Comparing Geometric Shape Representations in Humans and Baboons: A Language of Thought Perspective

Mathias Sablé-Meyer*[1,2,3], Joel Fagot[4,5,6,7], and Stanislas Dehaene[1,2]

[1]Collège de France, Université Paris-Sciences-Lettres (PSL), 11 Place Marcelin Berthelot, 75005 Paris, France
[2]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France
[3]Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London W1T 4JG, UK
[4]Aix Marseille Univ, CNRS, LPC, Marseille, France
[5]Aix Marseille Univ, ILCB, Aix-en-Provence, France
[6]Station de Primatologie Celphedia, CNRS, Rousset, France
[7]Aix Marseille Univ, CNRS, CRPN, Marseille, France
*Corresponding author. Email: mathias.sable-meyer@ucl.ac.uk

Abstract

In various cultures, across history and at many different spatial scales, humans produce a rich variety of geometric shapes. Recent work has put forward a concrete proposition for a Language of Thought (LoT) underlying the mental representation of geometric shapes in contemporary humans. Initial experiments, based on a comparison of baboon and human performance in an intruder task, suggested that this ability could be unique to humans. Here, to deepen our understanding of the evolutionary origins of geometric representations, we compared humans and baboons (*Papio Papio*) in a delayed geometric match-to-sample task with a broad array of shapes. Crucially, the presentation speed was manipulated so that, on slow-paced trials, animals were given more time to reflect on the shapes. The shapes were sampled from our proposed LoT and spanned a range of predicted geometric complexity under that model. Although the overall pattern of behavior was still strikingly different in humans and baboons, when presentations were slower, we did find a small contribution of the LoT representations in baboons, although weaker than in humans. In both species, longer looking time increased the effect of the LoT-based predictor. While humans used the ability to self-pace to modulate their looking time, thus benefitting from longer exposure times for more complex shapes and rendering this geometric complexity effect visible, baboons did not.

Introduction

Contemporary humans produce drawings and fill out missing parts from partial schematics depictions, but the same behavior has not been successfully prompted in great apes; and even stylized or impoverished images elicit very different behavior in apes compared to humans (Close & Call, 2015; Saito et al., 2014; Tanaka, 2007). Moreover, the human competence for symbolic, non-representative productions is not new: there is evidence of symbolic engravings from Homo erectus dating back ~500.000 years, with many other prehistoric evidence at different periods and locations (d'Errico et al., 2018; Henshilwood et al., 2018; Joordens et al., 2015; Texier et al., 2010). These productions are recognizable as intentional to contemporary humans because they have a lot of internal structure, often featuring repetitions of motifs as well as distinct geometric properties such as right angles and parallel lines. Such geometric signs are sometimes referred to as symbolic because they do not resemble real-world objects (although they might have been intended to refer to them). While recent work suggests that macaques sometimes produce graphic patterns with some geometric regularities (Sueur, 2025), these productions appear to lack the systematic, intentional and varied features that prehistoric and contemporary human productions systematically exhibit.

In previous work, we found that, when tested for the perception of geometric regularities in quadrilaterals under the very same conditions, baboons and humans performed very differently, with only humans attending to geometric regularities such as parallels, right-angles or equal sides (Sablé-Meyer et al., 2021). We therefore suggested that some aspects of mental representation of geometric shapes may be unique to humans and may therefore shed light on the emergence of geometric abilities and the human cognitive singularity (Dehaene et al., 2022) – all the while stressing that more data was needed.

The goal of the present study is to provide more evidence on the respective abilities of baboons and humans in tasks with geometric shapes, in order to shed light on the question of the emergence of symbolic behavior. Do humans, but not non-human primates, produce geometric engraving because of a lack of representational ability, which would indicate that this ability is evolutionarily recent? Or are there other factors at play, in which case we might be able to find evidence for some degree of symbolic representation in non-human primates?

A successful formal theory of the representational abilities in contemporary humans provides a useful benchmark against which to evaluate prehistoric findings or contemporary non-human primate abilities. Here, we rely on a formal description of the human representation of geometric shapes within the Language of Thought (LoT) framework (Sablé-Meyer et al., 2022) which we describe in a first section. We used stimuli derived from this model to probe the representational abilities of baboons (*Papio Papio*) and probed how they fare on a test for the mental representations of geometric shapes under different time constraints. First, we developed a training regime designed to ensure that the delayed-match-to-sample task was mastered independently of the stimuli. Then, we collected choice and response time in two experiments in baboons where the encoding time was varied (self-paced in 12 baboons; fixed and comparatively longer in 9 baboons that were part of the first experiment). Finally, we collected data in a similar fast-paced experiment in 560 humans: together with data from previously published work in a similar experiment with self-paced timing in humans (Sablé-Meyer et al., 2022), these data allowed us to investigate the effects of encoding time and population on behavior in this task.

We found that while the evidence for a LoT representation was much stronger in humans than in baboons, geometry regularity did influence behavior in both populations, particularly at longer encoding times. However, the baboons did not chose to self-pace their responses with longer looking times, which might explain why there was only a very small effect of the LoT representation on their behavior. Conversely, not only did humans self-pace their responses at longer looking times, but they adapted their looking times to the difficulty of a given trial in direct proportion to the predictions of the LoT model.

Language of Thought and Program Induction

We previously proposed that the mental representation of geometric shapes in humans is best described by mental expressions or "mental programs" in a Language of Though (LoT) for geometry (Fodor, 1975; Leeuwenberg, 1971; Sablé-Meyer et al., 2022). For example, the mental representation of a square might look like "repeat the following 4 times: [draw a line, then turn by 90°]". This representation attempts to capture the high degree of compositionality and productivity of human geometric shape representations, which include concepts such as "a square of circles". The theory postulates that the mental representation that people have of a shape is the shortest program in this internal language, and that the cognitive complexity of a shape depends on the size of this minimal program. Therefore, we predict that the Minimum Description Length (MDL) of a shape should be a major predictor of participants' performance in tasks involving these shapes.

This theory supposes that when humans perceive a geometric shape, they infer the "mental program" or the sequence of mental operations that could have drawn it. Such an inference is known as the program induction problem (Gulwani et al., 2017; Rule et al., 2020). It is a notoriously hard problem because the number of programs of length n increases exponentially with n, making systematic search in the space of programs difficult for anything but trivial program lengths. In the present work, we do not address how the program induction steps are performed, but merely assume that participants build the appropriate representations; however, induction algorithms that can apply to geometric shapes have been proposed (Ellis et al., 2023; Sablé-Meyer et al., 2022).

Shape Complexity and Minimal Description Length

To make things more concrete, we will briefly describe the language here; more details can be found in Sablé-Meyer et al. (Sablé-Meyer et al., 2022). The goal when designing the language was to capture simple shapes as attested in various cultures and (pre-)historic period with short programs. To avoid increasing the expressivity of the language, which would make every shape have a simple representation, we symmetrically tried to ensure that short programs generate only shapes that are simple for humans.

Our language is conceptually similar to the Logo computer language (Abelson et al., 1974) and to other similar proposals (Leeuwenberg, 1971) in that some instructions move a pen on paper, either tracing or not, while other instructions organize the control flow. The language comprises three categories of instructions. First, control primitives organize the execution of other programs. Our language has three control primitives: *Repeat*, which makes it possible to repeat a program n times (n=2 by default); *Concatenate*, which allows to chain two programs, and *Subprogram*, which allows for temporarily executing another program by calling it from within the parent program. The generative grammar is illustrated in Fig. 1A.

Second, the language comprises drawing primitives. The pen has an internal heading direction, and three primitives can move it : *Move*, which sends the pen ahead a certain distance without tracing, *Turn*, which changes the heading without moving, and *Trace*, which takes several parameters (the pen's speed, acceleration, turning speed, and a duration) and integrates over these parameters while tracing the corresponding curve.

Finally, since some of these primitives require numbers, the language includes a minimal mechanism for arithmetic. As a starting point, we simply included three mechanisms to build them from number 1 : the ability to increment any number by one, the ability to change the sign of a number, and the ability to take the fraction of two integers. In our language, ¾ is therefore represented as frac(1+1+1, 1+1+1+1) and has a cost corresponding to the number of operations required to construct it, here 13 ("frac", seven "1", and five "+").

In our previous work (Sablé-Meyer et al., 2022), we tested the cognitive plausibility of this language by sampling shapes of different MDL (Figure 1B) and testing the prediction that the human capacity to memorize and recognize them should decrease monotonically as MDL increases. We used a self-paced delayed-match-to-sample task. On each trial, participants first saw a target shape and studied it for as long as they thought that they needed to memorize it. The shape appeared when they pressed a key and disappeared upon release, providing us with a first measure of subjective complexity which we called the "encoding time". Once the shape disappeared, there was a 2s delay with an empty screen (in the experiments reported here this delay varies between experiments), then a choice screen with six choices shuffled in six possible fixed locations, including the target shape and five distractors. Participants had to select the target shape, and we measured their accuracy and choice time. We found that all three dependent measures (self-paced encoding time, choice time, and choice accuracy) were well predicted by the shapes' MDL, even after accounting for other visual properties of shapes such as gray level (average pixel value, or luminance) and visual frequency.

In the present work, we extend this work to baboons to try to better understand the evolutionary origins of the representation of geometric shapes as mental programs.

Experiment 1: Baboons, Self-Paced

We first replicated, in baboons, the exact method that we used in human adults except for a different delay duration. Thus, we designed a training procedure to introduce the baboons to a self-paced match-to-sample task robust to changes in size and number of distractors. We confirmed their

A

```
Program :=
  | Program ; Program                        Concatenate : run one program and then another
  | Repeat([Int=2]) { Program }              Repeat a program a certain number of times
  | Subprogram { Program }                   Execute a program, then restore the original state
  | Trace([t=Int=1],                         Trace a curve by moving according to the parameters
          [speed=Num=+1],
          [acceleration=Num=+0],
          [turningSpeed=Num=+0])
  | Move([t=Num=+1])                         Move a certain distance without tracing anything
  | Turn(angle=Num)                          Rotate the current heading
Int :=
  | one                                      Number 1
  | Next(Int)                                Successor function
Num :=
  | +Int     | -Int                          Return a signed number
  | +Int/Int | -Int/Int                      Return the signed fraction of two integers
```

B

MDL	Image	Nex	Ni2	Ni3	Ni4+	Closure	Ndis
1		2	0	0	0	0	1
2		0	0	0	0	1	1
3		2	0	0	0	0	1
3		4	0	0	0	0	2
3		1	0	1	0	1	1
4		0	0	0	1	1	1
4		2	0	0	1	1	1
4		8	0	0	0	0	4
4		2	1	0	0	0	1
4		2	0	0	0	0	1
4		2	0	0	0	0	1
4		0	0	0	0	1	2
5		2	1	0	0	0	1
5		2	0	0	0	0	1
5		0	0	0	0	1	4
5		0	0	0	2	1	1
5		0	4	0	0	1	1
5		2	0	0	0	0	1
5		0	0	0	1	1	1
6		1	0	1	0	1	1
6		1	0	1	0	1	1
6		0	0	0	5	1	1
6		6	0	0	0	0	3
6		2	1	0	0	0	1
6		2	0	0	0	0	1
6		2	0	0	0	0	1
7		0	2	0	0	1	1
7		2	1	0	0	0	1
7		2	3	0	0	0	1
7		4	0	0	0	0	2
7		2	0	0	0	0	1
7		2	1	0	0	0	1
7		0	0	0	0	1	1
8		2	1	0	1	1	1
8		1	0	1	0	1	1
8		1	0	1	0	1	1
8		2	0	0	0	0	1
8		2	0	0	1	1	1
8		1	0	1	0	1	1
8		2	0	0	0	0	1
9		1	0	1	0	1	1
9		2	0	0	1	1	1
9		0	0	0	4	1	1
9		1	0	1	0	1	1
9		4	0	0	0	0	2
9		4	0	0	0	0	2
9		2	0	0	0	0	1
10		2	3	0	3	1	1
10		2	1	0	0	0	1
10		2	0	0	0	0	1
10		0	2	0	0	1	1
10		2	1	0	0	0	1
10		2	2	0	0	0	1
10		0	0	0	3	1	1
11		4	0	0	0	0	2
11		1	0	1	0	1	1
11		2	0	0	3	1	1
11		2	1	0	1	1	1
11		2	0	0	0	0	1
11		2	0	0	0	0	1
11		4	0	0	0	0	2
12		0	0	1	5	1	1
12		2	2	0	1	1	1
12		0	0	0	9	1	1
12		2	2	0	0	0	1
12		0	0	0	8	1	1
12		2	4	0	0	0	1
12		1	0	0	2	1	1

Fig. 1. A. Generative grammar of our proposition for a language of thought for geometric shapes. We separate primitives that generally structure the program (top rows) from primitives that move the pen and trace the shape (middle rows) and primitives related to numbers (bottom rows). Adapted from (Sablé-Meyer et al., 2022) **B.** List and visualization of all of the geometric shapes ranked by MDL, together with some of the regressors used in the various analyses throughout the article: number of 4-or-more-way intersections (Ni4+), number of angles (Ni2), whether the shape is closed or not (Closure), number of 3-way intersections (Ni3), number of extremities (Nex) and number of disconnected parts (Ndis). Two additional continuous regressors are not shown in the table: gray level (luminance, called Greys) and spatial frequency (Spatial Freq.). Adapted from (Sablé-Meyer et al., 2022).

understanding of the task on a novel set of stimuli, before evaluating their performance on our geometric stimuli.

Method

Participants

Participants were 25 Guinea baboons (*Papio Papio*, 18 females, age range 1.5-23 years, mean age 11 years) from the CNRS primate facility (Rousset-sur-Arc, France). Baboons lived in a 700m^2 outdoor enclosure with access to indoor housing and had a permanent access to ten Automated Learning Devices for Monkeys (ALDM, (Fagot & Bonté, 2010) equipped with a 19-inch touch screen and a food dispenser. Note that the baboons' environment contains a mixture of natural features (e.g. trees, congeners) and artificial tools and buildings with rectangular shapes (e.g. prefabricated rooms, testing booths, computer screens, etc).

A key feature of ALDM is a radio-frequency identification (RFID) reader that can identify individual baboons through microchips implanted in their arm (Fagot & Bonté, 2010). The baboons therefore participated in the research at will, without having to be captured, as the test programs could recognize them automatically. The experiment was controlled using EPrime software (Version 2.0, Psychology Software Tools, Pittsburgh). Ethical Standards: the baboon experiment received ethical approval from the French Ministry of Education (approval APAFIS 2717-2015111708173794 v3).

Twelve baboons were able to successfully learn the delayed match to sample task all the way to the geometric stimuli. Others were removed from this project at various stages of the learning process because they were not engaged enough, often due to poor performance leading to low reward rate.

Procedure and Stimuli

Procedure The task structure is shown in Fig. 2A (training) and Fig. 3A (geometry trials). Each trial started with a waiting image of a yellow cross at the bottom of the screen. When baboons were ready to start a trial, they touched it. Upon touching it, the target stimulus for this trial appeared centered on screen, on the bottom half of the screen. As soon as the target shape was touched, it disappeared. Unlike in our previous experiment (Sablé-Meyer et al., 2022), as soon as the target disappeared, we displayed the choice screen with up to six shapes. To prevent the subject's hand, which just touched the target stimulus, from occluding specific stimuli in the choice screen, stimuli are presented on a semi-circle centered on the target shape in one of six possible locations equidistant from the center, with an angle between the center of two adjacent shaped and the center of the previously displayed target shape of 36°. In all trials, each stimulus was randomly positioned to one of these locations.

Baboons were rewarded for clicking on the stimulus that was identical up to scaling to the target shape displayed previously. For training purposes, trials were presented in blocks with increasingly challenging trial types (see training procedure in Fig. 2). There were eight consecutive training blocks with a first set of non-geometric stimuli, followed by one step with a different set of non-geometric stimuli to test for an understanding of the task beyond our training set of stimuli (generalization), and finally the block with geometric shape trials. When they performed 80% correct trials on a given block, they moved on to the next step, otherwise they performed an additional block of the same step. Each training block comprised 108 trials (see below), and each test block 68 trials, one per shape used in our human experiment (seven shapes for each MDL ranging from 4 to 12, plus one shape with MDL=1, one with MDL=2, and three with MDL=3). At the end of each block, behavior was evaluated and baboons were moved to the next block (or, eventually, to another unrelated experiment) when they reached 80% success on a given block. Some baboons were put back on this task a few days later, in which case they were restarted at the generalization step in order to ensure that they remembered the task well before collecting more data on the geometry stimuli.

Stimuli **Training stimuli** belonged to three categories, as shown in Fig. 2A: clipart of fruits, patches of colors, and single letters in white on black background. There were 12 items in each category, and the match-to-sample was only performed with distractors from the same category in order to force the baboons to pay attention to the task. Out of the 12 items, 6 were used for the consecutive training steps, and 6 were saved for the "Generalize" step in order to confirm that the subject understood the match-to-sample task and could generalize beyond the specific training stimuli.

Geometric shape stimuli were generated by enumerating geometric shapes from our language by increasing order of MDL, as previously described (Sablé-Meyer et al., 2022). Since, the number of programs increases exponentially with their size, we could only reasonably enumerate all programs up to length 12. Since some shapes could be generated by multiple programs of different complexity, we selected, for each shape, the program of lowest complexity and removed copies. Finally, we used all shapes from MDL 1, 2 and 3 as there were few of them, and we randomly sampled 7 of them for MDLs 4 to 12. Detailed descriptions of the procedure for generating the stimuli can be found in Sablé-Meyer et al. (Sablé-Meyer et al., 2022) together with summary statistics of some of the features of these shapes (see Fig. 1B for a description of the shapes' properties).

In order to make the task sufficiently difficult, the distractors in the main task were not chosen completely randomly. Of the five distractor items on-screen during the choice period, one was selected as being the closest to the target among all available shapes, according to a metric based on the Euclidean distance between their internal representations in the last layer (IT) of a convolutional neural network of object recognition, CORnet (Kubilius et al., 2018). We used this as a way to approximate the visual processing of the ventral pathway, and therefore display a distractor that should be visually quite confusable with the target according to this metric. Similarly, another distractor was selected as the one with the closest overall gray level. The three other distractors were chosen without replacement amongst the remaining shapes.

The different stimuli forming the choice display were arrayed in a semi-circle (see figure 2B), thus ensuring that the correct choice was always at a different location and at a fixed distance from the stimulus previously presented in the target screen. To further block low-level matching strategies, the size of the target was twice that of the items in the choice screen—we gradually trained subjects to ignore size in several steps, see below. Because some shapes had rotational invariance while others did not (e.g., circle versus segment), no rotation was performed with the stimuli (unlike in Sablé-Meyer et al., 2021).

Training Scheme

To train animals, they went through a series of training steps where the task became increasingly similar to the test task, starting with a simple version with a high reward rate even for poor performance in order to encourage animals to keep trying and learn the task. Each baboon was moved onto the next stage whenever their success rate exceeded 80% on a block of 108 trials: six repetition of each of three visual stimuli types (cliparts, colors, and letters in each of the six possible intruder locations, 6x3x6=108). The steps, schematized in Fig. 2B, ramped up in difficulty along two successive dimensions. First, the sample was presented with two distractors (a total of three possible choice items, thus ensuring a minimal reward rate of 33% even for chance performance) and we added more distractors one by one until we reached the total of 5 distractors and therefore six items on the choice screen. Then we progressively increased the size of the target from a scaling of 100% to a scaling of 200% in a total of four steps (125%, 150%, 175%, 200%). By the end of such a training procedure, the subjects should have understood the task, expected the target to be twice the size of the items to choose from, and known that they would have to look for the target within six possible shapes to choose from.

To confirm that baboons understood the structure of the task beyond the set of items used for training, we added an additional generalization step wherein all the stimuli were swapped with brand new stimuli never seen before, although superficially similar to the training ones, before switching to the target geometric shapes.

Data analysis

We report results from the 12 subjects that completed the entire training regime and the geometry block. We collected several dependent variables: the "encoding time", which corresponds to the duration between the click on the yellow cross and the click on the target stimulus; the "choice

step	chance	estimate	conf.int	statistic	df	p.value
1	1/3	0.73	[0.67, 0.79]	14.75	11	$< .001$
2	1/4	0.82	[0.77, 0.87]	25.63	11	$< .001$
3	1/5	0.80	[0.76, 0.84]	33.31	11	$< .001$
4	1/6	0.79	[0.76, 0.82]	40.40	11	$< .001$
5	1/6	0.76	[0.71, 0.81]	26.35	11	$< .001$
6	1/6	0.73	[0.70, 0.76]	41.76	11	$< .001$
7	1/6	0.71	[0.66, 0.75]	27.67	11	$< .001$
8	1/6	0.72	[0.68, 0.75]	32.58	11	$< .001$
9	1/6	0.61	[0.58, 0.64]	34.18	11	$< .001$
10	1/6	0.46	[0.42, 0.51]	15.59	11	$< .001$

Table 1: Average success rate on the first block of each step, and comparison to chance level. Steps 1–4 are training steps with increasing number of distractors, steps 5–8 are training steps with increasing size of the target, step 9 is the generalization to new stimuli, and step 10 is generalization to geometric shapes

time" which corresponds to the duration between the appearance of the choice stimuli and a click on one of them; and finally the identity of the chosen stimulus which we turn into a correct/incorrect independent variable. Additionally, to make comparison with humans more direct, we estimated the LISAS score (Linear integrated speed–accuracy score). The formula is $LISAS = RT_j + \frac{S_{RT}}{S_{PE}} \times PE_j$ with $\mathrm{RT_j}$ and $\mathrm{PE_j}$ the choice time and proportion of error in condition j, and $\mathrm{S_{RT}}$ and $\mathrm{S_{PE}}$ the overall standard deviations of choice time and proportion of error respectively (Liesefeld & Janczyk, 2019; Vandierendonck, 2017, 2021).

Each trial was characterized by the target shape and the set of distractors: each target shape was coded as a factor which was used as independent variables in some of our analyses. Additionally, each geometric stimulus possessed several properties: of main interest to our questions, the shape's MDL which we included as a numerical predictor in our analyses. Since we expected other strategies to be operative when performing a match-to-sample (for example, one can reject distractors based on their luminance without resorting to a rich LoT-like representation) and these other strategies partially correlate with MDL, we also included numerical estimates of the visual properties that support these strategies as additional regressors in some of our analysis—borrowing from previous work (Sablé-Meyer et al., 2022). To control for low-level visual properties, we computed and used as regressors the luminance (Grey) and spatial frequency (Spat. Freq.) of each shape. Additionally, to take into account other visual properties of shapes which may correlate with MDL, but ultimately are not as fine-grained as our proposition, we manually counted and included as regressors the number of extremities (Nex), intersections where two lines meet (singularities, variable Ni2), intersections where three lines meet (Ni3), intersections where four or more lines meet (Ni4+); we also included topological properties such as connectedness (number of disconnected parts, or Ndis), and finally the presence of a closed shape or not (closure) (Chen, 1982; Roelfsema & Singer, 1998; Wagemans et al., 2012).

Results

Training and Generalization

Fig. 2C shows the average accuracy across primates for each training step, separately for their first block and their last block. Note that chance level is not constant in these plots: since we slowly increased the number of distractors, chance was 1/3 in the first block, then 1/4, 1/5 and finally 1/6 for the rest of the experiment. At each step, some learning was required–however, even during the first block of each step, performance was already better than chance across subjects as shown in table 1.

Baboons also generalized immediately to new stimuli, since their performance was better than chance on the first block of the "generalizing" step (average success = 0.61, chance = 0.167; see Table 1.) Since blocks contained several repetitions of the same target stimulus, we also looked at

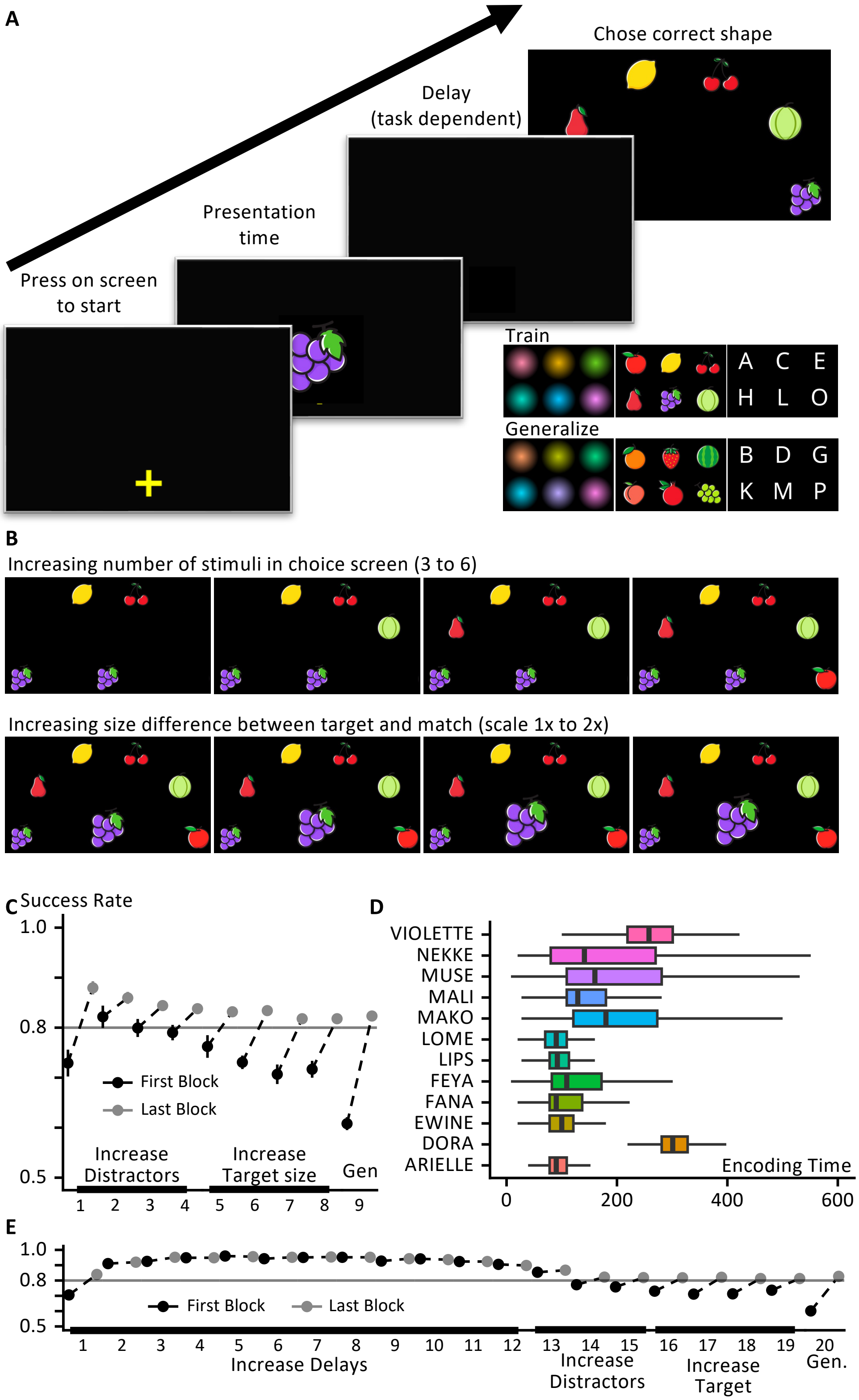

Fig. 2. A. Training stimuli and procedure. Pressing the yellow cross started a trial and displayed a target, which disappeared when clicked or after a fixed delay depending on the condition. An empty screen of fixed duration (or no empty screen in Exp. 1; see method) followed, and finally the choice screen. The 36 stimuli used for training are displayed in the inset. **B.** Main training steps: first the number of distractors, and then the size of the target was slowly increased to make the task as close as possible to the human task. This display overlaps the target screen and the choice screen to make the size difference visible, but the two never appeared at the same time on screen, as shown in Fig. 2A. **C.** Average performance at each step, with standard error across subjects, during the training procedure in experiment 1. **D.** Boxplot summarizing each primate's encoding time across the geometric shape testing blocks. **E.** Average performance at each step, with standard error across subjects, during the training procedure in experiment 2 which includes more intermediary steps than experiment 1.

the average success rate for the first presentation of each stimulus. Again, performance was better than chance (average success rate = 0.43, 95% CI $[0.35, 0.50]$, $t(11) = 8.02$, $p < .001$).

Furthermore, baboons did not take significantly longer to reach the success criterion for that step than for the previous training step: the mean number of trials for the last step before generalization was 1368, SEM=240.52, min=216, max=6480, while for the generalization step it was 1971 (SEM=574.59, min=216, max=6480; paired t-test on the number of trials, two tailed: difference is -603.00, 95% CI $[-1,923.40,\ 717.40]$, $t(11) = -1.01$, $p = .336$).

Geometric shapes

Baboons were significantly better than chance as early as the first geometric shape block (average success rate of 0.46, 95% CI $[0.42, 0.51]$, $t(11) = 15.50$, $p < .001$). Still, they performed significantly worse on the first block with geometric shapes than on the first block of generalization (paired t-test: difference of 0.14, 95% CI $[0.09,\ 0.20]$, $t(11) = 6.10$, $p < .001$). Note that in the case of the geometry trials, a block consisted of 68 trials with different target shapes, so there was no repetition within a block, which makes learning at the block level harder: if we compare the performance on a geometry block with the performance in the first presentation of each stimulus, they are not significantly different (difference -0.04, 95% CI $[-0.09,\ 0.02]$, $t(11) = -1.43$, $p = .179$). Although they all eventually reached the 80% target accuracy, it also took baboons significantly longer to reach that criterion in the geometry condition than it did for the generalizing step (number of trials: mean= 3337.67, SEM=530.47, min=476, max=6528; paired t-test on the number of trials, two tailed: -1,366.67, 95% CI $[-2,265.01,\ -468.32]$, $t(11) =\ -3.35$, $p = .006$).

We next turned to analyzing the baboons' performance across shapes during the geometric shape period. Our first observation was that encoding time was not a meaningful metric for this cohort. As shown on Fig. 2D, the baboons used very short "encoding time," averaging at 193ms (distribution of the subject's means: SEM= 18.58ms, min=126.00ms, max=319.60ms). This duration was not modulated by the target shape: we conducted a one-way between subjects ANOVA to compare the effect of shape on encoding time and found no significant effect [F(67, 737) = 1.07, p = .334]. These observations suggest that the animals touched the sample as fast as possible in order to make it disappear and move to the choice screen: indeed, since the starting cross and the target stimulus were in the same position, it was possible to "double-click" the screen fast to move on to the choice screen while briefly seeing the target stimulus. Contrary to humans, baboons did not modulate their actions in order to vary the duration of the target display according to the complexity of the item on display, making no use of the possibility to look at some shapes longer than others, or in fact to look for long at any shape at all. Consequently, for this experiment we did not analyze this dependent measure further–this is why we designed experiment 2 with a longer presentation period (now imposed, not self-paced).

We then turned to choice time and choice accuracy and correlated both metrics to the target shape MDL. To compare performance across animals with different learning speed, we only kept the blocks where performance was above 70% correct (but below 80%, since this threshold moved them to another task.) This means that different animals had different numbers of trial, and we first averaged the data for each shape, and then for each MDL value, and finally across subjects before correlating it with the MDL. For these blocks, the average choice time across shapes was 1112.40ms (SEM=20.50, min=707.70, max= 1452.70).

In a simple linear regression analysis, we found no effect of MDL on either accuracy or response time (see Fig.3B; choice time: F(1,10)=1.82, p=.21, R^2=.15; accuracy: F(1,10)=2.00, p=.19, R^2=.17). However, since we expected other predictors such as the luminance or the number of sub-parts of a shape to also predict behavior, we next turned to more sensitive analyses with mixed-effect general linear models. We used each shape's MDL and other visual features as additive fixed effects on either accuracy or choice time, and an intercept as random effect (adding additional random effects caused singular fits). All variables were z-scored across the entire dataset to make the fitted coefficients comparable. The model was estimated on a total of 816 observations across 12 participants and successfully converged using the Satterthwaite method; p-values were estimated with Type 3 tests, with F(1, 795.00) degrees of freedom for each predictor. Estimation was performed using R version 4.5.1 and afex version 1.4 (Singmann et al., 2015). The coefficients and their statistical significance

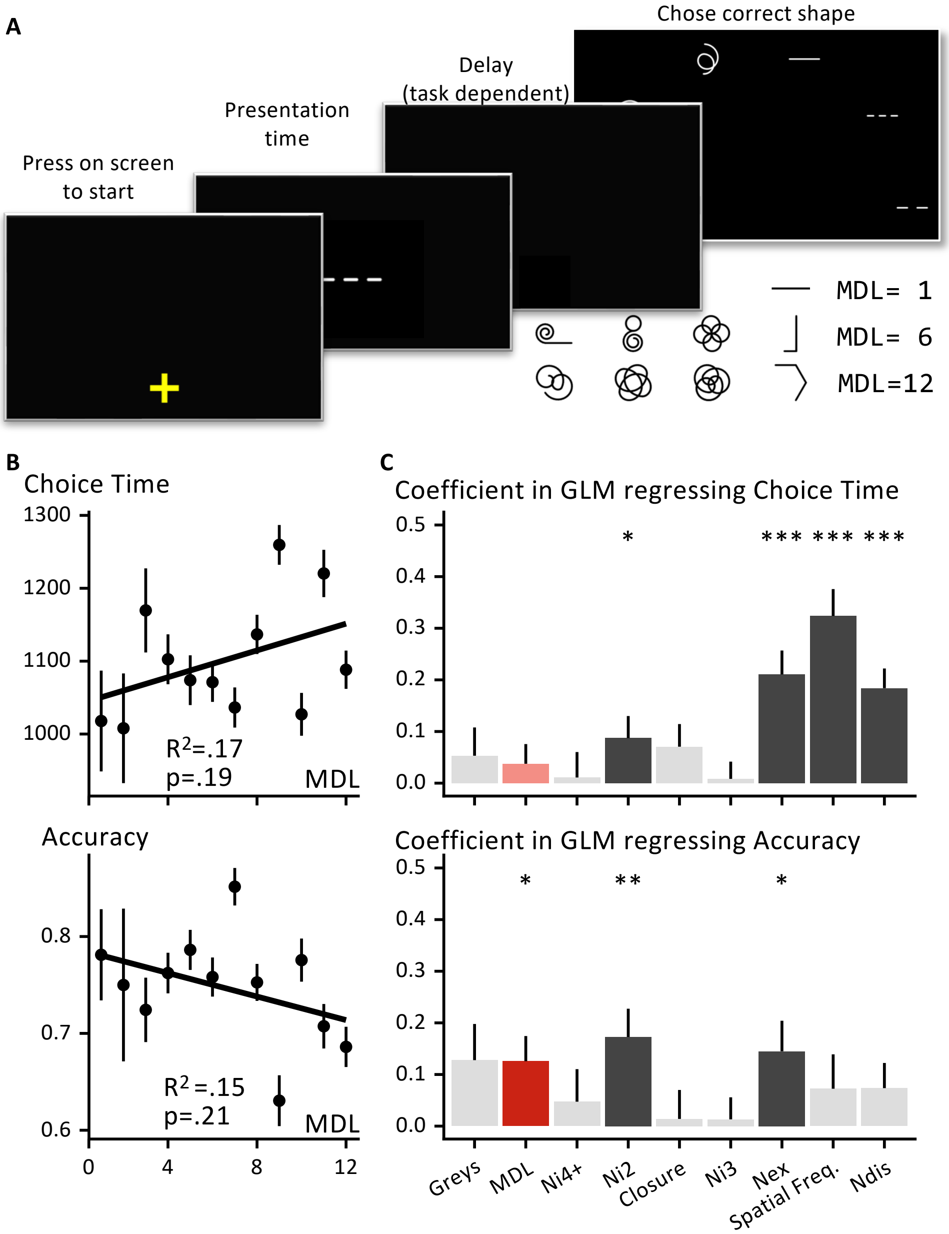

Fig. 3. **A.** Trial structure for the test condition, plus examples of stimuli of different MDL on the bottom right (full list of shapes and their MDL is provided in Fig. 1). **B.** Linear model between Choice Time (top) and MDL, or Accuracy (bottom) and MDL, together with the coefficient of determination R^2 and p-value of the slope against 0. **C.** GLM predicting respectively the Choice Time (top) or the Accuracy (bottom) from all the regressors we consider (same abbreviations as Fig. 1). Light colors are predictors that are not significant from 0 at the $p<.05$ level, and we highlight the MDL regressor in red. Start indicates p-value thresholds (*** for $p<.001$, ** for $p<.01$, and * for $p<.05$).

are shown in Fig. 3C. While the MDL predictor did not reach significance for the choice time in this analysis, it significantly correlated with accuracy, together with two other variables: number of 2-way intersections, and number of extremities.

Unfortunately, direct comparison with human data is made complicated by the fact that (i) baboons did not self-pace the encoding time, preventing us from modeling that dependent variable, and (ii) the two populations made different speed-accuracy trade-offs. Comparing to previously collected data in a human self-pace condition with the same task (Sablé-Meyer et al., 2022), the accuracy of the human group was virtually perfect: the average success was 99%, SEM=.12%; 34.9% of the participants made no mistake at all, yielding no variance to explain. On the other hand, their choice time was comparatively slower, with an average choice time of 1.52±.03s in humans compared to 1.11±.05s in baboons (two-sample t-test, p<.001).

Because of such differences, and to combine the information about both choice time and accuracy in the same meaningful manner in both populations, we unified these two dependent measures by using an integrated measure of speed and accuracy, the LISAS score. There was one LISAS score per participant and shape, which we submitted to the same analyses as above, see Fig. 4A, top row. Both a simple linear model with MDL and an intercept as predictors (F(1,10)=2.51, p=.14, R2=.20) and a mixed-effect GLM similar to the one described above but predicting the LISAS score (coefficient associated to the MDL term estimated as .051, 95% CI $[-0.03, 0.13]$, p=.21) indicated that MDL was not a significant predictor of baboon behavior at the p<.05 level (see below, experiment 3, for human data).

Experiment 2: Baboon, Slow-Paced

The results of experiment 1 showed that baboon behavior, unlike humans, was at best weakly influenced by geometric complexity as measured by MDL: our main dependent measure, the LISAS score, did not correlate with MDL. However, baboons did not make use of their ability to self-pace the encoding time to modulate it as a function of the shape complexity. Since they did not look at complex shapes longer than simple shapes, they might not have had time to encode them properly, which would bias the behavior in favor of low-level perceptual strategies that can be computed with a quick glance. To correct for this and provide a more meaningful comparison with humans, we decided to replicate the experiment with longer fixed, subject-independent looking times and wait times, set to the average of the human condition in the self-paced condition. The hope was that this manipulation would provide baboons with enough time to encode a richer representation of the target shape and use it in the choice period. We could then examine if, under those slower conditions, their behavior would be more correlated with the shapes' MDL.

Method

Participants

The cohort of participants was the same cohort of 25 baboons as the previous study. In this experiment, 9 baboons eventually reached the geometry task: we report data from these 9 participants in this section.

Procedure and Stimuli

Procedure The procedure was identical to experiment with one exception: during the main task, both the display of the target item, and the waiting time before the choice screen were set to 1000ms. We adapted the training procedure as described below to slowly introduce this change to the participants.

Stimuli The stimuli were fully identical to the one used in experiment 2.

Training Scheme

We introduced 12 additional training steps that took place before the first training step of experiment 1. Starting with a 1-in-3 choice screen, a 100ms presentation time and no delay, we slowly built up the display duration (200ms, 300ms, 500ms, 700ms, 1000ms) and then the delay duration (0ms, 100ms, 200ms, 300ms, 500ms, 700ms, 1000ms). Once both had reached 1000ms, we resumed the previous training procedure at the 5th step of experiment 1 where we increase the number of distractors, while keeping those delays fixed for the rest of the experiment.

Results

As shown in Fig. 2E, while the baboons had to relearn the first step of the training (200ms presentation duration, 0ms delay to match), they then very quickly moved through the successive steps where we increased target duration and delay duration (the median number of blocks required to move to the next step was 2 for all of these steps except the first one). The animals then learned the next steps very similarly to experiment 1.

On the crucial geometric test trials, the average performance across shapes was very comparable to experiment 1: after filtering for blocks with over 70% correct as before, the Pearson coefficient between the two experiments was r=.90 (in a linear model using the data from the fast condition to predict the data from the slow condition across all shapes, F(1,66)=294.3, p<.001, R^2=.82). Choice times were also very similar (distribution of average choice time across shapes: mean=1197ms, SEM=20.65, min= 755.60, max=1544.50; Pearson correlation across shapes between experiments: r=.81). Fig. 4B shows the Pearson correlation coefficients across shapes of the LISAS score for each pair of participants, including in different population and conditions, as well and the average correlations across population and conditions.

In this experiment, both choice time and choice accuracy showed a small but significant correlation with MDL in a simple linear regression (choice time: F(1,10)=6.54, p=.028, R^2=.40; accuracy: F(1,10)=7.13, p=.023, R^2=.42). The LISAS score summarizing those two effects was also correlated with MDL (LISAS score: F(1,10)=7.10, p=.024, R^2=.42; shown in Fig. 4A, 2nd row). Finally, MDL also predicted the LISAS score in a mixed-effect GLM with other predictors as described above, with 612 observations in 9 subjects (coefficient associated to the MDL term estimated as .25, 95% CI $[0.15, 0.35]$, p<.001 in a one sample t-test against 0). However, MDL was not the sole predictor of baboons' LISAS score, since the number of 2- and 4- way intersections, the shape's closure, and the number of disjointed parts were also significant predictors.

Experiment 3: Humans, Fast-Paced

Increasing the target encoding time changed baboons' behavior in measurable ways. In particular, the LISAS score of baboons in the slow-paced condition was now correlated with the shapes' MDL, a result previously found only in humans (Sablé-Meyer et al., 2022). This finding led us to wonder whether, conversely, with very short encoding times, humans would lose their correlation with MDL and revert to a baboon-like behavior. To answer this question, we collected data in human participants while the display and delay durations were fixed to short duration instead of self-paced (experiment 1). Together with our prior human data in the self-paced condition (Sablé-Meyer et al., 2022), this experiment provided us with a full 2x2 design matrix crossing pace (fast or slow) with species (humans versus baboons) – with the only difference that the slow-paced condition in humans was self-paced, with a delay that varied across shapes, while it was the fast condition that was self-paced in baboons.

Method

Participants

We recruited 560 participants online for this experiment. Demographics were self-reported and a few aberrant values were given but we did not remove participants on the basis of that information. The

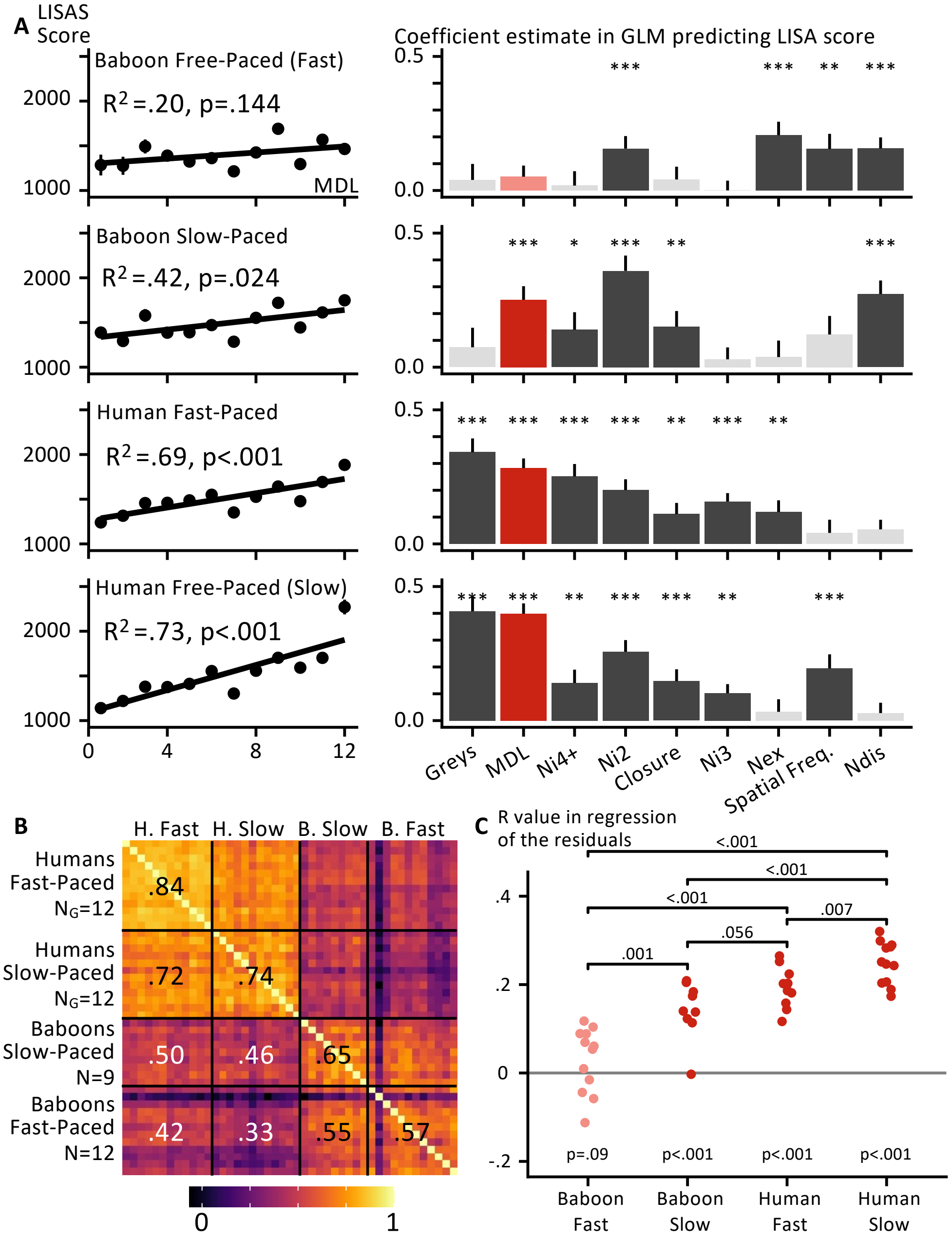

Fig. 4. A. Simple regression and GLM for our four conditions: two populations (humans and baboons) and two speed conditions (slow-paced and fast-paced). In all cases, we regress the LISAS scores; the simple regression is against the MDL, and the GLM against all of the regressors identified previously. **B.** Correlation coefficients across 68 shapes between the behavior across shapes for each individual baboon, in both conditions, and against each human group in both conditions. Values inside indicate the average correlation coefficient inside a particular cell (for diagonal cells, we restrict the average to the lower-triangular matrix without the diagonal), which corresponds to the average correlation between the corresponding populations and conditions. **C.** Coefficients in a simple regression of the MDL (x-axis) against the residuals in a regression of the LISAS score and all of the other regressors. Both for the GLM stage, and for the simple regression stage, both the predictor(s) and the value to predict have been z-scored, so that the resulting coefficients can be directly compared. Stars and p-values indicate pairwise two-tailed t-test, between each population and condition.

average age was 37.7 years old (SD=19.1; 1st quartile 28, 3rd quartile 45). The sample comprised 130 females, 408 males, and 22 who did not identify as either.

Following exactly our criteria for filtering participants from previous work (Sablé-Meyer et al., 2022), we removed participants that were on average three standard deviation above the average for the choice time across participants. Similarly, for each shape, we removed the responses that were three standard deviation above the average for that shape. We did not filter on accuracy. We computed these statistics separately and on the entire dataset: some datapoints were marked as outliers for both criteria. Overall, four participants were removed from the analysis, and 1.85% of the data points were removed.

Procedure, stimuli and training scheme To test the behavior of humans under time pressure when performing this task, this new group of participants was tested with a fixed display duration of 100ms. We also removed the waiting duration entirely to mimic the overall structure of our experiment 1. Additionally, we reused the semicircle display used in the baboons instead of the 3x2 grid used in the original study. Participants received explicit instructions to find the shape matching the target (see supplementary text), and took the same training steps as the baboons, in the same order but substantially accelerated to avoid wasting participants' time on a task that they easily understood. Thus, any success at a given training step moved them to the next step; otherwise that step (with new stimuli) was repeated. The trials started with a yellow empty circle encompassing where the target would appear, instead of a yellow cross. The stimuli were identical to those used in experiment 2.

To compare individual baboons, who went through many repetitions per individual and condition, with humans who only received a single trial per condition, we created pseudo-participants by pooling participants together and match the number of non-human primate subjects. We aggregated the data of our human participants into twelve groups based on a random ID (random strings of 10 alphanumeric characters) assigned to each participant when they took the experiment. We sorted the participants' ID by lexicographic order and assigned them to groups 1 to 12 based on the ID rank in an interleaved fashion (1st ID goes to group 1, 2nd to 2, etc., 12th to group 12, 13th to group 1, 14th to group 2, etc. until the last participant). The procedure avoids any order effect, and because the IDs are generated at experimentation time, it is deterministic when running the analysis. This ensured a systematic but randomized way to pool participants and created as many pseudo-participants as we have subjects in experiment 1. Note that using all participants separately does not change substantially the results: see Fig. S1 for a comparison of the estimates before and after this procedure. This also allowed us to estimate a fine-grained LISAS score in humans – otherwise, since each participant performed a single trial per shape, the second term of the LISAS estimate would either be multiplied by 0 or 1 as we could not reliably estimate the error rate to a single geometric stimulus. Then, when we performed direct comparisons between each individual baboon and other baboons or humans, instead of reporting the results of individual human participants, we compared baboons to each *group*, treating it (and its average, etc.) as an individual participant. In these analyses, we lose the ability to model inter-individual differences between humans, but this makes direct comparison between the different conditions more straightforward. We applied the same procedure to the data from our previously published human slow-paced condition (Sablé-Meyer et al., 2022).

Results

Following the baboon training procedure, humans went through 9 successive training steps. Having read explicit instructions, human participants performed extremely well even during this training period. The average number of trials required per training step was 1.4 (SEM=.016 across participants), significantly better than chance (esp. for step 4 onward where chance is one item in six). This suggests that despite the accelerated training procedure, participants did not have trouble performing the fast-paced match-to-sample task. Indeed, overall accuracy for the test period was 96.3% (SEM=0.27%; 22.8% made no mistake at all, and 481 out of 556 participants made fewer than 5 mistakes during the testing period; nevertheless, the performance is slightly below those of the human self-pace condition: since normality cannot be assumed with so many participant with perfect performance, we performed a Wilcoxon rank sum test: W=42986, p<.001)

Next, we investigated the behavior with geometric shapes. Again, we aggregated error rates and choice times with the LISAS score, and correlated it with MDL. MDL significantly predicted LISAS scores in a simple linear regression (F(1,10)=21.76, p<.001, R^2=.69; Fig. 4A). We also found that the MDL also predicted the LISAS score in a mixed-effect GLM with other predictors like the one described above, with 816 observations in 12 pseudo-subjects (coefficient associated to the MDL term estimated as .28, 95% CI $[0.21, 0.35]$, p<.001). In this regression, MDL was the second-best predictor, after gray levels (i.e. luminance; see figure 4A, 3rd panel).

Overall Results

To measure the impact of presentation pace on behavior, we estimated the similarity between the betas in our mixed effect GLM across our four conditions. For each condition, we estimated the mixed-effect GLM model described above and extracted the coefficient associated to each of our 9 predictors. Then, we modeled the distribution of coefficients in the conditions using a simple linear regression. In humans, the correlation between the fast- and slow-paced conditions was strong and highly significant (F(1,7)=39.48, p<.001, R2=.84). The same was not true of baboons, however (F(1,7)=2.97, p=.13, R2=.30). This suggests that the difference in looking time does not substantially change the strategy that humans use, but makes a different in the strategies that baboon use.

Figure 4B shows the correlation matrix between the LISAS of each individual baboon and human group, across the 68 shapes, separately for the two conditions (fast and slow pace). The behavior of any two baboons in those two conditions was more correlated with one another than with humans in either condition. Likewise, the behavior of any two humans in those two conditions was more correlated with one another than with the baboons in either condition. Thus, in our geometry task, baboons differed from humans, regardless of task conditions (Figure 4B). Nevertheless, forcing humans to encode the shape in 100 ms made their behavior slightly more correlated with the baboons' in either condition than when humans took the self-paced condition.

To measure the extent to which humans are more correlated with baboons in the fast condition than in the slow condition, we performed a non-parametric test: since different cells in the correlation matrix are not independent, we performed a bootstrap analysis by shuffling the scores of each shape for each (pseudo)-subject before estimating the empirical null distribution of the average difference in correlation between humans in one condition and baboons in both conditions. We sampled 1,000 such shuffled correlation matrices and found that the true difference was greater than every single shuffle, putting an upper bound on the p-value and indicating that the average correlation between data from the fast-paced human group and data from baboons across conditions is significantly higher than the correlation with data from the self-paced human group (p < .001).

To test more specific predictions, we report in Fig. 4C a regression of MDL with the residuals of a GLM including all the other predictors. To estimate this, we first ran, separately for each (pseudo-)participant, a GLM with all the z-scored predictors from our mixed-effect GLMs except for MDL. We then extracted the residuals from that fitted model and, separately for each (pseudo-)participant, performed a simple linear regression with the z-scored MDL as only predictor. We extracted the estimated coefficient for each (pseudo)-participant, which we then analyzed.

A joint analysis of all the available data also revealed interesting patterns across and within populations. To make data across populations comparable, we systematically relied on the LISAS scale. Figure 4A shows the results of simple linear regressions and the full GLM. In both populations, the fast-paced condition elicited a significantly weaker MDL effect on behavior, suggesting that the mental operations required for estimating a shape's mental program require some time—an unsurprising fact in humans given our previous findings on encoding time and the MDL of geometric shapes.

First, by comparing those values against the null hypothesis, we again found a significant effect of the shape's MDL in both human conditions, but also, crucially, in the slow-paced baboon data (see Table 2). We also performed pair-wise comparisons between populations (two-sample t-test, statistics reported on Fig. 4C). We found a significant difference between the fast-paced baboon

Population	estimate	CI (lower bound)	statistic	df	p.value
Baboons, Exp. 1	0.03	-0.01	1.43	11	.090
Baboons, Exp. 2	0.14	0.10	6.62	8	$< .001$
Humans, Exp. 3	0.19	0.17	15.96	11	$< .001$
Humans, previous work	0.25	0.22	18.14	11	$< .001$

Table 2: Average, confidence intervals and statistics associated to the distribution of coefficient estimated in (pseudo)-subject level analysis of behavior in a regression of the shapes' MDL against the residuals of the LISAS score on our set of confounding predictors. The datapoints are visible in Fig. 4C.

and the fast-paced humans condition and a significant difference between the slow-paced baboon condition and the self-paced human condition: in both cases, the magnitude of the MDL predictor was significantly higher in humans. However, while there was a trend for the slow-paced baboon condition to exhibit smaller MDL correlation weights than in the fast-paced human condition, it was not significant at the p<.05 level, and appeared driven by a single individual.

To separate the effect of population and encoding duration on the effect of MDL, we conducted a two-way ANOVA with species (humans or baboons) and pace (slow or fast) as between-subject predictors of the coefficients of the MDL predictor in the regression against the residuals from other regressors. We found a large main effect of species ($F_{1,41} = 60.8$, $\eta^2_G = .60$, p<.001) and a significant main effect of encoding duration ($F_{1,41} = 23.1$, $\eta^2_G = .36$, p<.001), but no significant interaction ($F_{1,41} = 2.8$, $\eta^2_G = .063$, p=.10). This result did not depend on our choice of accuracy cutoff in baboons (nor did most results; see Fig. S1). Keeping all the baboon data yielded very similar results, as we again found a very significant main effect of species ($F_{1,41} = 76,4$, $\eta^2_G = .65$, p<.001) and a significant main effect of encoding duration ($F_{1,41} = 35.25$, $\eta^2_G = .46$, p<.001), but no significant interaction ($F_{1,41} = 3.2$, $\eta^2_G = .071$, p=.08). Very similar results came out of a mixed effect model of the LISAS score across all (pseudo-)participants with both population and condition as categorical predictors, together with their interaction with one another and with the continuous predictors used throughout. This time, there was a significant main effect of population, but not of condition or the interaction of the two. There was also a main effect of all of our predictors except spatial frequency. The effect of MDL on baboons was lower than its effect in humans, and the effect of MDL was higher in the slow conditions than in the fast conditions, but there was no interaction of these effects—see Table S1 for details of the model and its fit.

Overall, those results indicate that baboons were less correlated with MDL than humans, and that in both cases, a slower pace of shape presentation boosted the correlation with MDL, yet without any interaction between these two factors. Note that the explained variance (the generalized eta squared, η^2_G, an estimation of the proportion of variance explained by a factor in ANOVAs (Olejnik & Algina, 2003)) was much higher for the species factor than for the encoding duration factor.

Discussion

Across experiments, we have found that a model of human geometric-shape representation also accounts for non-null variance in baboons' behavior on a comparable task with geometric shapes. Indeed, when baboons perform a delayed match to sample task, with presentation duration set to one second and geometric shapes sampled from a generative language for geometric shapes, their pattern of performance on a given shape was slightly but significantly predicted by the length of the shortest program that generated that shape (its MDL). We have also found that this effect of the length of the shortest program was strongly modulated by the viewing time of the target shape before matching: when participants looked longer at the target shape, their behavior correlated more strongly with the MDL predictor.

We also found several interesting differences between species. First, the effect of the MDL was significantly weaker in baboons overall, to the point of being non-significant in the self-paced condition. Second, the two populations self-paced in opposite directions: whereas baboons self-paced equally fast with all shapes (and overall responded very fast), humans look longer at shapes with higher MDL. This is why, to make the two populations more comparable, we slowed down the presentation pace in baboons while we sped it up in humans.

Task Demands, Inhibition and Meta-Cognition

When given the opportunity to study the shapes for as long as they wished, human participants chose to self-pace the task and slow down to observe geometrically complex shapes for a longer time. This is a rational strategy to facilitate memory, and indeed, we found that even though the fast-paced human group of experiment 3 already achieved extremely high success rates, participants in the self-paced condition performed even better. It is of course unsurprising that longer looking time yields better performance, but it is interesting and relevant that in the self-paced condition, longer looking times were specifically allocated to the more complex shapes. This indicates that human adults possess an introspection of shape complexity, as previously report for quadrilaterals (Sablé-Meyer et al., 2021), and a metacognitive ability to adapt behavior to task difficulty in real time.

Baboons, by contrast, failed to use the self-paced condition in order to extend the time available to study the most complex shapes. This finding suggests that the two populations made different trade-offs when approaching this task. This species difference could arise from several, non-necessarily incompatible sources. First, baboons may be deficient in meta-cognitive abilities or inhibitory control and respond primarily on an impulsive basis. However, there is evidence that baboons score well on metacognitive monitoring and task control (Loyant et al., 2025; Malassis et al., 2015). A second possibility is that baboons' cognitive control systems do not have introspective access to the complexity of the shapes. Thus, shape complexity impacts their behavior, but not in a way that they can anticipate or use for inhibitory control—unlike humans, whose explicit reports of shape complexity correlates strongly with their performance (Sablé-Meyer et al., 2022).

A third, related hypothesis comes from the neuroimaging of geometric shape perception in humans. By recording MEG signals while human adults viewed quadrilaterals, we found two consecutive stages in the visual processing of shapes (Sablé-Meyer et al., 2025): an early stage, peaking around 90ms, is well captured by convolutional neural models of image processing, while a later stage, peaking around 230ms, relates to a geometric representation of shapes. Given the impulsiveness and fast response time of the baboon population, it is possible that in the fast condition, the animals took their decisions before the geometric representation could emerge and influence behavior. This account is also compatible with the observed difference between the fast and the slow condition in baboons.

Impact of the Choice of language

While we found only a weak effect of MDL in baboons, it is important to keep in mind that our MDL measure is based on the choice of a particular language-of-thought for geometric shapes. This is a strength of our original model: once the language is fixed, the model becomes parameter-free, which makes it easier to validate or invalidate in humans. On the other hand, the choice of the primitives was partially arbitrary, based on its capacity to capture the most basic shapes that humans preferentially produce across cultures. It could be that our set of primitives is not well adapted to non-human primates, even if they also possess LoT-like representations.

Indeed, work in other domains investigating LoT models of human cognition found that, even within humans, the set of primitives that captured the data varied slightly with age and education (Amalric et al., 2017). Even in a fixed adult population, previous work has established the need for incorporating some flexibility in LoT proposals to either select among multiple proposals (Piantadosi et al., 2016)–and in development data across learning, to incorporate the learning of new primitives with experience (Piantadosi et al., 2012; Rule et al., 2020). With no obvious primate productions to bootstrap the search for a language, starting from a human-based proposition is a valid starting point, but further research could look for a better LoT for non-human primates. This would require exploring not only the space of programs within a given language, but also the space of languages, and repeating the present experiment with many more shapes sampled from several competing languages.

Local versus Global effects in perception

In complex displays, such as a square made up of small circles, humans tend to notice global changes (e.g. when the large square is changed into a triangle) faster than local changes (e.g. when the circles are changed into crosses). Interestingly, the reverse is true in baboons (Deruelle & Fagot, 1998; Fagot & Deruelle, 1997). Although the roots of this finding are still being investigated (Santacà, 2024), it suggests that complex displays and shapes are encoded differently in both species. This is compatible with the hypothesis that recursive nesting might be particularly costly, if at all available, in baboons (Dehaene et al., 2015, 2022; Hauser et al., 2002).

Specific primitives in our language affect the global properties of a shape more often than the local properties of a shape. For example, the repeat and embed primitives can be combined to generate the outline of a big shape from many small shapes (e.g. “a line of squares”). Conversely, the draw primitive is the one that generates lines and curves, and if embedded in repeat or embed primitives it will impact local properties more strongly. Since baboons are more sensitive to local features than humans, an interesting hypothesis would be that they have access to a similar representation format but with a different emphasis on which primitives are easily available, and which are not—and therefore we should adjust not the language but the relative weight of the primitives as a function of population and condition. Most crucially, the small correlation with our current MDL that we observed here therefore does not imply that non-human primates possess the same geometric compositionality as humans do.

Comparison with other shape skeleton theories

How does our proposal of a model of geometric shape perception compare with others? Here, we briefly describe how our theory the overlap and complementarity with another theory of shape perception, the shape skeleton.

The idea that we identify shapes, based on their internal structure, by finding their skeleton, principal axis or medial axes, is influential in the neuropsychology of vision (Ayzenberg & Lourenco, 2019; Blum, 1973; Lowet et al., 2018). While the skeletal model has mostly been applied to the contour of real-life animal or objects, people’s perception of even simple geometric shapes seems to be attuned to their skeleton, in the absence of a task (Firestone & Scholl, 2014). Models to estimate the skeleton of shapes share some properties with our approach: they are using minimal description length as a model selection criterion to trade the prior of a given skeleton with its likelihood to be a good fit for a shape (Feldman & Singh, 2006), and they can construct complex objects from simpler parts (Froyen et al., 2015).

Still, our geometric shapes are very unlike the kind of shapes for which skeletal perception has been postulated—often object or animal contours. Our proposal is very limited in scope to geometric shapes, but the shapes that we model are not straightforwardly described by approaches based on skeletons. In fact, the two approaches can be complementary, as objects that are well modelled by skeletal theories are very poorly explained by our theory (e.g. the contour of a person), while the shapes that we are generating are poorly captured by skeletons. Most interestingly, once the skeleton of a shape is extracted (e.g. a snake), the LoT theory that we developed could explain how humans compress it further, thus ending up depicting it with simple geometric shapes such as a zigzag or a spiral – drawings abilities which are not observed in other non-human primates (Close & Call, 2015; Saito et al., 2014; Sueur, 2025; Tanaka, 2007).

Comparison with existing results

The current finding contrasts partially with the results that were previously obtained when comparing the perception of quadrilaterals in humans and baboons (Sablé-Meyer et al., 2021). What is common to both studies is that baboons exhibit much greater difficulties than humans in processing geometric shapes–and the present data indicate that this large difference between species is robust to changes in task conditions, even when humans are asked to process the stimuli faster than baboons. However, while baboon quadrilateral perception was entirely explained by a convolutional network model of visual processing, without any need for a symbolic or geometry-based language-of-thought, the present study identified a condition in which, when baboons are forced to consider the shapes for longer time before responding, their behavior is slightly but significantly predicted by MDL.

We see two possible interpretations of this finding. First, it is possible that the hypothesis of a human singularity for geometry is too strong, and that baboons also possess a language-of-thought for geometry, at least in rudimentary form. Alternatively, it is possible that some perceptual variable, correlated with MDL, affected baboon behavior and made it superficially like human behavior. While our previous quadrilateral study focused on specific properties (parallelism, right angles, equal sides, equal angles), which our study showed are not attended by baboons, the present work used a much greater diversity of shapes whose properties are therefore harder to control. We controlled for many perceptual properties of our geometric shapes (see Figure 1B), but we might have missed some. This possibility is made more likely by the fact that, as shown in figure 4A, baboon and human behaviors were captured by very different predictors, even when the MDL regressor appeared in both.

Conclusion

In this work, we present some evidence for the possibility that a rudimentary Language-of-Thought representation of geometric shapes may be present in baboons. The effect is weaker than in humans on comparable tasks, and in both species, it is modulated by encoding duration. Notably, baboons do not self-pace in a way that makes their behavior better predicted by LoT models; instead, they tend to respond as quickly as possible unless explicitly delayed. These findings suggest that the cost of accessing or deploying a LoT strategy may be higher for baboons than for humans, that baboons have reduced inhibitory control, or that their LoT differs from the one we proposed for humans.

Taken together, the results indicate that some of the mental representations we had previously postulated in humans may be partially present in non-human primates as well. This, in turn, leaves open the question of why humans naturally produce geometric pictorial productions whereas baboons do not. Beyond providing a rich dataset for benchmarking theories of symbolic perception across species, our work replicated the finding that human behaviour, much more than baboon behaviour, is governed by Minimum Description Length (MDL) even under severe time constraints. By identifying the rudiments of this effect in baboons, the present research may bring us a step closer to understanding the evolutionary trajectory of symbolic behaviour in humans.

Acknowledgements

This work was supported by INSERM, CEA, CNRS, Collège de France, an ERC advanced grant “MathBrain” to S.D, and by a FYSSEN post-doctoral study grant to M.S.M. The primate research facility is also supported by the Brain and Language Research Institute Labex (ANR-11-LABX-0036), Convergence Institute Language Communication and the Brain (ANR-16-CONV-0002).

M.S.M. would like to express his special gratitude to Kristian Tylén for having been such a positive force in pulling together researchers from psychology, archeology, and anthropology for the benefit of science. Science can thank him for his ability to do the hard but necessary work of lifting ideas, questions and method from one field and using them to inform thinking in another field, thereby uniting academic circles. Every opportunity that M.S.M. has had to spend with him in person led to uplifting, energetic discussions on the origin of symbols in human cognition, and how to make sense of contemporary data to understand the evolutionary trajectory of *thought*, a subject he cared deeply about.

Data availability

Data and analysis code are available on the following OSF repository: https://osf.io/3qwxa/?view_only=091c8a0ec371412786f9421612bd4087

References

Abelson, H., Goodman, N., & Rudolph, L. (1974). Logo manual.

Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., & Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology*, *13*(1), e1005273.

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports*, *9*(1), 9359. https://doi.org/10.1038/s41598-019-45268-y

Blum, H. (1973). Biological shape and visual science (part I). *Journal of Theoretical Biology*, *38*(2), 205–287. https://doi.org/10.1016/0022-5193(73)90175-6

Chen, L. (1982). Topological structure in visual perception. *Science*, *218*(4573), 699–700.

Close, J., & Call, J. (2015). From colour photographs to black-and-white line drawings: An assessment of chimpanzees'(Pan troglodytes') transfer behaviour. *Animal Cognition*, *18*, 437–449.

d'Errico, F., Doyon, L., Colagé, I., Queffelec, A., Le Vraux, E., Giacobini, G., Vandermeersch, B., & Maureille, B. (2018). From number sense to number symbols. An archaeological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740), 20160518. https://doi.org/10.1098/rstb.2016.0518

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences*, *26*(9), 751–766. https://doi.org/10.1016/j.tics.2022.06.010

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, *88*(1), 2–19. https://doi.org/10.1016/j.neuron.2015.09.019

Deruelle, C., & Fagot, J. (1998). Visual search for global/local stimulus features in humans and baboons. *Psychonomic Bulletin & Review*, *5*(3), 476–481.

Ellis, K., Wong, L., Nye, M., Sable-Meyer, M., Cary, L., Anaya Pozo, L., Hewitt, L., Solar-Lezama, A., & Tenenbaum, J. B. (2023). Dreamcoder: Growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning. *Philosophical Transactions of the Royal Society A*, *381*(2251), 20220050.

Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behavior Research Methods*, *42*(2), 507–516. https://doi.org/10.3758/BRM.42.2.507

Fagot, J., & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialization in humans (Homo sapiens) and baboons (Papio papio). *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 429.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019.

Firestone, C., & Scholl, B. J. (2014). "please tap the shape, anywhere you like" shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological science*, *25*(2), 377–386.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychological Review*, *122*(4), 575–597. https://doi.org/10.1037/a0039540

Gulwani, S., Polozov, O., Singh, R., et al. (2017). Program synthesis. *Foundations and Trends® in Programming Languages*, *4*(1-2), 1–119.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–79. https://doi.org/10.1126/science.298.5598.1569

Henshilwood, C. S., d'Errico, F., van Niekerk, K. L., Dayet, L., Queffelec, A., & Pollarolo, L. (2018). An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature*, *562*(7725), 115–118. https://doi.org/10.1038/s41586-018-0514-3

Joordens, J. C., d'Errico, F., Wesselingh, F. P., Munro, S., de Vos, J., Wallinga, J., Ankjærgaard, C., Reimann, T., Wijbrans, J. R., Kuiper, K. F., et al. (2015). Homo erectus at Trinil on Java used shells for tool production and engraving. *Nature*, *518*(7538), 228.

Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018, September 4). *Cornet: Modeling the Neural Mechanisms of Core Object Recognition*. Neuroscience. https://doi.org/10.1101/408385

Leeuwenberg, E. L. J. (1971). A Perceptual Coding Language for Visual and Auditory Patterns. *The American Journal of Psychology*, *84*(3), 307. https://doi.org/10.2307/1420464

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs(?) *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/s13428-018-1076-x

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, *80*, 1278–1289.

Loyant, L., Collins, L., & Joly, M. (2025). Inhibitory control tests in non-human animals: Validity, reliability, and perspectives. *Biological Reviews*.

Malassis, R., Gheusi, G., & Fagot, J. (2015). Assessment of metacognitive monitoring and control in baboons (Papio papio). *Animal Cognition*, *18*(6), 1347–1362.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological methods*, *8*(4), 434.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*(2), 199–217. https://doi.org/10.1016/j.cognition.2011.11.005

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological review*, *123*(4), 392.

Roelfsema, P. R., & Singer, W. (1998). Detecting connectedness. *Cerebral cortex (New York, NY: 1991)*, *8*(5), 385–396.

Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The Child as Hacker. *Trends in Cognitive Sciences*, *24*(11), 900–915. https://doi.org/10.1016/j.tics.2020.07.005

Sablé-Meyer, M., Benjamin, L., Watkins, C. P., He, C., Pajot, M., Morfoisse, T., Al Roumi, F., & Dehaene, S. (2025). A geometric shape regularity effect in the human brain. *eLife*. https://doi.org/10.7554/elife.106464.1

Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2022). A language of thought for the mental representation of geometric shapes. *Cognitive Psychology*, *139*, 101527.

Sablé-Meyer, M., Fagot, J., Caparos, S., Kerkoerle, T. v., 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). https://doi.org/10.1073/pnas.2023123118

Saito, A., Hayashi, M., Takeshita, H., & Matsuzawa, T. (2014). The Origin of Representational Drawing: A Comparison of Human Children and Chimpanzees. *Child Development*, n/a–n/a. https://doi.org/10.1111/cdev.12319

Santacà, M. (2024). Some like it "local": A review of hierarchical processing in non-human animals. *Learning & Behavior*, *52*(2), 143–161.

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M., Højsgaard, S., Fox, J., Lawrence, M., Mertens, U., Love, J., et al. (2015). Package 'afex'. *URL http://afex. Singmann. Science/, https://github. Com/singmann/afex*.

Sueur, C. (2025). From stones to sketches: Investigating tracing behaviours in Japanese macaques. *Primates*, 1–6.

Tanaka, M. (2007). Recognition of pictorial representations by chimpanzees (Pan troglodytes). *Animal cognition*, *10*, 169–179.

Texier, P.-J., Porraz, G., Parkington, J., Rigaud, J.-P., Poggenpoel, C., Miller, C., Tribolo, C., Cartwright, C., Coudenneau, A., Klein, R., et al. (2010). A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. *Proceedings of the National Academy of Sciences*, *107*(14), 6180–6185.

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior research methods*, *49*(2), 653–673.

Vandierendonck, A. (2021). On the utility of integrated speed-accuracy measures when speed-accuracy trade-off is present. *Journal of cognition*, *4*(1), 22.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., Van der Helm, P. A., & Van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological bulletin*, *138*(6), 1218.

Supplementary Materials

Task instruction

Supplementary text: task instruction for humans, experiment 3

English:

In this experiment, you have to memorize a shape and then identify it within others. You have to:

1. click inside the yellow circle to start a trial. As soon as you click, a shape briefly flashes on the screen, and **you have to memorize it. Pay close attention!**
2. Then several shapes appear, and you have to **click on the shape you just memorized**. Warning: you cannot change your answer once you've clicked!

Some trials are easy, others are hard, this is intended. Thanks for you time, and stay focused to help us!

French:

Dans cette expérience, vous devez mémoriser une forme puis l'identifier parmi d'autres. Pour cela :

1. cliquez dans le cercle jaune pour commencer un essai. Une forme s'affiche alors très brièvement : **vous devez y faire attention, concentrez vous !**
2. parmi les formes qui s'affichent ensuite, **cliquez sur la forme que vous venez de voir**. Attention : vous ne pouvez pas changer votre réponse une fois que vous avez cliqué !

Certains essais sont faciles, certains plus difficiles, c'est normal. Merci pour votre participation, et restez concentrés pour nous aider !

Supplementary analysis: Mixed Effect GLM

To make best use of the entire dataset, we pulled it in a single mixed-effect regression where our continuous regressors can interact with the population and duration condition. The model structure is as follows:

$$\begin{aligned} LISAS \sim & (pop \times condition) \\ & \times (Greys + MDL + Ni4 + +Ni2 + Closure + Ni3 + Nex + SpatialFreq + Ndis) \\ & + (1 \mid ID) \end{aligned}$$

There was a total of 3060 observations across 36 (pseudo)-individuals: note that while there was no overlap in participants across the human experiments, the identity of the subjects in experiment 1 and 2 was shared (experiment 2 saw a subset of subjects from experiment one, yielding 12 distinct participants for the baboon experiment, and 12 per human experiment for a total of 36 different values for the ID variable). The standard deviation for the remaining parameters was as follows: for ID, sd(Intercept)=112.; overall, sd(Observation)=340.

Below we report the fit of the model as well as the associated statistics.

Table 3: Fitted value and associated statistics for a mixed effect model of the LISAS score including population and condition as co-variates. Linear mixed model fit by REML using the same R and afex version as for the other models. t-tests use Satterthwaite's method.

term	estimate	std.error	statistic	df	p.value	p.value.stars
(Intercept)	1499.35	20.75	72.26	32.09	p<.001	***
pop=baboons	-68.32	20.75	-3.29	32.09	p=.002	**
cond=fast	-17.00	13.07	-1.30	45.29	p=.2	
Greys	86.44	12.70	6.81	2986.96	p<.001	***
MDL	100.45	8.84	11.37	2986.96	p<.001	***
Ni4+	50.67	11.43	4.43	2986.96	p<.001	***
Ni2	-99.58	9.93	-10.03	2986.96	p<.001	***
Closure	-36.66	10.16	-3.61	2986.96	p<.001	***
Ni3	26.44	7.74	3.42	2986.96	p<.001	***
Nex	39.01	10.73	3.64	2986.96	p<.001	***
spatial_freq	-16.80	12.07	-1.39	2986.96	p=.164	
Ndis	-44.84	8.84	-5.07	2986.96	p<.001	***
pop=baboons:cond=fast	0.30	13.07	0.02	45.29	p=.982	
pop=baboons:Greys	-62.70	12.70	-4.94	2986.96	p<.001	***
pop=baboons:MDL	-38.45	8.84	-4.35	2986.96	p<.001	***
pop=baboons:Ni4+	-18.56	11.43	-1.62	2986.96	p=.104	
pop=baboons:Ni2	-7.59	9.93	-0.76	2986.96	p=.445	
pop=baboons:Closure	15.53	10.16	1.53	2986.96	p=.127	
pop=baboons:Ni3	-20.62	7.74	-2.66	2986.96	p=.008	**
pop=baboons:Nex	14.76	10.73	1.38	2986.96	p=.169	
pop=baboons:spatial_freq	26.96	12.07	2.23	2986.96	p=.026	*
pop=baboons:Ndis	-44.98	8.84	-5.09	2986.96	p<.001	***
cond=fast:Greys	-31.80	12.70	-2.50	2986.96	p=.012	*
cond=fast:MDL	-51.04	8.84	-5.77	2986.96	p<.001	***
cond=fast:Ni4+	-12.87	11.43	-1.13	2986.96	p=.26	
cond=fast:Ni2	37.68	9.93	3.79	2986.96	p<.001	***
cond=fast:Closure	30.93	10.16	3.04	2986.96	p=.002	**
cond=fast:Ni3	-5.35	7.74	-0.69	2986.96	p=.49	
cond=fast:Nex	23.11	10.73	2.15	2986.96	p=.031	*
cond=fast:spatial_freq	56.94	12.07	4.72	2986.96	p<.001	***
cond=fast:Ndis	17.03	8.84	1.93	2986.96	p=.054	
pop=baboons:cond=fast:Greys	25.72	12.70	2.03	2986.96	p=.043	*
pop=baboons:cond=fast:MDL	12.06	8.84	1.36	2986.96	p=.173	
pop=baboons:cond=fast:Ni4+	-11.06	11.43	-0.97	2986.96	p=.333	
pop=baboons:cond=fast:Ni2	-0.52	9.93	-0.05	2986.96	p=.958	
pop=baboons:cond=fast:Closure	8.63	10.16	0.85	2986.96	p=.396	
pop=baboons:cond=fast:Ni3	-0.62	7.74	-0.08	2986.96	p=.936	
pop=baboons:cond=fast:Nex	15.48	10.73	1.44	2986.96	p=.149	
pop=baboons:cond=fast:spatial_freq	2.17	12.07	0.18	2986.96	p=.857	
pop=baboons:cond=fast:Ndis	2.68	8.84	0.30	2986.96	p=.762	

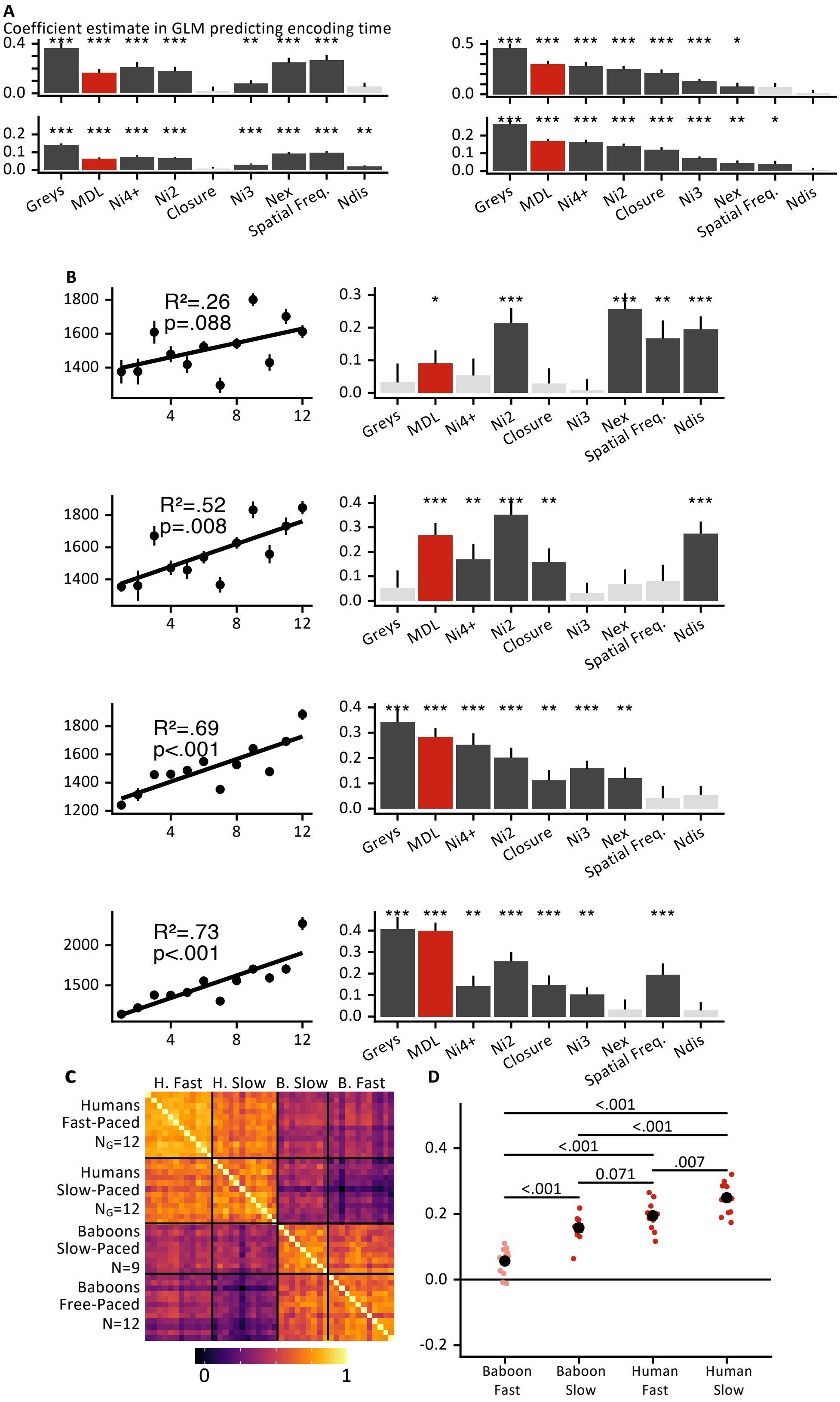

Fig. S1 A. Comparison between predicting the choice time of participants with our predictors after creating pseudo-population of 12 pseudo-participants (top row) or each participant separately (top row), in experiment 3 (left) and data from Sablé-Meyer et al. (Sablé-Meyer et al., 2022). The estimates are overall smaller on the bottom row, but so is the standard error of the mean around them: this is expected from the fact that while the pseudo-population procedure effectively de-noises the data, making the fraction of its variance that a model can explain higher, it reduces the number of participants over which to estimate parameters. And indeed, the bottom model is more sensitive: while the estimates are very similar across models, the strength of the statistical tests against zero are stronger in the model with all participants. **B., C., and D.** follow the exact structure of Fig. 4, but use all of the baboon data on the geometry task instead of first filtering on success rate to model a homogeneous population.