

Constituency tests in human adults' language of thought for geometry

Barbu Revencu (barbu.revencu@cea.fr)

NeuroSpin Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay

Stanislas Dehaene

NeuroSpin Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay

Abstract

Humans can remember arrays or sequences of stimuli that exceed working memory limits in many domains, from auditory sequences to geometric shapes. This ability has been interpreted as evidence for language-like representations that compress stimuli into compact descriptions. We extend this evidence in the domain of geometry by showing that representations of geometric shapes are not only compressed but also syntactically structured. Experiment 1 shows that different representations can be induced for the same geometric shape, indicating structural representation. Experiment 2 shows that subparts of a shape are easier to recognize when they belong to the same subtree than when they do not, indicating hierarchical organization. Taken together, the results indicate that geometric shapes are encoded in representations that possess internal syntax, just like natural language sentences.

Keywords: geometric shapes; language of thought; syntax; structural ambiguity; Gestalt perception

Introduction

A growing body of evidence indicates that humans use language-like representations to encode many nonlinguistic stimuli, from arbitrary categories in concept learning tasks to auditory and visual sequences (Al Roumi, Marti, Wang, Amalric, & Dehaene, 2021; Wang et al., 2019; Dehaene, Al Roumi, Lakretz, Planton, & Sable-Meyer, 2022). According to a recent proposal, human adults also encode geometric shapes in a dedicated language of thought (LoT) (Sablé-Meyer, Ellis, Tenenbaum, & Dehaene, 2022). This geometry LoT, illustrated in Figure 1, consists of a handful of primitives, among which TRACE (a segment), TURN (the drawing angle), REPEAT (take a set of drawing instructions as input and repeat them n times), and SUBPROGRAM (execute one or more drawing instructions, then return to the original position). These primitives can be combined in various ways to recreate many of the geometric shapes attested across human cultures. Moreover, Sablé-Meyer et al. (2022) found that adults' encoding times, choice times, and error rates in a match-to-sample task can be predicted from a shape's minimum description length—the number of primitives in its shortest generative program (see also Leeuwenberg & van der Helm, 2013).

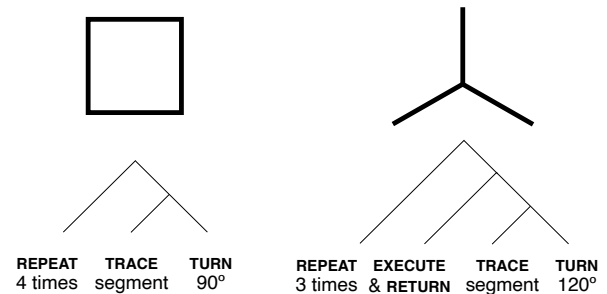


Figure 1: The main ingredients of the geometry language of thought in Sablé-Meyer et al. (2022), illustrated on two geometric figures.

Sablé-Meyer et al. (2022) used this evidence to make two claims about adults' representations of geometric shapes. The first claim concerns the *format* of representation: an LoT with semantic content (given by the primitives) and syntactic structure (given by the combinations of these primitives). The second claim concerns the *selection* among equivalent representations. For instance, a square can be represented not only by the program [REPEAT (4 times, (TRACE, TURN))] in Figure 1 but also by a program such as (TRACE, TURN, TRACE, TURN, TRACE, TURN, TRACE, TURN). However, the REPEAT program will be selected because it is more compact.

Note, however, that inferring the format of a representation from the minimum description length in the space of possible programs is highly indirect. First, this inference relies on the correlation between behavior and a numerical value without showing further evidence of the hypothesized structure. Second, format and selection are partly orthogonal: Humans could use an LoT to encode geometric shapes even if they do not always select the shortest program for a given shape.

For these reasons, we devised a series of constituency tests that allow for a more direct investigation of syntactic structure in shape representations. We present two online experiments showing, first, that adults can represent the very same geometric shape in different mental representations. Second, we show that shape subparts are easier to recognize when their segments belong to the same subtree than when they straddle different subtrees.

Experiment 1: Methods

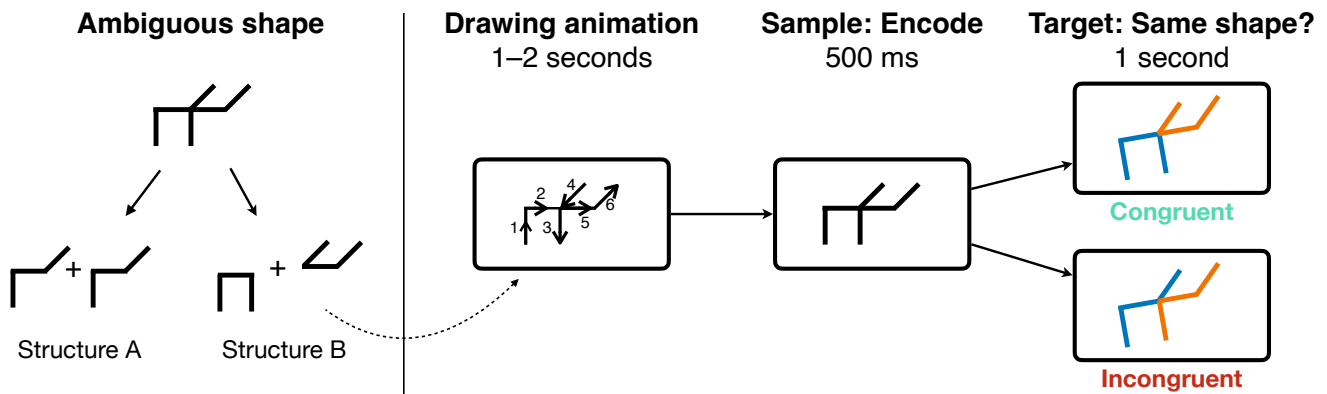


Figure 2: **Left** Example of a stimulus shape, together with its two decompositions. We induced both decompositions via drawing animations across and within participants. **Right** Task overview, illustrated with one of the two structures on the left. Trial congruence is defined by the match/mismatch between target coloring and the structure implied by the animation. Only Match targets are shown.

Experiment 1: Structural Ambiguity

In language, there are many sentences for which more than one syntactic parse is possible (e.g., *I saw [the man with the binoculars]* versus *I saw [the man] [with the binoculars]*). This ambiguity has long been used to argue that sentence representations have internal structure (Chomsky, 1967). In Experiment 1, we sought to provide an analogous test case for geometric shapes by testing whether different parses can be induced for the same surface geometric shape. If true, this would allow us to draw two conclusions. First, it would show that the internal representations of geometric shapes can be structured—if they had no structure, the same shape would always give rise to the same (flat) representation. Second, it would show that the structured representations of geometric shapes can be influenced by factors other than minimum description length (van der Helm, 2014; Sablé-Meyer et al., 2022; Feldman, 2016). Otherwise, the same shape would always give rise to the same representation (the most compact one).

Methods

Transparency and openness The hypotheses, methods, exclusion criteria, analyses, and statistical analyses were preregistered at the Open Science Framework (Experiment 1: <https://osf.io/yrkzm/>; Experiment 2: <https://osf.io/k2cx8/>). The local ethical committee approved the experiment, and informed consent was obtained from the participants at the beginning of the testing session. The data for both experiments were analyzed using R 4.4.0 (R Core Team, 2022), and the packages tidyverse 2.0.0 (Wickham et al., 2019), ggplot 3.5.1 (Wickham, 2016), and rstatix 0.7.2 (Kassambara, 2023).

Task and trial structure A schematic task depiction is shown in Figure 2 (right).¹ Participants had to decide whether two stimuli have the same shape. Each trial started with an animation that generated a line drawing (1–2 seconds) by sequentially drawing two shapes (Segments 1–3 and Segments 4–6). A short break (333 ms) was interposed between the two shapes that formed the final drawing. The last frame of the animation depicted the **sample** image participants had to memorize and was shown for 500 ms before being removed. After a short break (250 ms), a fixation circle was displayed for 500 ms. The fixation circle was then replaced by one of the **target** images. Once the target image was displayed, participants used their keyboard to indicate whether the target had the same shape as the sample, ignoring differences in color, scale, and orientation. They were given 1 second to respond. If they made a mistake or did not respond fast enough, a feedback message was displayed for 1.5 seconds. In the experiment instructions, animations were presented in a neutral way: Participants were told that they would have to remember a black-and-white shape generated by a drawing process.

Stimuli Building on the LoT program in Sablé-Meyer et al. (2022), we created 16 **sample** stimuli that could be decomposed in two ways: Either by joining two identical sub-shapes (Figure 2, left: Structure A) or by joining two different sub-shapes (Figure 2, left: Structure B). Both structures were turned into short animations depicting how the shape was drawn. For each sample shape, the two animations started from the same position and lasted equally long. The length and width of each sample varied between 90 and 180 pixels. The samples were presented at the center of the screen in an invisible bounding box of 200 × 200 pixels.

¹A sample video of the procedure is accessible at <https://vimeo.com/1045772620>.

For each sample image, we generated 8 colored **Match targets** (Figure 2, right), which had the same shape as the samples. Based on their coloring pattern, the targets could be divided in two subtypes, depending on the decomposition they reflected. For instance, the target on the top screen of Figure 2 (**Congruent** with respect to the animation) obeys structure B, while the target on the bottom screen (**Incongruent** with respect to the animation) obeys structure A. The match/mismatch between the animation and coloring is thus the independent variable, which allowed us to test whether (i) adults use structured representations to encode the sample shapes; (ii) whether animations can induce different representations for the same shape. The targets could appear in one of four color pairings (blue–orange, orange–blue, red–green, and green–red), at different scales relative to the sample image (0.92, 0.94, 0.96, 0.98, 1.02, 1.04, 1.06, or 1.08), and in different orientations (-25° , -20° , -15° , -10° , 10° , 15° , 20° , or 25°).

To create the **Deviant targets**, we modified the colored Match targets either by moving a segment to a different position or by adding a new segment.

Design The task was divided into two equal-length blocks. Across the two blocks, each sample image was present on 16 trials, evenly paired with Match and Deviant targets. In the first block, half the sample images were assigned to Structure-A animations, half to Structure-B, randomly for each participant. In the second block, the assignments were reversed. Thus, all participants saw every shape in both guises over the entire experiment.

To generate the sample–target pairings in the first block, we randomly sampled a red-and-green color pair (red–green or green–red, specifying the color of the first shape) and a blue-and-orange one (blue–orange or orange–blue). This established the eight targets that would be shown for each sample image in Block 1: 4 Match targets shown in both color pairings, 2 move-segment Deviant targets shown in one color pair, and 2 add-segment Deviant targets shown in the remaining color pair. The eight targets were evenly divided into **Congruent** and **Incongruent** ones, depending on whether their coloring reflected the same structure as the animation program of the sample. For each sample image, the 8 targets in a block were randomly assigned to a unique scale–orientation tuple. Trials were presented in a pseudo-random order, such that each sample image appeared exactly once every 16 trials. Block 2 followed the same logic but reversed the animation programs for each sample and the color pairs for the associated targets, and re-randomized the scale–orientation assignments and the order of presentation. The randomization was generated in the experiment script for each participant anew.

There were 256 test trials in total (16 sample images \times 2 animation programs \times 2 target coloring patterns \times 2 target types \times 2 color pairs). At the beginning of the experiment, participants completed a 16-trial Training block to get acquainted with the task. The Training block was identical to

the main task but used unambiguous stimuli (e.g., a concatenation of a line and a spiral).

Participants The final sample consisted of 50 English-speaking adult participants, recruited via Prolific ($M_{\text{age}} = 31.5$ years, $\text{range}_{\text{age}} = 19\text{--}62$). The sample size was chosen based on a pilot with 17 participants to detect an effect on reaction times with 80% power. Five additional adults were tested but not included in the final sample because they failed to meet the inclusion criterion.

Data exclusion Following the preregistration, we excluded the trials in which participants responded in less than 200 ms ($n = 20$ trials; 0.16%) or in which they provided no response ($n = 300$ trials; 2.34%). We also excluded the participants who scored lower than 75% in accuracy or response rate ($n = 5$). Reaction-time analyses take into consideration correct-response trials only.

Results

Participants solved the task correctly and above chance (overall accuracy: 87%, $t(49) = 58.85$, $p < .0001$). For the sake of brevity, we focus on the results for the Match trials only, as these are the most relevant to our hypothesis. As preregistered, reaction time analyses included only the trials in which participants were correct.

As predicted, incongruent Match trials, in which the coloring of the target did not correspond to the structure of the animation, were harder than congruent Match trials (Figure 3). The effect of trial type was insignificant for reaction times ($M_{\text{difference}} = 6.1$ ms, $t(49) = 1.96$, $p = .056$, 95% CI $[-0.16, 12.34]$, Cohen’s $d = 0.28$), but strong for error rates ($M_{\text{difference}} = 0.04$, $t(49) = 5.42$, 95% CI $[0.03, 0.06]$, $p < .0001$, Cohen’s $d = 0.77$). Incongruent trials were more difficult than congruent trials for 82% of the participants (41 out of 50) and for 88% of the shapes (14 out of 16).

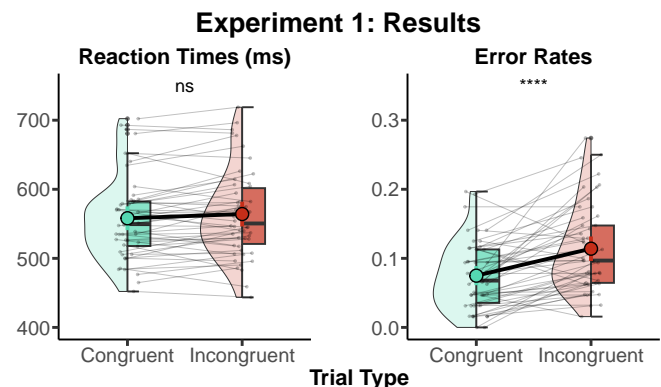


Figure 3: Congruency effect on the Match trials of Experiment 1. Gray circles and the lines connecting them represent individual averages as a function of trial type; colored circles represent group averages and 95% CIs; box plots mark the median and interquartile range; violin plots depict the shape of the distribution. **Left** Reaction times. **Right** Error rates.

To test whether our manipulation worked not only *across* but also *within* participants, we conducted an additional exploratory analysis by experimental block (also preregistered). Recall that the two animation programs for each sample shape were blocked for each participant. Thus, in each half of the experiment, any given shape was drawn by one kind of program only. Did participants' repeated exposure to the same program in the first block cause them to ignore the different program (for the same shape) in the second block? While the effects were higher in the first block (Reaction times: $M_{\text{difference}} = 11.28$ ms, $t(49) = 2.51$, $p = .016$, 95% CI [2.23, 20.32], Cohen's $d = 0.35$; Error rates: $M_{\text{difference}} = 0.05$, $t(49) = 4.15$, $p < .001$, 95% CI [0.026, 0.074], Cohen's $d = 0.59$), the incongruence effect persisted in Block 2 as well, as evidenced by participants' errors (Reaction times: $M_{\text{difference}} = 2.58$ ms, $t(49) = 0.59$, $p = .557$, 95% CI [-6.18, 11.33], Cohen's $d = 0.08$; Error rates: $M_{\text{difference}} = 0.03$, $t(49) = 3.78$, $p < .001$, 95% CI [0.013, 0.042], Cohen's $d = 0.54$). The Block \times Trial Type interaction was not significant, $F(1, 49) = 2.75$, $p = .1$, $\eta_p^2 = .05$. Thus, representations of shapes can be experimentally manipulated within participants as well—a signature of structural ambiguity.

Discussion

By using the geometric analog of ambiguous sentences, Experiment 1 showed that adults encode geometric shapes in structured representations. Moreover, Experiment 1 introduced a new Gestalt-like cue to grouping—drawing animation—that successfully manipulated the representation of the same shape across and within participants. Finally, Experiment 1 suggests that different grouping cues—drawing animation and coloring—tap onto the same internal representation. In the absence of such a process, we would not have observed any effect of trial congruence.

Experiment 2: Subtree Facilitation

In Experiment 2, we explored humans' representation of geometric shapes with a different constituency test. Inspired by the evidence that subtrees are faster to process than merely adjacent configurations in language and arithmetic (Landy & Goldstone, 2010; Nakai & Okanoya, 2018; Schneider, Maruyama, Dehaene, & Sigman, 2012), we ask whether adults are more likely to recognize that a shape is a part of the figure when it corresponds to a subtree in the geometry LoT put forth by Sablé-Meyer et al. (2022). This paradigm is closely related to the hidden-figure task in the Gestalt psychology tradition (Palmer, 1977; Reed & Johnsen, 1975; Leeuwenberg, Mens, & Calis, 1985) but adds to it in several ways. First, the sub-shapes we are investigating do not correspond to subtrees at the first level of the hypothesized representation but at the second one—they are subtrees of subtrees. Second, we use a rigorously controlled set of stimuli, such that identical targets stem from the same subtree in one sample image but straddle different subtrees in another. Third, we designed the stimuli to minimize salient Gestalt cues to parsing (e.g., no stimulus could be parsed based on closure).

Methods

Task and trial structure A schematic depiction of the task is shown in Figure 4 (right).² Participants had to decide whether a two-segment image (target) stimulus was a subpart of a previously shown six-segment image (sample). Each trial started with the sample image presented for 3 seconds. After a short break (200 ms), a fixation circle was displayed for 500 ms. The target image was then flashed for 300 ms above or below the fixation circle (200-pixel displacement). The circle and target disappeared, and participants used their keyboard to indicate whether the target was a subpart of the sample. Participants had 1 second to respond. If they made a mistake or did not respond within the one second, a feedback message was displayed for 2 seconds.

Stimuli As in Experiment 1, we created 16 sample images using a modified version of the language in Sablé-Meyer et al. (2022). Each image consisted of 6 equal-length segments and was generated by concatenating two subfigures of 3 equal-length segments each (Figure 4). In turn, each subfigure was created by one of the two signature primitives of the shape LoT in Sablé-Meyer et al. (2022)—repetitions and subprograms. Importantly, participants were **not** shown an animation or any other cue to the underlying structure of the shapes. Rather, we hypothesized that participants would decompose them in accordance with the compact LoT programs used to generate them.

Unbeknownst to the participants, the 16 sample images came in pairs (Figure 4, left). The two images in each pair were designed to contain a common 3-segment connected subfigure. This subfigure was used to generate two target Match images [(left segment + center segment) and center segment + right segment]. These two Match targets were present in both images, but differed in their status in the underlying structure of their programs: One target aligned to the tree structure of the program (**Same subtree**), whereas the other did not (**Different subtrees**).

The Deviant targets (not shown) were created by rotating the two Match targets by 90, 180, or 270 degrees. These targets were concatenated to create another 3-segment connected subfigure, which was used as the basis for the images in the second pair of each group. This allowed us to perfectly balance the targets, as the Match targets for one stimulus pair become the Deviant targets for another.

Design Unlike Experiment 1, the targets did not vary in color, scale, or orientation. However, to avoid direct matching, the targets were flashed above or below the central position where the sample image had been presented. Within each pair, samples were presented to display the common 3-segment subfigure at the same on-screen position. To fix the display coordinates, we centered the images in each pair, then moved them toward each other until the common subfigures overlapped. To equalize the distance between fixation and the

²A sample video of the experiment can be found at <https://vimeo.com/1045773638>.

Experiment 2: Methods

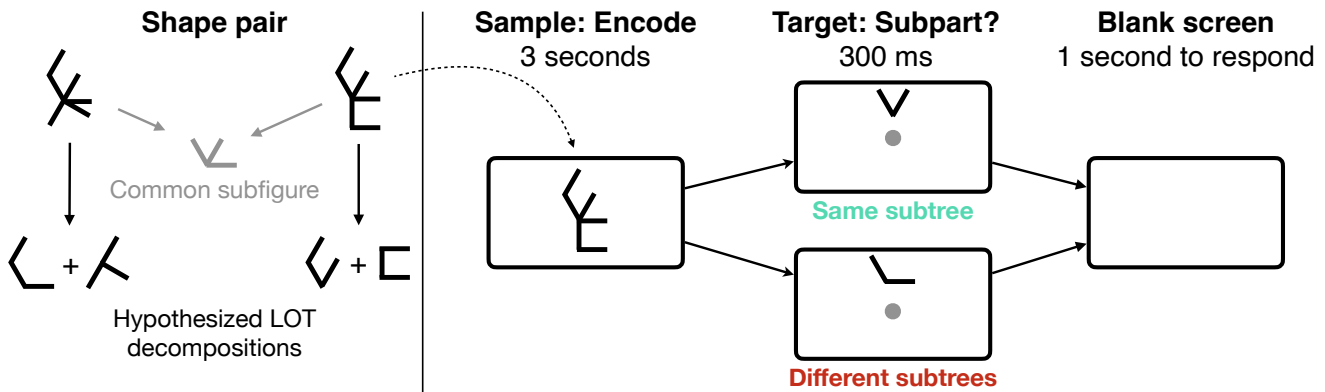


Figure 4: **Left** Example of two paired stimuli shapes in Experiment 2, together with their hypothesized structure (*not* shown to participants). The stimuli have a 3-segment common subshape that was split to generate two 2-segment targets. The targets contain segments that belong to the same subtree or to different subtrees, counterbalanced for every pair. **Right** Task overview, illustrated with one of the two shapes on the left.

3-segment subfigure, we presented each sample (and corresponding target) rotated by 180 degrees as well. Within each pair, every target appeared exactly twice at each of the two screen positions and in each of the two orientations.

There were 128 test trials in total (16 sample images \times 4 targets \times 2 orientations). The order of trials was pseudorandomized such that all samples appeared exactly once every 16 trials. As in Experiment 1, participants completed a 16-trial Training block with unrelated stimuli before the main task.

Participants The final sample consisted of 50 English-speaking adult participants, recruited via Prolific ($M_{\text{age}} = 28$ years, $\text{range}_{\text{age}} = 18\text{--}66$). The sample size was chosen based on a pilot with 9 participants to detect an effect on reaction times with 80% power. Four additional adults were tested but not included in the final sample because they failed to meet the inclusion criterion.

Data exclusion As preregistered, we excluded from the final analysis the trials in which no response was provided ($n = 482$ trials, 7.5%), as well as the participants who provided responses on less than 50% of the Match trials ($n = 4$). We did not include an accuracy threshold, as we knew from the pilot that this task would be much harder than the task in Experiment 1. We did not exclude any fast-response trial, as participants were not allowed to respond in the 300-ms time window during which the target was displayed.

Results

Participants solved the task correctly and above chance (overall accuracy: 55%, $t(49) = 4.42$, $p < .001$). We focus on the results for the Match trials, as these are directly relevant to the question under investigation. As predicted, accuracy was higher on same-subtree Match trials than on different-subtrees Match trials (Figure 5, left). This subtree-facilitation

effect was not present in reaction times ($M_{\text{difference}} = 2.23$ ms, $t(49) = 0.27$, $p = .787$, 95% CI $[-14.2, 18.7]$, Cohen's $d = 0.04$) but was robust in error rates ($M_{\text{difference}} = 0.08$, $t(49) = 4.59$, $p < .001$, 95% CI $[0.05, 0.12]$, Cohen's $d = 0.65$). The subtree facilitation effect on errors was present in 74% of the participants (37 out of 50) and for 81% of the shapes (13 out of 16).



Figure 5: Subtree facilitation effect on the Match trials of Experiment 2. **Left** Reaction times. **Right** Error rates.

Discussion

Experiment 2 provides a second piece of evidence for constituent structure in geometric shape representations. Participants were consistently better at detecting the subparts of an encoded image when the subpart segments corresponded to (sub)subtrees in the image than when they straddled different subtrees.

General Discussion

The presence of both structure and hierarchy in representations provides compelling evidence of an underlying internal syntax. Experiment 1 provided evidence of structure by highlighting that structurally ambiguous shapes can be parsed in different ways by human adults. Experiment 2 strengthened the evidence for structural representations, while also providing evidence of hierarchy. Participants were better at detecting subparts of shapes when the subparts belonged to the same subtree than when they did not.

Tree structure versus chunk concatenation It can be argued, however, that chunk concatenation (e.g., Nassar, Helmers, & Frank, 2018) can explain the data without resorting to syntactic trees. According to the chunking account, the two sub-figures on the bottom left of Figure 4 would not have been further subdivided into component elements but encoded as primitives. These primitives could have then been concatenated without any nesting to recreate the shapes on the upper left of Figure 4. The chunking account would thus accept our claim about structure but reject the claim about hierarchy.

Experiment 2 speaks against this alternative in two ways. First, a chunk-based account would need to come up with an explanation of why the shapes were chunked the way they were over other potential decompositions. (Recall that participants were never exposed to individual chunks at any point in the experiment.) On the syntactic account, the decomposition patterns are explained by the geometry LoT hypothesized to underlie the shape representations. On the chunking account, these patterns would receive a post-hoc explanation. Second, we did not pit chunks against non-chunks but sub-subtrees against segments that did not belong to a single subtree. In other words, the targets did not correspond to chunks but to chunk subparts. But if the chunks had been represented as (structure-less) primitives, detecting *their* subparts should not have been easier at all.

That said, chunks might be constructed by an initial syntactic process that is eventually discarded in the outcome representation. A template-matching process could then be employed to check whether the two-segment targets are part of either of the two chunks. Therefore, we cannot definitively rule out the chunk-based account at this stage. To achieve this, we would need evidence that aspects of human performance vary continuously with aspects of the hierarchical structure of the hypothesized representation (e.g., processing time with syntactic tree depth). We are currently developing a paradigm that would allow us to test this, thereby providing a more stringent test between the two competing accounts.

Automaticity A further question our study leaves open is the automaticity of the LoT representations. One possibility is that such representations are recruited only when the current task requires them. In both experiments, the memorization task may have prompted participants to build a decomposed, compressed representation for efficient storage.

Alternatively, shapes may be encoded in syntactic representations automatically, even when the task does not prompt them to. Indeed, in Experiment 1 it is not clear why participants were influenced by the animation at all, especially in the second half of the experiment, where the drawing animations switched the implied structure of every shape. By that time, participants could have learned the shapes, encoded them in their preferred representation, and ignored the animations altogether. Yet the animations still influenced their behavior. To tease these possibilities apart, a possible solution would be to remove the need for memorization altogether and test, for instance, for a same-subtree advantage on attention (e.g., Duncan, 1984) using the stimuli in Experiment 2. If attention can move faster within a subtree than across subtrees even when stimuli need not be remembered, this will favor the automaticity of syntactic representations.

Perception versus cognition Finally, our results raise interesting questions about the nature of the geometric shape representations: Are they driven by perceptual processes or by higher-level ones? In other words, does vision itself furnish the geometry language of thought? Note that while the stimuli in our tasks are fully visual, one cannot conclude that the representations themselves are produced by vision. Instead, it could be that the encoding of shapes relies on visual information without being a visual process itself.

We can think of two directions of investigation for approaching this question. One possibility would be to test convolutional and transformer networks on the same task, as these networks model several aspects of the human visual system reasonably well (Tuli, Dasgupta, Grant, & Griffiths, 2021; Eickenberg, Gramfort, Varoquaux, & Thirion, 2017). Another option would be to turn the visual tasks in Experiments 1 and 2 into tactile tasks (e.g., 3D contours that can be traced by hand) for blindfolded and/or congenitally blind individuals. Analogous results in a different sensory domain would show that these representations are amodal and would rule out vision as their primary source.

If it turns out that geometric shape representations do not originate in vision, after all, but in a higher cognitive process, this would have important implications for the Gestalt tradition in object perception. In particular, it would raise the possibility that many of the Gestalt principles and findings concern geometric shapes (rather than objects) and are driven by cognitive processes (rather than perceptual ones).

Acknowledgments

This work was supported by the Fyssen “Syntax–semantics mappings in the shape LoT” grant to Barbu Revenu, by the ERC MathBrain grant (ERC-2022-ADG101095866) to Stanislas Dehaene, and by INSERM, CEA, Collège de France, and Université Paris-Saclay.

References

Al Roumi, F., Marti, S., Wang, L., Amalric, M., & Dehaene, S. (2021, 8). Mental compression of spatial sequences in

- human working memory using numerical and geometrical primitives. *Neuron*, 109(16), 2627–2639. doi: 10.1016/j.neuron.2021.06.009
- Chomsky, N. (1967). *Aspects of the theory of syntax*. MIT Press.
- Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sable-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Science*, 26(9), 751–766. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/35933289> doi: 10.1016/j.tics.2022.06.010
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4), 501–517. doi: 10.1037/0096-3445.113.4.501
- Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017, 5). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, 152, 184–194. doi: 10.1016/j.neuroimage.2016.10.001
- Feldman, J. (2016). The simplicity principle in perception and cognition. *WIREs Cognitive Science*, 7(5), 330–340. doi: 10.1002/wcs.1406
- Kassambara, A. (2023). *rstatix: Pipe-friendly framework for basic statistical tests*.
- Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Quarterly Journal of Experimental Psychology*, 63(10), 1953–1968. doi: 10.1080/17470211003787619
- Leeuwenberg, E., Mens, L., & Calis, G. C. (1985). Knowledge within perception: Masking caused by incompatible interpretation. *Acta Psychologica*, 59, 91–102.
- Leeuwenberg, E., & van der Helm, P. (2013). *Structural Information Theory: The Simplicity of Visual Form*. Cambridge University Press.
- Nakai, T., & Okanoya, K. (2018, 12). Neural evidence of cross-domain structural interaction between language and arithmetic. *Scientific Reports*, 8(1). doi: 10.1038/s41598-018-31279-8
- Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018, 7). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological Review*, 125(4), 486–511. doi: 10.1037/rev0000101
- Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441–474.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reed, S. K., & Johnsen, J. A. (1975). Detection of parts in patterns and images. *Memory & Cognition*, 3(5), 569–575.
- Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2022, 12). A language of thought for the mental representation of geometric shapes. *Cognitive Psychology*, 139, 101527. doi: 10.1016/j.cogpsych.2022.101527
- Schneider, E., Maruyama, M., Dehaene, S., & Sigman, M. (2012, 12). Eye gaze reveals a fast, parallel extraction of the syntax of arithmetic formulas. *Cognition*, 125(3), 475–490. doi: 10.1016/j.cognition.2012.06.015
- Tuli, S., Dasgupta, I., Grant, E., & Griffiths, T. L. (2021). Are convolutional neural networks or transformers more like human vision? *arXiv preprint arXiv:2105.07197*.
- van der Helm, P. A. (2014). *Simplicity in vision*. Cambridge University Press. doi: 10.1017/CBO9781139538268
- Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., ... Dehaene, S. (2019, 2). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage*, 186, 245–255. doi: 10.1016/j.neuroimage.2018.10.061
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. doi: <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... Yutani, H. (2019, 11). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. doi: 10.21105/joss.01686