. http://dx.doi.org/10.25080/Majora-92bf1922-011.

Sheghava, S., & Goel, A. (2020). Symmetry as an Organizing Principle for Geometric Intelligence. arXiv:2004.07879.

Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences*, *24*(4), 581–601.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. http://dx.doi.org/10.1111/j.1467-7687.2007.00569.x.

Spelke, E. S., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, *34*(5), 863–884. http://dx.doi.org/10.1111/j.1551-6709.2010.01110.x.

The pandas development team (2020). *Pandas-dev/pandas: pandas*. Zenodo, http://dx.doi.org/10.5281/zenodo.3509134.

Tommasi, L., Chiandetti, C., Pecchia, T., Sovrano, V. A., & Vallortigara, G. (2012). From natural geometry to spatial cognition. *Neuroscience & Biobehavioral Reviews*, *36*(2), 799–824.

Vallortigara, G. (2017). Comparative cognition of number and space: The case of geometry and of the mental number line. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1740), http://dx.doi.org/10.1098/rstb.2017.0120.

van der Ham, I. J. M., Hamami, Y., & Mumma, J. (2017). Universal intuitions of spatial relations in elementary geometry. *Journal of Cognitive Psychology*, *29*(3), 269–278. http://dx.doi.org/10.1080/20445911.2016.1257623.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, *47*(2), 599–604.

Weiss, Y. (2022). *High resolution time*: *W3C working draft*, W3C.

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in Macaque inferotemporal cortex. *Nature Neuroscience*, *11*(11), 1352–1360. http://dx.doi.org/10.1038/nn.2202.