Mathias Sablé–Meyer
supervised by
Stanislas Dehaene

Human Cognition of Geometric Shapes, A Window into the Mental Representation of Abstract Concepts

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée au Collège de France

La cognition des formes géométriques chez l'humain, une fenêtre sur la représentation mentale des concepts abstraits

Soutenue par
Mathias Sablé-Meyer
Le 18/10/2022

École doctorale n° 158
Cerveau, Cognition, Comportement

Spécialité
**Neurosciences ;
Sciences Cognitives**

Composition du jury :

Pascal, MAMASSIAN Directeur de Recherche CNRS, École Normale Supérieure	*Président & Examinateur,*
Adrien, MEGUERDITCHIAN Chargé de Recherche CNRS, Aix Marseille Université	Rapporteur & *Examinateur,*
Nora, NEWCOMBE Professor of Psychology, Temple University	*Rapporteur,*
Tecumseh, FITCH Professor of Cognitive Biology, University of Vienna	*Examinateur,*
Stanislas, DEHAENE Professeur de psychologie cognitive, Collège de France	*Directeur de thèse*

Table of Contents

Formal Requirements

Titles

English title: "Human cognition of geometric shapes, a window into the mental representation of abstract concepts"

Titre français : « La cognition des formes géométriques chez l'humain, une fenêtre sur la représentation mentale des concepts abstraits »

English Summary (4000 characters)

Natural language is not the only hallmark of humans' singular cognitive abilities. In my PhD work, and in line with the language of thought literature, I argue that there could exist several internal languages; in particular I show that cognition involving geometric shapes requires a set of discrete, symbolic mental representations that act as an internal mental language. Under that view, perceiving a shape is comparable to the process of program induction: finding the shape's shortest representation in the internal mental language. I test this hypothesis by studying the perception and working memory of geometric shapes in a series of studies featuring both cross-cultural, developmental, and cross-species behavioral experiments, diverse neuroimaging techniques in adults and infants, and computational modeling with both symbolic models and neural networks.

First, leveraging an intruder task with quadrilateral shapes, I show

that humans of diverse education, age and culture share a sense of geometric complexity wherein some shapes are consistently simpler than others, but that baboons lack this sense even after adequate extensive training. Artificial neural networks of object recognition fit all baboons' data well, but explaining humans' behavior requires using additional symbolic properties such as the presence of right angles. Symbolic models of the humans' behavior generalize to several related tasks, indicating robustness in the choice of primitives the models leverage. This sharp dissociation suggests that two strategies are available to encode geometric shapes: both humans and non-human primates share a perceptual strategy, well captured by models of the ventral visual pathway, but only human have access to an exact, symbolic strategy. Using magnetoencephalography during a passive shape perception task, I confirm this by identifying the neural dynamics of both a visual and a symbolic strategy. I shed light on an early occipital response that resembles the neural network models, followed by a slower, more dorsal response similar to the symbolic model. At the same time, using electroencephalography, I provide preliminary evidence for the existence of the symbolic strategy already in three-month-old infants. Using functional MRI, I argue that geometric shape perception also recruits areas previously argued to belong to a non-linguistic network for mathematical reasoning.

Going beyond a small set of highly controlled shapes, I set to try to account for all geometric shapes produced by humans: I make a concrete proposition for a generative mental language of geometric shapes inspired by attested human productions. I argue that perceiving a shape means finding the shortest program in this language that generates the shape, connecting shape perception in humans to the literature on program induction. With this language, I show that program induction is in principle a tractable problem using the DreamCoder algorithm I helped implement. Then, I use this language to generate shapes of increasingly high complexity and show that humans' performance in a match-to-sample task for a shape correlate with the length of its shortest program. To decouple this result from the exact language proposition, I also derive more general additive rules that any alternative languages must obey and I provide empirical evidence for the validity of these laws. Finally, I show that if a simple perceptual strategy is enough to answer, participants may not deploy a symbolic strategy even if it exists – which suggests new ways of making human and non-human primates converge, and in turn refine theories.

Taken all together, these results support the existence of a discrete, symbolic set of mental representations for geometric shapes, which coexists with bottom-up visual representations. This work paves the way for integrated neuro-symbolic models of shape perception while challenging currently dominant object-recognition based models of visual perception.

English Summary (1000 characters)

Natural language is not the only hallmark of humans' singular cognitive abilities. I propose that cognition involving geometric shapes requires a set of discrete, symbolic mental representations that act as a mental language, and that perceiving a shape is performing program induction: finding the shape's shortest representation in this internal language. First, I show that all humans share a sense of geometric complexity, but that baboons lack this sense even after training. Artificial neural networks of object recognition fit baboons' data, but explaining humans' behavior requires using additional symbolic properties such as the presence of right angles. Then, I identify the neural dynamics of both a visual and a symbolic strategy of shape perception using brain imaging methods, and I provide preliminary evidence for the existence of the symbolic strategy in infants. Finally, I propose and test a mental language of geometric shapes inspired by attested human geometric productions.

Résumé français (4000 caractères)

Le langage naturel n'est pas la seule capacité cognitive qui distingue les humains. Dans la lignée de la littérature sur les langages de la pensée, je défends l'idée qu'existent de nombreux langages internes qui participent de la singularité cognitive humaine. Je concentre mon travail sur la perception des formes géométriques pour montrer qu'elle requiert un ensemble de représentations mentales symboliques qui ressemblent à un langage mental interne ; selon cette hypothèse, la perception d'une forme géométrique déclenche un mécanisme cognitif d'induction de programme. Je teste cette proposition via une série d'expériences comportementales cross-culturelles, développementales, et interespèces ; via plusieurs méthodologies d'imagerie cérébrale chez l'adulte et le nourrisson ; et via de la modélisation computationnelle soit formelle, soit fondée sur des réseaux

de neurones artificiels.

Dans une tâche de détection d'intrus parmi des quadrilatères, le comportement d'humains d'âges, éducations et cultures variées est homogène, de plus il se distingue du comportement de babouins même après un entrainement adéquat et intensif. Des réseaux de neurones artificiels pour la reconnaissance d'objets sont de bons modèles des babouins, mais il faut y adjoindre des propriétés symboliques pour rendre compte du comportement des humains. En contrastant ces modèles, il apparaît que pour réaliser cette tâche, les deux populations partagent des mécanismes visuels, mais que seuls les humains ont en plus accès à des propriétés symboliques. Chez l'humain, je confirme cette théorie en modélisant les processus neuronaux de la perception de formes obtenus en magnétoencéphalographie en rendant compte d'une première réponse cérébrale, occipitale et rapide, qui ressemble aux modèles fondés sur les réseaux de neurones, suivie d'une réponse lente et dorso-frontale qui partage des propriétés avec le modèle symbolique. Je fournis aussi des preuves préliminaires de l'existence de la stratégie symbolique chez le nourrisson de trois mois. Grâce à une étude en IRM fonctionnelle, je montre également que la perception de formes géométriques fait aussi appel à un réseau cérébral précédemment identifié pour la cognition mathématique, et distinct du réseau classique du langage.

Pour ne pas me limiter à un petit ensemble de formes, je tente de modéliser toutes les formes géométriques : à cette fin, je propose une version explicite de langage mental de la géométrie en m'appuyant sur des productions géométriques humaines attestées. Je propose de relier le problème de la perception des formes géométriques au problème de l'induction de programme : étant donné uniquement le résultat d'un programme, il s'agit d'en trouver le meilleur candidat dans un langage donné. Grâce à cette proposition concrète de langage, il est possible de générer les formes de complexité croissante : dans une tâche de reconnaissance différée, je vérifie alors que les performances dépendent de la longueur du programme le plus court, même après avoir corrigé pour les propriétés visuelles de bas niveaux. Pour découpler ce résultat de ma proposition concrète de langage, je dérive aussi des lois additives de la perception des formes géométriques que doit respecter toute proposition alternative, et je montre qu'elles sont vérifiées dans une autre expérience. Enfin, je montre que lorsqu'il existe une stratégie visuelle aussi appropriée qu'une stratégie symbolique,

les humains ne déploient pas la stratégie symbolique, ce qui suggère de nouvelles façons de concevoir des expériences comparatives entre les humains et des primates non humains.

Ensemble, ces résultats suggèrent l'existence d'un ensemble discret et symbolique de représentations mentales pour les formes géométriques qui coexiste avec une représentation visuelle classique. Ce travail ouvre la voie vers des modèles neuro-symboliques de la perception des formes, et remet en question les modèles dominants qui n'utilisent que la représentation visuelle.

Résumé français (1000 caractères)

Le langage naturel n'est pas la seule capacité cognitive qui distingue les humains. Dans cette thèse, je défends l'idée que la cognition humaine des formes géométriques passe par un langage mental qui leur est exclusif. Dans une tâche de détection d'intrus, le comportement d'humains est homogène et se distingue de celui de babouins dans son utilisation de propriétés symboliques comme la présence d'angles droits. En contrastant un modèle fondé sur des réseaux de neurones et un modèle symbolique, je rends compte de cette différence : les deux populations partagent les mécanismes visuels, mais seuls les humains ont accès aux propriétés symboliques. Chez l'humain, je confirme cette théorie en modélisant les processus neuronaux de la perception de formes obtenus en magnétoencéphalographie et IRM fonctionnelle. Je fournis aussi des preuves préliminaires de l'existence de la stratégie symbolique chez le nourrisson. Enfin, je propose une version explicite de langage mental de la géométrie.

Originality of the Work

All the work presented in this manuscript is mine and was performed during my PhD. The writing is mostly original to the thesis, except for chapters 1 and 4, which were respectively published as follows:

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

- **Sablé-Meyer, M.**, Ellis, K., Tenenbaum, J., & Dehaene, S. (2021). A language of thought for the mental representation of geometric shapes. Minor revisions to be submitted to the journal "Cognitive Psychology"

Some elements of the introduction and conclusion were borrowed from the article hereafter, as well as from a grant proposal submitted to the FYSSEN Foundation.

- 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.

In addition to the work presented in this manuscript, two additional first-author articles were published, one from continued collaboration after an internship supervised by Pr. Salvador Mascarenhas at ENS, and the other one in a collaboration with Lorenzo Ciccione, a friend and colleague:

- **Sablé-Meyer, M.**, & Mascarenhas, S. (2021). Indirect illusory inferences from disjunction: a new bridge between deductive inference and representativeness. Review of Philosophy and Psychology, 1-26.

- Ciccione, L., **Sablé-Meyer, M.**, & Dehaene, S. (2022). Analyzing the misperception of exponential growth in graphs. Cognition, 225, 105112.

Ainsi certes, nous ne pourrions jamais connaître le triangle géométrique par celui que nous voyons tracé sur le papier, si notre esprit d'ailleurs n'en avait eu l'idée

Descartes, Cinq Rép., AT VII, 382 ; OC IV-1, 574.

What is a point? Euclid famously kickstarted geometry as we know it today with his definition n°1, "A point is that which has no part" ("Σημειον εστιν, ου μερος ουθεν"; (Byrne & Euclid, 1847)). There is no physical entity to which this definition would apply; a point must therefore exist only in the mind of the beholder. What mental and neural mechanisms make it possible to entertain such concepts? Are those mechanisms only available to humans, and are they deeply tied to natural language? In the present work, I propose that even the simplest geometric concepts are uniquely human, and that they lie at the foundation of a rich generative system of shapes that behaves like an internal mental language. I argue that humans across ages, cultures and education levels share this sense of geometry, and I explore its neural mechanisms.

Observe a child drawing a person: the drawing will surely look very different from the visual percept. What is depicted has a structural and conceptual relation to the visual world, but is a poor rendition of what is perceived. **Figure 1** presents three examples of drawings from children: the top one is an explicit "draw a person" test taken by a four-and-a-half-year-old child, while the two bottom ones, taken from (Saito et al., 2014), were respectively produced by a 2y5m girl and 2y7m boy, in free drawings sessions, and were described a posteriori by the young

Figure 1: Examples of visual productions by children. Bottom row borrowed from (Saito et al., 2014)

artists respectively as "railroad" and "train": a fitting drawing for that age, but quite a perceptual leap of faith.

The obvious motor limitations cannot account for the *nature* of what is visually implausible, nor the striking consistency in choosing what to depict. Choosing what to depict could come from many sources during development and might be very dependent on culture, yet the very fact that an implausible representation constitutes an acceptable production both for the child and for its tutors indicates a remarkable ability to engage in a make-believe state where erratic lines on paper represent something in the world.

But if perception, at a high level, superimposes structures onto visual percepts, then productions such as stick figures for a person can be seen as producing the underlying structure instead of the perceived object. In fact, some common advice about learning to draw as an adult involves copying other drawings upside down (both the original and the copy) to avoid being distracted by what we recognize and instead focus on what we see.

Studying the perception of complex scenes and rich visual environments requires controlling for innumerable confounding factors. Luckily, human production has featured for millennia depictions that are not obviously related to visual percepts *at all*, but rather seem like abstract productions in and of themselves: geometric shapes. Whether engraved on stones or laid on the ground with giant boulders, geometric productions cover the world and span over at least half a million years – and they are unmistakably human. Furthermore, the shapes featured can be very different, both across paleonto-archeological sites and within sites: Lascaux, in the south of France, has been argued to feature over 13 different types of non-symbolic representations, in-

cluding crosshatches and squares (Von Petzinger, 2009).

In this thesis I explore the possibility that humans, and possibly humans only, mentally represent shapes as much more than their visual impressions by superimposing structure onto their perception. More specifically, I argue that all humans possess the ability to represent geometric shapes using an internal language specialized for that purpose. I argue that they naturally deploy this mechanism when faced with geometric shapes, in addition to other mechanisms of perception. As a contrasting point, I provide evidence that baboons either do not possess this competence, or do not deploy it even when it would be extremely useful. I model the difference between humans and non-human primates with two very different classes of models.

First, I will provide a brief overview of the production of geometric shapes across history and culture, together with a review of the experimental cognitive science literature regarding geometry in humans. Then I will introduce the Language of Thought Hypothesis (LOTH), with an emphasis on "programs" as candidates for mental representations. I will connect the LOTH to cognitive science research performed using information theory, with introductory elements on Kolmogorov complexity and Minimum Description Length (MDL). Then I will connect the LOTH to the problem of program induction in computer science, and provide a general introduction to a few dominants programming language paradigms. Finally, will outline and briefly summarize the structure of the work reported in this document.

Evidence of Geometrical Productions in Humans

Paleontological Evidence

Evidence for abstract concepts of geometry, including rectilinearity, parallelism, perpendicularity and symmetries, is widespread throughout prehistory. About 70,000 years ago, **Homo Sapiens** at Blombos cave carved a piece of ocher with three interlocking sets of parallel lines forming equilateral triangles, diamonds and hexagons (Henshilwood et al., 2002). Much earlier, approximately 540,000 years ago, **homo Erectus** in Java carved a zig-zag pattern on a shell (J. C. A. Joordens et al., 2015). Such a zig-zag may look simple, but it approximately respects geometric constraints of equal lengths, equal

Figure 2: Geometric shapes in human cultural history. A, examples of small- and large-scale geometric drawings and constructions (From left to right and top to bottom: an engraved slab from Blombos caves dating about 70.000 years ago (Henshilwood et al., 2002); zigzag pattern engraved on a shell in Java approximately 540.000 years ago (J. C. A. Joordens et al., 2015); Boscawen-Ûn's Bronze Age elliptical cromlech in Cornwall; spiral stone engraving on Signal Hill in Saguaro National Park, Arizona, dated 550 to 1550 years ago; geometrical shapes below the painting of a Megaloceros in Lascaux, France, typically dated to be 17,000 years old)

angles and parallelism, and is undoubtedly attributed to the *homo* genus.

Even earlier, about ~1.8 million years ago, ancient humans have been carving spheroids (sphere-like stones) and bifaces — stones possessing two orthogonal planes of symmetry (Le Tensorer, 2006). The vast number of bifaces, their near-perfect symmetry (which is not required for them to operate as efficient tools (Le Tensorer, 2006)), and the archeological evidence that many were never used as tools, suggest that an aesthetic drive for symmetry was already present in ancient humans.

What is more, a single site may feature many different human productions, mixing representative ones with more geometric ones. Lascaux, home of the famous Megaloceros and its accompanying geometric shapes (featured in **Figure 15**), features 13 types of geometric shapes in a taxonomy of 26 possible productions, and there are as many as 153 sites in France alone which feature at least one geometric production (Von Petzinger, 2009).

Anthropological and Cross-Cultural Evidence

Contemporary cognitive anthropology corroborates those findings. Cognitive tests performed in relatively isolated human groups such as the Mundurucu from the Amazon, the Himba from Namibia, or indigenous groups from Northern Australia, show that in the absence of formal western education in mathematics, adults and even children already possess strong intuitions of numerical and geometric concepts (Amalric et al., 2017; Butterworth et al., 2008; Dehaene et al., 2006; Izard et al., 2011; Pica et al., 2004; Sablé-Meyer, Fagot, et al., 2021).

Indeed, adults without formal western education share with Western preschoolers a large repertoire of abstract geometric concepts (Dehaene et al., 2006) and use them to capture the regularities in spatial sequences (Amalric et al., 2017) and quadrilateral shapes such as squares or parallelograms (Sablé-Meyer, Fagot, et al., 2021). They also possess sophisticated intuitions of how parallel lines behave under planar and spherical geometry, such as the unicity of a parallel line passing through a given point on the plane (Izard et al., 2011).

11

Developmental Evidence

Another piece of evidence arises from developmental data. Preschoolers and even infants have been shown to possess sophisticated intuitions of space (Hermer & Spelke, 1994; Landau et al., 1981; Newcombe et al., 2005), spatial sequences (Amalric et al., 2017), and mirror symmetry (Bornstein et al., 1978). Indeed, preschoolers' drawings already show a tendency to represent abstract properties of objects rather than the object itself. Although they look primitive, drawings of a house as a triangle on top of a square, or a person as a stick figure with a round head, suggest a remarkable capacity for abstracting away from the actual shape and attending to its principal axes, at the expense of realism. Numerous tests leverage this geometric competence to assess a child's cognitive development by counting the number of correct or incorrect abstract properties, for instance when asked to draw a person (Goodenough, 1926; Harris, 1963; Long et al., 2019; Prewett et al., 1988; Reynolds & Hickman, 2004). There is some evidence, however limited, that this ability may be specifically human: when given pencils or a tablet computer, other non-human primates do not draw any abstract shapes or recognizable figures, but mostly generate shapeless scribbles (Saito et al., 2014; Tanaka et al., 2003).

From these observations, one is left wondering what could be the origins and evolutionary advantage of the human competence for abstract geometry. I propose that it is a specific case, in the visual domain, of a general human ability to decompose complex percepts and ideas into composable, reusable parts – an ability which led to a massive enhancement of human productions, from architecture to tool building, and of the capacity to understand abstract features of the environment.

Geometry as a Research Topic in Psychology

As early as in Plato's *Meno*, Socrates is reported to have led an uneducated slave in a Greek household throughout a long maieutic process to reach accurate conclusions about the relation between the area of two squares; Socrates then concluded that "his soul must have always possessed this knowledge". Geometry as a segue into a deeper understanding of human cognition is, therefore, a process at least 2350 years old.

Research projects involving geometric shapes are deeply intertwined with many other research topics. The first, remarkably influential research project that intersects with geometric shape perception is that of Gestalt Principles of perception. Shape perception has also leaped forward thanks to an extensive research on medial axis theory, or skeleton, of shapes. Spatial Cognition, the ability to orient oneself in space, has also found a major role of geometric features in spatial representations of animals. And of course, the literature on mathematical cognition and symbolic thinking has turned its attention on geometric shapes throughout its entire history. I provide below a brief review, together with pointers to more extensive review, of each of these research projects and how they relate to the work I focus on in the rest of the manuscript.

Gestalt Perception
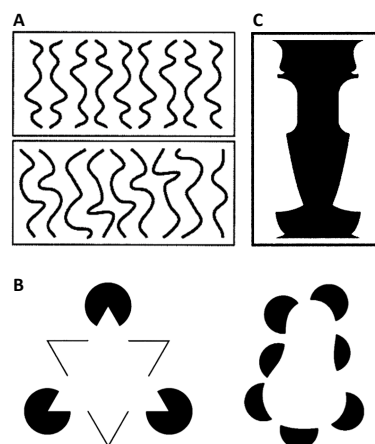
Figure 3: **A.** Perceptual grouping induced by various properties. Observe how symmetry (on the top) and parallelism (on the bottom) structure the perception of isolated line. **B.** Illusory, subjective contour integration, borrowed from (Kanizsa, 1976). Despite the striking classical example on the left, the central observations were about good continuation and is not related to geometric shapes. **C.** Example of figure-ground stimulus borrowed from (Peterson et al., 1991) which contrasts such stimuli with their upside-down reversed counterpart and argues that object recognition must precede figure-ground computations in vision.

It is remarkable how effortlessly the visual system of humans' group parts of a visual scene into individual, coherent objects. Historically, a set of rules have been devised from introspection, these rules form the basis of "Gestalt Perception": they were first described in (Wertheimer, 1912) and greatly impacted vision research. In 2012, for the anniversary of 100 years since the publication of the original article, two ex-

tensive reviews were published (Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012).

Gestalt Psychology puts forward many features of visual scenes which give rise to percepts that go beyond the visual input. In particular, the literature concerns itself with (i) perceptual grouping, where items are perceived to belong to coherent groups by virtue of various properties (see **Figure 3.A** for examples of grouping induced by symmetry (top) and parallelism (bottom))., (ii) contour integration, where separate items are perceived as a whole outlining a contour (see **Figure 3.B** for example of illusory contour integration), and (iii) figure-ground relation, where ambiguity of simple stimuli are resolved by assuming a superposition of different layers (see **Figure 3.C** for an example of figure-ground ambiguity). All of these phenomena have received extensive experimental support, as well as neural mechanism accounts, extensively reported in (Wagemans, Elder, et al., 2012).

In addition to the evidence in support of the Gestalt principles, the principles have received a lot of attention from a conceptual and theoretical point of view, in particular framing early observations in terms of information processing. Gestalt principles have been modeled in terms of attractors in dynamical systems (Leeuwen, 1990), possibly optimizing for neural resources, self-adapting properties, Bayesian inference given some information (Kersten et al., 2004), and optimal coding efficiency (starting with (Attneave, 1954)): an in-depth history, including the successes, limitations, and differences between these models is provided (Wagemans, Feldman, et al., 2012).

One important notion is that of good continuation and path integration: good continuation is a Gestalt grouping principle stating that separated objects that form a smooth contour tend to be grouped together (Wertheimer, 1938). Models and neural evidence for path integration as a mechanism for good continuation, notably (Hubel & Wiesel, 1965; Iacaruso et al., 2017; Ledgeway et al., 2005), are developed in **chapter 4** with regard to the categorical perception of right angles suggested by non-intersection segments.

The literature on Gestalt perception is immense, and while it has proven very insightful and fertile in vision research, it is not often clear what it has to say about geometric shapes specifically. Examples such as the Kanizsa triangle in **Figure 3.B** (left) could lead one to believe that geometric shapes play a central role in the field, but as

highlighted in the original article about the example on the right in the same figure, "Geometric regularity is not a necessary condition for the formation of subjective surfaces and contours. Amorphous shapes are possible and irregular figures can generate contours" (Kanizsa, 1976). That holds true for many other phenomena described, where some geometric properties may prove useful, but there is no account of geometric shapes as such.

Medial Axis Theory and shape skeleton
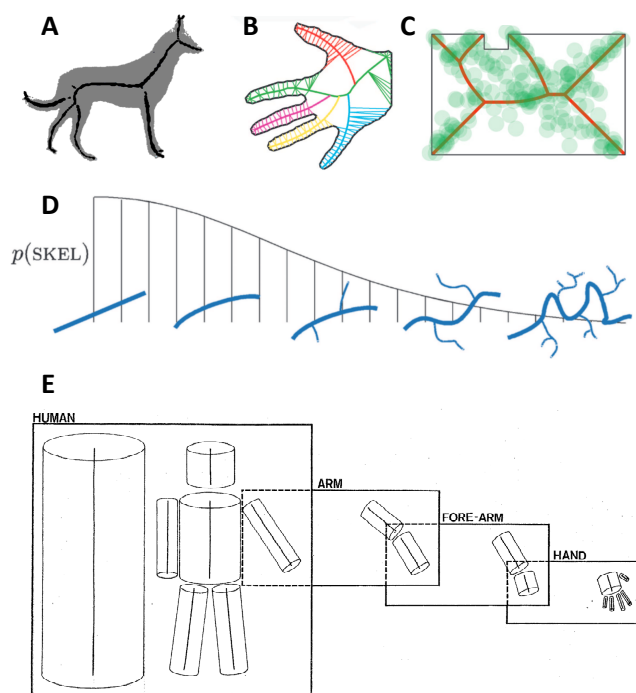
Figure 4: **Examples of medial axis theory shapes. A.** Example borrowed from (Philbrick, 1966) introducing the first procedure for deriving the skeleton of a shape. **B.** Example of shape skeleton derived using Bayesian estimator (borrowed from (Feldman & Singh, 2006)). Notice how the colors, corresponding to main skeleton axes as opposed to minor "ribs", correspond well to perceptually distinct parts of the shape. **C.** Results of an unguided shape tapping (borrowed from (Firestone & Scholl, 2014)). Participants were presented with a shape on a portable touch-screen enabled tabled and could tap anywhere they wanted: the density of tapping corresponds to important locations according to the Medial Axis Theory. **D.** Example of probability distribution over branching when finding the optimal medial axis from a Bayesian perspective, connecting this approach to the notion of Minimum Description Length (borrowed from (Feldman & Singh, 2006)). **E.** Recursive decomposition of an approximate 3D "human" into subparts featuring systematic medial-axis structure; borrowed from (Marr & Nishihara, 1978)

Another very influential idea in shape processing is medial axis theory, or shape skeletons. It was first introduced in (Blum, 1967; Philbrick, 1966), see **Figure 4.A**, as a tool to enable automatic analysis of visual

scenes by computers, and it provides a dense and rich representation of shapes that can then be used either in computational tools or for neural modeling.

Medial axis theory has a long history in vision science: it has been shown to predict neural activity in the visual cortex (Lee, 2003), and in fact neural firing patterns in cats has been cited as an inspiration when the method was introduced (Hubel & Wiesel, 1965). The complexity of a shape's medial-point description, a representation similar to its skeleton, has been argued to correlates with behavior in a detection task using Gabor patches to outline a shape (Kovács et al., 1998). Even when merely tapping inside a shape without any additional instructions, participants tend to choose locations that fall along the medial axes, as displayed in **Figure 4.C** (Firestone & Scholl, 2014).

Medial axis theory also bears close relations with theories of the cognitive processes underlying whole-part decomposition, see **Figure 4.E** from (Marr & Nishihara, 1978) and (Biederman, 1987); in particular, it has been argued that the inferred axes separating an object into parts systematically cross skeleton axes (Singh et al., 1999)

The first models devised to compute the skeleton of a shape were mathematical approximations of "grassfire propagation" (planar diffusion of fire, in a homogenous material, from point sources), and could otherwise be mechanically implemented with successive defocusing of an optic mechanism. But these early methods were not robust to noise on the contours, and produce implausible skeleton in the presence of jagged boundaries.

To break free of these issues, and in line with the idea of perception as Bayesian inference (Kersten et al., 2004; Knill & Richards, 1996), a Bayesian method for estimating the skeleton of shapes has been proposed (Feldman & Singh, 2006), see **Figure 4.B** for an example of decomposition of a (mostly smooth) hand. At its core, the model optimizes the trade-off between the complexity of a skeleton (the Bayesian prior, see **Figure 4.C**) as approximated by its number of branches and its curvature, and how well that skeleton fits a given shape (the Bayesian likelihood).

Such methods are at the core of information-theoretic approach of understanding visual complexity, started with (Attneave, 1954); see also (Donderi, 2006) for a review of these approaches. Information-theoretic considerations are also central to recent modeling of Gestalt

principle, see (Wagemans, Feldman, et al., 2012). We reconsider these notions again in **chapter 5** when I compare the predictions of medial axis theory to that of a symbolic generative language of geometric shapes.

I want to point out that while skeletons of shapes have proven remarkably useful for many sorts of shapes, and in particular contours of real objects as depicted in **Figure 4.A and .B**, they may not be the most appropriate for dealing with geometric shapes specifically. This observation is discussed more extensively in **chapter 5**, where I consider shapes such as circles and spirals: the former has a non-informative skeleton, and the latter has no skeleton per se.

Geometry and Spatial Representation

Humans, as well as other animals, are able to navigate their environment: in order to do so, they must form internal mapping of the surrounding space, an ability referred to as spatial representation. There are many mechanisms at play in spatial representation, and many animals have been argued to possess some form of spatial representation: of interest to this work is the use of purely geometric information as one of the key mechanisms, first highlighted in (Cheng, 1986).
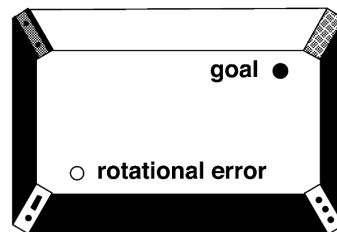
Figure 5: Illustration of the original experiment from (Cheng, 1986): the goal and its counterpart after invariance by rotation are indicated with dots inside a rectangular room. Despite the four different corners, and the disambiguating white wall, rats' pattern of error support a geometry-based spatial representation system.

In the experiment featured in (Cheng, 1986), rats were looking for a reward in a rectangular room with distinctive features on all four corners, one wall colored differently, and two corners with distinctive smells. The patterns of reward-seeking mistakes indicate that rats confused two diagonally opposite corners, indicating that while they could use the general geometry of the room as a cue, they couldn't narrow the research more.

This observation has spawned an important literature about the possible existence of a cognitive module for geometry, and its relation to other, nongeometric features used in spatial representations. In-depth review of the literature on the use of geometry as a tool for spatial re-orientation can be found in (Cheng et al., 2013), (Cheng & Newcombe, 2005) and (Twyman & Newcombe, 2010).

A central debate in the literature on geometry and spatial representation has been about the existence of a cognitive module for geometry, and how it would integrate with other modules, memory, and non-geometric feature perception. Under the "geometry module" view of spatial orientation, humans (and other animals) are endowed with a module, in the sense of Fodor (Fodor, 1983), which would underlie the geometric aspect of spatial orientation: this module would be limited in abilities, but through education and language could serve as the foundation of higher-level geometric concepts in humans – while constraining how these concepts develop (De Cruz, 2009).

The literature on spatial cognition is gigantic, and I confess to having a superficial knowledge of its ramifications outside of its very direct relation to the cognition geometric shapes in and of themselves (see next section). Indeed, what is at the center of the work I present here is cognition *about* geometric shapes, where the mental representations are representations *of* the shapes, rather than geometric properties being *used as part of* spatial orientation. I am interested in understanding how objects such as "a quadrilateral with exactly one right angle" (**chapter 1**), or "a square of circles" (**chapter 5**) are represented. In principle such geometric objects could be used for spatial navigation (resp. in a room with exactly one right angle, or in a sequence of circular rooms arranged in a square, for example), but the literature on spatial orientation has mostly focused on either fully regular or fully irregular quadrilateralsé and triangles (for an exception see (Hupbach & Nadel, 2005)), across many species or age ranges, and tried to understand how the geometric features interact with spatial orientation.

Symbolic Thought, Mathematics, and Core Knowledge

Unlike many visual objects and categories, geometric shapes have normative definitions: a circle has a number of defining features, of which features like its color or its position are not. Of course, real-world representations of circles are by definition never exact (e.g., because no

curvature is ever mathematically constant), but under some assumptions about the inference abilities of humans such considerations can be mostly ignored. Thanks to this, geometric shapes can be used to probe the understanding of defining features (Satlow & Newcombe, 1998). But what are the mechanisms that enable our ability to understand geometric shapes? Do other animals share this sense? Is it rooted in natural language, or perhaps abilities for either approximate numbers, or proto-mathematics?

Jean Piaget published two seminal books in 1948, "La représentation de l'espace chez l'enfant" and "La géométrie spontanée de l'enfant": both books break down cognitive development of children regarding spatial and geometric concepts into successive stages, where each stage is associated with the ability to manipulate new concepts, or the same concepts more and more abstractly. For example, when asked to compare the length of two immovable objects set apart, children in the first stage might answer intuitively, while children in later stages may rely on keeping hands at a fixed distance while moving between the two objects, or use a third, reference object to compare the two target objects, successively. Piaget's work is very empirically driven but provides little theoretical insights as to the nature of the cognition in geometry – instead, it focuses on taxonomizing "developmental stages" related to many properties such as measurements, metrics for angles and curves.

Piaget puts forward the theory that development of geometric abilities goes through a specific order, wherein topological properties are constructed first (nearness, enclosure, etc.), and then gradually Euclidean properties emerge as relation between shapes become increasingly important. In that work and in many subsequent studies, the question of the Euclidean nature of intuitive geometry in humans was central; the proposition is that shape is the defining property of a geometrical object while size, location, rotation and sense are disregarded.
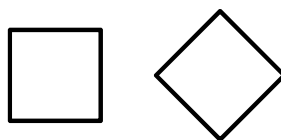
Figure 6: Retraced from (Mach, 1914); "two figures may be geometrically congruent, […] but could never be recognized as the same."

Yet, it has been noted as early as the beginning of the XXth century that

rotations do impact identification of a shape (Mach, 1914), by contrasting the introspective perception of a square against that of a rotated square, which is perceived categorically differently. Indeed, while the two shapes featured in **Figure 6** are identical from the point of view of Euclidean geometry (both are squares), the second one tends to be perceived as categorically different (and preschoolers may fail to identify it as a square in a set of shapes (Halat & Dağli, 2016)) to the point that a 45° rotated square may colloquially be referred to as a "diamond". This holds true despite extensive empirical evidence that geometric objects have a rich internal representation which allows mental operations. Roger Shepard in (Shepard & Metzler, 1971) famously showed that the response time in deciding whether two rotated 3D tetris-like objects (as shown in **Figure 7**) were identical almost perfectly correlated with the angular difference. This correlation strongly supports the idea that not only the mental representation carries information about the orientation of the objects, but participants are performing mental rotations on the representation themselves.
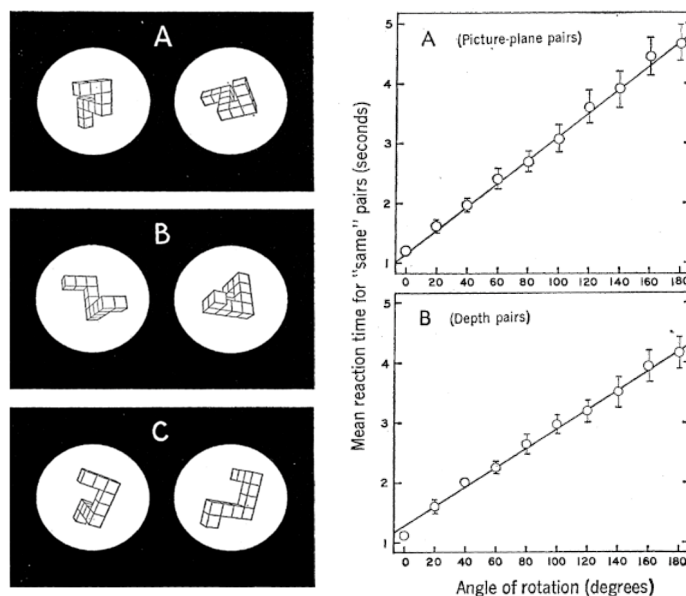
Figure 7: Borrowed from (Shepard & Metzler, 1971): stimuli and results from a mental rotation task involving three-dimensional tetris-like objects.

More recently, the claim from Piaget that symbolic geometric properties are learned in stages throughout education, from topological concepts such as "inside", or "closed", to Euclidean properties, has been contested from another angle. Experiments suggest that in fact adults and children from very different environments, including one without

formal Western education, perform remarkably similarly on many geometric intruder tests (Dehaene et al., 2006) with categories as diverse as topology, Euclidean properties, geometric figures, symmetry and sense, metric properties and geometric transformations. Interestingly, the language of the Mundurukus has few words for the geometric concepts, and therefore this article also suggests that natural language is not the vessel for geometric thoughts: a discussion of this Whorfian problem is provided in (Newcombe & Uttal, 2006).

Another important question is whether geometry forms a "core knowledge" (E. S. Spelke, 2003), i.e an evolutionary ancient, innate ability we could share with other animals. I provide some arguments weighting against this view in **chapter 1** and in **chapter 3**, but there is still a lot of evidence in favor of a version of this view, see (Tommasi et al., 2012) for a review. A more nuanced view is that geometry would result from the interaction of two lower-level systems: one dedicated to navigation, as discussed in the previous section, and one for visual forms and object recognition; those two systems, when interacting in humans, give rise to our ability for geometry (Carey, 2009; E. Spelke et al., 2010).

This proposition prompts the question of whether natural language is what enables this interaction. A similar hypothesis was formulated for mathematical cognition, with support from in several ways: notably because language plays a significant role in the development of spatial reasoning (Pyers et al., 2010), and because deaf individuals who did not learn a conventional language have a harder time manipulating exact numbers (Hyde et al., 2011; Spaepen et al., 2011, 2013). But recent neuroimaging studies find a systematic dissociation between language networks and math networks (Amalric & Dehaene, 2016, 2017). Furthermore, infants have been shown to be able to solve geometric tasks without being able to provide a justification for their decisions, either through language or gesture (Calero et al., 2019).

Another important question is why humans appear to agree that the "natural" form of geometry corresponds to the geometry entailed by Euclid's axiomatization. While the axioms formulated by Euclid are formulated in an abstract fashion, many of its constituents can be approximately drawn and observed, or can be experienced in space: one can draw a line and a point outside that line, and observe that it is impossible to draw two distinct lines that are both parallel to the first one

and go through the point[1].

To what extent, then, is our geometric knowledge constrained by visual experience? On the one hand of the spectrum, one might postulate that perception is the very process by which we internalize whether certain geometric statements are true or not (Giaquinto, 2005; Mumma, 2009): under that view then geometry must follow our visual experience. Yet not all visual information is relevant for geometrical concepts (e.g., size, or orientation). A review of the arguments regarding the question of the visual origin of Euclidean geometry is provided in (Izard, 2022).

Programs as Candidates for Mental Representations: A Take on the Language of Thought Hypothesis

> "Humans have a multi-domain capacity and proclivity to infer tree structures from strings, to a degree that is difficult or impossible for most non-human animal species"
>
> (Fitch, 2014)

Geometric Primitives of Cognition

When it comes to cognitive science, the first application of information theory to visual perception comes from (Attneave, 1954). He observed that most of a visual field is redundant, i.e. many portions may be hidden and yet successfully recovered, suggesting that the mental representation might compress the visual information. For geometric shapes specifically, the earliest theory of programs representing geometric shapes comes from Leeuwenberg and colleagues (E. L. Leeuwenberg, 1969, 1971; Boselie & Leeuwenberg, 1986), who proposed a formal coding language for 2- and 3-dimensional shapes. They argue that the mental representation of a shape is as complex as the smallest program in that language, a property I also defend in **chapter 5**. In fact, they already observe that some elements of

---

[1]Whether this holds true on curved surfaces or otherwise matters little with regard to the fact that a simple visual observation on a "base case" will strongly feed one's intuition about this axiom.

their proposed language are quite general, and that they could be applied to the compression of auditory sequences as well, a property that was found again in recent work on geometric sequences in the visual and auditory domain (Amalric et al., 2017; Planton et al., 2021). In addition to the theoretical claims, they provide some empirical support for their proposition, but conclude in saying, "[...] for the time being [the proposed coding procedure] will hardly lend itself to computer programming."

Independently, Leyton (Leyton, 1984, 2003) argued that the shapes that humans generate arise from a set of primitives (points, lines, planes) together with the repeated mental application of a sequence of group transformations that duplicate, stretch, rotate, or skew space. The proposal is mathematically remarkably elegant, and has shown to be very influential in the design of software for graphical design, but while that line of work, unlike the one from Leeuwenberg, does lend itself quite well to computer programming, it remained partially disconnected from the experimental psychophysical or neurophysiological literature on shape perception (for exceptions, see Brincat & Connor, 2004, 2006; Hung et al., 2012)

Kolmogorov Complexity & MDL

A long-standing cognitive hypothesis is that the brain excels at compressing information. Indeed, in the presence of structure in stimuli, either visual or auditory or other, participants' score improves in a wide variety of tasks. The first observation of this phenomenon comes from (Miller, 1956), who states: *"I have fallen into the custom of distinguishing between bits of information and chunks of information. [...] The span of immediate memory seems to be almost independent of the number of bits per chunk, at least over the range that has been examined to date."* Immediate examples include remembering words, where the main factor is the number of words and not the number of letters, but similar observations are pervasive in psychology.

A strong version of this hypothesis states that the brain finds structure in a richer way than chunking, and relies on generative (programming) languages: the complexity of a given piece of information is the length of the shortest program that generates that information. For now, this is underspecified, but this will become more concrete in a few sections. This hypothesis strongly connects to the predictive coding hypothesis,

as the ability to predict and generate can be deeply tied to the generative language. Furthermore, for probabilistic programing languages, this class of theories can account for both success and mistakes in human behavior, a crucial feature of a fitting theory of cognition, and can entertain the coexistence of several possible mental representation with different probabilities.

But a lot hinges on the choice of the programming language, as various propositions will make wildly different predictions and it's unclear that one can find a unifying proposition that can account for very different stimuli (auditory, visual, intuitive physics, etc.)

Thankfully, an entire research domain exists to study a related problem: information theory. While most of its applications are far from psychological considerations, it includes theoretical questions such as "what is the most information-efficient way to represent a given set and its elements", "how much information do these two representations share". I will introduce two related notion from that field and show how they relate. Then, I will relate these propositions back to cognition.

Kolmogorov Complexity

To characterize the complexity a priori of a given piece of data, information theory has come up with a useful metric: Kolmogorov complexity. Consider the following strings of zeros and ones: (i) "0101010101010101" and (ii) "0110011101101000". While they have the same number of zeros and ones, intuitively the first one has a "lower complexity" than the second, because it can be expressed succinctly with an algorithm akin to "repeat '01' 8 times", while the second requires an exact description. How can we systematically characterize this apparent difference? Kolmogorov's original proposition (Kolmogorov, 1963) appears convoluted by today's standards: for a sequence of 0s and 1s, it finds a set of indices such that the corresponding extracted subsequence is not random (for some statistical test), and measures complexity as the number of symbols required to define such a set of indices using set operations. The intuition is that the smallest such set for example (i) is going to be simple to build, and for (ii) will more or less requires encoding exact positions of 0s (or 1s).

This can be simplified: considering a programming language, such as

python, let's describe our sequences:

```python
def seq1():
  print("01"*8)

def seq2():
  print("0110011101101000")
```

The first program requires 13 characters, while the second requires 25, and in that sense the first sequence is shorter, or less complex. One might think that the specific choice of language matters enormously: python's ability to "multiply" strings, for example, is far from common, and having to unroll this operation with a while loop might get us over the length of the sequence, nullifying any compression effect!

The question becomes: is there a single language in which, for any sequence, the program for that sequence is the smallest in that language across all possible alternative languages? This is almost the case, since any language expressive enough can be used to implement any other language. Because of that, we can start from a very small computational device, such as a Universal Turing machine, and use that to measure the length of the smallest program for a given output. Maybe an alternative language would yield a smaller program, but we can implement that alternative device for a constant cost in the reference machine, and in doing so ensure that the length in the reference machine is at most the shortest program in the new machine plus a cost for the translation between the two machines.

Indeed, consider a new programing language, "python-prime", identical to python in everything except there is a single additional instruction "short()" which is semantically equivalent to `print("0110011101101000")`. Our sequence (ii) now has a complexity of 7 in that new language, while leaving the complexity of (i) untouched. This is puzzling, as it feels like this has made the "complex" sequence less complex than the other – but the Kolmogorov complexity doesn't depend on the choice of language! Now, observe that both "python" and "python-prime" can be implemented in a Turing machine, and that the cost of implementing "python-prime" will be greater than the cost of implementing "python". In fact, it will be greater almost exactly by the cost of implementing our sequence (ii), since it needs to define the "short()" instruction. Therefore, if relying on a python implementation, in the Turing machine the order

is again as expected, with (i) having a lower complexity than (ii).

However, these are upper bounds: the Turing machine itself can output these sequences without having to resort to implementing a complex language such as python, and a typical machine would do so in such a way that again, (i) has a lower complexity than (ii). Unfortunately, it is impossible de devise a function that would output the Kolmogorov complexity of an arbitrary string: it is uncomputable, and therefore researchers often resort to upper bounds. This limitation does not hold for the alternative metric introduced below: Minimal Description Length.

Minimum Description Length

There is no reason to believe that a minimal, universal Turing machine is a good candidate for cognitive processes, and therefore Kolmogorov complexity isn't the most suited metric to study humans' idea of complexity. Instead, it is useful to come up with specific propositions for languages for a given domain, typically referred to as DSL for Domain Specific Languages, and measure the **minimum description length (MDL)**: the length of the shortest program, *in that particular language*, that computes the desired output.

Traditionally, MDL minimization has been used in statistics for model selection, and has been described as a mathematical version of Occam's razor. The principle is that given a noiseless dataset and a generative language of models to account for the data, the best model is the one with the shortest description that generates the data. In practice, as data is often noisy, the goodness of fit needs to be considered, and the Bayesian Information Criterion (BIC) is used instead of the MDL.

Interestingly, in the absence of a formal generative language to target, it is possible to fall back to indirect measure of "minimum description length" using natural language: this has been successfully done in (Sun & Firestone, 2022). The downside of this approach is that there is no reason to think that natural language is the right language to encode these shapes, and therefore operating using "verbal description length" may be too indirect. Consider for example face recognition: while humans are very good at discriminating faces, the reason why two given faces are confusable or not may be hard to describe, as our ability for facial recognition is mostly cognitively impenetrable – and

therefore, and introspective "verbal description length" may be too indirect for actual complexity.

Cognition and Program Induction

If mental representations have program-like properties, then it is crucial to offer a theory of how programs are inferred: how do we go from the sensory inputs to the structured representation? What are good models or "mental program building"?

In computer science, the subfield tackles this question is that of "Program Induction": the problem of program induction is the problem of finding a program given, typically, a set of input-output example pairs. A first observation is that this is in principle an impossible problem: there are arbitrarily many programs that may work, but some will behave differently on new input – in the absence of which the notion of "correct program" cannot be decided. Given enough examples, the trivial program that encodes explicitly the input/output pairs becomes very costly, and MDL can be used as a selection strategy: the goal is to find the shortest program that works for the examples given.

A baseline for solving program induction is program enumeration: recursively enumerate all possible programs in a programming language's grammar until you hit a program that satisfies the constraints. But this procedure is a poor candidate for a cognitive implementation of program induction under the MDL hypothesis: if we consider that the complexity of a certain program is a function of its MDL, then under that baseline approach the complexity of *finding* that program would grow exponentially with its MDL.

However, several methods have been devised to try and keep this combinatorial explosion in check, initiative that stems primarily from the computer science literature so far. In particular, under the umbrella term "neuro-symbolic" models is the idea of bringing the advantages of neural networks to the problem of program synthesis – see (Chaudhuri et al., 2021) for direct application to program induction, and (Garcez & Lamb, 2020) for a more general review of unifying neural networks and symbolic representations.

When it comes to program induction, the unifying idea is to use neural networks to either (i) directly suggest programs for a given task, which can then be refined, or (ii) suggests smart enumeration strategies, or

policies, to make the brute-force search efficient, the idea featured in (Ellis et al., 2021) which I use in **chapter 5**.

The Possibility of Several Internal Languages
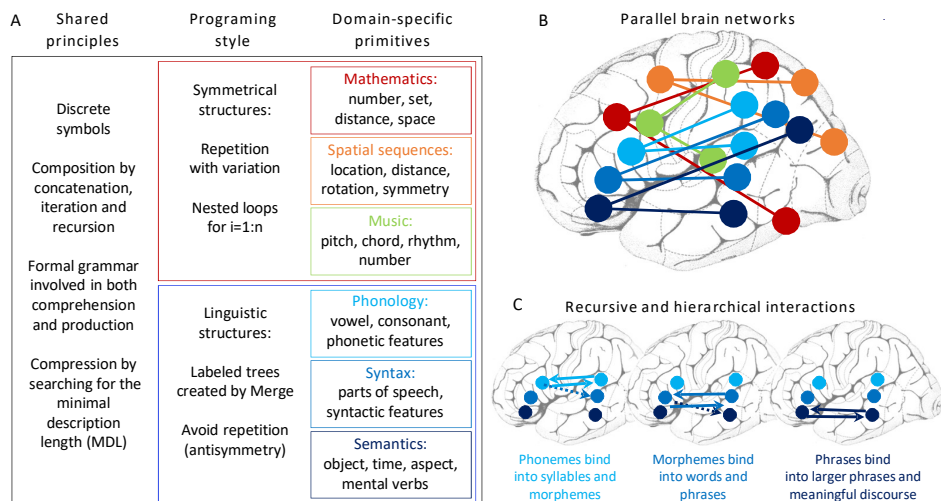
Figure 8: Taken from (Dehaene et al., 2022); **Main hypotheses of the present proposal. A, Multiple mental languages**, all based on symbols and recursive mental programs. Various domains of human cognition rests upon several distinct internal languages, each capable of compressing different types of inputs. Those languages share the same design principles, but differ in their primitives. Two broad styles may be distinguished: one based on the capacity to detect repetition with variation, thus appropriate to encode symmetrical patterns and mathematical structures; and another based on asymmetrical Merge, appropriate to encode the structures of communicative language at multiple levels (this part of the figure is inspired by a previous proposal by Pieter Hagoort [19]). Distinct languages emerge once these general instructions are combined with domain-specific primitives. **B, Multiple parallel cortical circuits**. The proposed languages do not rely on a single localized brain area for recursion, but on multiple parallel brain networks with primitives in temporal and parietal cortex and control structures in prefrontal cortex. For simplicity, only a left hemisphere is shown, but the postulated brain circuits are generally bilateral. **C, interactions within and between languages**. The mental expressions formed in one language become available as primitives for the same or for another language, thus allowing for the formation of complex recursive and hierarchical thoughts (bottom right).

The observations that humans are exceptional are, of course, not restricted to the production of geometric shapes. This can be referred to by using the term "human cognitive singularity" – the word singularity being used here in its standard meaning (the condition of being singular) as well as its mathematical sense (a point of abrupt change, a discontinuity in some parameters). Hominization was certainly a singularity in biological evolution, so much so that it opened up a new geological age: the Anthropocene. Even if evolution works by small continuous change (and sometimes it doesn't (Gould & Eldredge, 1977)), it led to a drastic cognitive change in humans.

Many accounts of this singularity have been proposed, notably spe-

cial competences for analogies (Hofstadter, 2001), theory of mind (Premack & Woodruff, 1978), teaching (Csibra & Gergely, 2009), cultural memory (Tomasello, 2000), and natural language (Morgan et al., 2015). What I argue here is that much simpler *tasks* will already yield visible differences, for example the ability to recognize an outlier amongst squares. I posit that all these competences are made possible thanks to emergence of an underlying ability to manipulate discrete symbols and compose them like a language to build increasingly complex mental representations. In the domain of geometric shape, I posit that when most animals perceive the difference between a square and a circle, they do so with classical visual recognition mechanisms, but that on top of that humans can encode these shapes in terms of exact properties. For the discrimination of squares and rectangles, the two strategies are similar, but the second one makes comparing a square and a visually close shape that is missing a right angle much easier, as shown in **chapter 1**.

This proposal finds its root in a tradition of emphasizing the role of recursive, nested structures in explaining humans' cognition, and language in particular (Dehaene et al., 2015; Fitch, 2014; Hauser et al., 2002; Penn et al., 2008). This ability has sometimes been referred to as "dendrophilia", the love for trees, to highlight how cognition seems to represent any data with tree structures. This hypothesis typically posited a single, core competence for recursion: see (Dehaene et al., 2022) for a recent description of a version of this hypothesis that features several languages, possibly sharing some core mechanisms but differing on others, and involving different brain areas.

Figure 8 summarizes the core hypotheses in favor of several languages of thought, fully developed in (Dehaene et al., 2022); however, in the work presented here I focus on the cognition of geometric shapes specifically, and only touch considerations about the link to other languages in **chapter 3** when the fMRI data provides evidence of a dissociation from natural language.

A Primer on Programing Languages

The following subsections are not crucial to any of the chapters in the work presented after; however, they help map the space of programming paradigms, and are useful to enrich one's view about the nature of the programs over which one could perform program induction.

Imperative and Functional Style

In imperative styles, programs are procedural, step-by-step description of the actions a machine must take in order to reach a desired goal. Most commonly used programing languages fall in this category, perhaps because typical machine code follows a similar pattern. Many high-level abstractions can be added, such as functions to avoid code redundancy, complex looping mechanisms, rich assignments, but ultimately the program describes a sequence of steps to reach a goal. Consider the following python program:

```python
radius = 2
pi = 3.14
circumference = 2*pi*radius
area = pi*radius*radius
print(f"For {radius=}, {area=} and {circumference=}")

# Output:
# For radius=2, area=12.56 and circumference=12.56
```

Approximately, the language parser turns this into an Abstract Syntax Tree (AST) and then executes the instructions in the order in which they appear: store a few values in variables, and then compute a new value using the previous one. The language allows for many different kinds of abstractions, such as loops, or functions – the following code computes the radius and area of several circles:

```python
pi = 3.14

def compute_area(radius): return pi*radius*radius

def compute_circumference(radius): return 2*pi*radius

for radius in range(3):
    area = compute_area(radius)
    circumference = compute_circumference(radius)
    print(f"For {radius=}, {area=} and {circumference=}")

# Output:
# For radius=0, area=0.0 and circumference=0.0
# For radius=1, area=3.14 and circumference=6.28
# For radius=2, area=12.56 and circumference=12.56
```

This modification introduces indirectness in the execution path, where the linear, top-down order is not respected anymore. But ultimately this program is still describing a sequence of steps to perform to get to a certain result, and makes use of high-level primitives to avoid code repetition. In this style, the state of the program is constantly modified by the code: variables can be changed by functions, the order of execution is very rigid and must be anticipated, and the attention of the programmer is on how to perform a task and how to track change of what is being performed as the execution flows.

Imperative style is often contrasted with functional style, in which functions are first-class citizens, the state of the program is largely immutable and instead the focus of the program is on what information is required, and what transformations this entails, in order to get to a certain result. Many modern languages include elements of functional paradigms, for example the ability to pass function as arguments to other functions, or to recursively call functions. In a functional paradigm, for example, it is possible to write the following code:

```python
def map(f, l):
  if l == []: return []
  else: return [f(l[0]), *map(f, l[1:])]

def pi(): return 3.14

def area(radius): return pi()*radius*radius

def circumference(radius): return 2*pi()*radius

for f in [area, circumference]:
  print(map(f, [0, 1, 2]))

# Output:
# [0.0, 3.14, 12.56]
# [0.0, 6.28, 12.56]
```

While the computation ends up being identical, the focus has shifted a lot: no variable is ever allocated, and instead at the core of this code is the notion of function: we can loop over functions ("for f in […]"), we can give functions as arguments to a function (the first argument of map is a function), and we can recursively call functions (map calls itself to range over the list). While uncommon in python, these con-

structions are pervasive in other languages.

Loosely, the imperative style of program is a derivative of the formalisms introduced with Turing's machines (Turing, 1936): instructions control the movement of a machine that reads and write symbols on tapes, each operation giving rise to the subsequent one. The code describes the movement of the machine. On a somewhat other end of the spectrum, λ-calculi, first described by Church a few years prior (Church, 1932), can be seen as the original idea for functional programming paradigms: everything is a function application, and the task of the programmer is to describe functions. Because actual computers are sequential in nature, they lend themselves more naturally to imperative paradigms, but the two models are as expressive and it is possible to express one in the other and vice versa.

Languages like C, Logo or Python are good examples of languages that are imperative at heart although they feature some properties of functional languages. Lisp is a good example of a programming language that is first and foremost functional. Functional languages more often offer high-level properties such as reflection (the ability of a program to introspect its own structure and behavior) and "metaprogramming" (the idea that programs themselves are part of what a program can manipulate, including itself). Such features could be relevant for cognitive considerations: without such techniques, it is very hard to conceive of a program such as "a square of squares" that could both (i) somehow manage to express the embedding and embedded "square" program identically, incurring a saving, and (ii) generalize this feature across arbitrary shapes. This consideration is featured in the discussion section of **chapter 5** together with considerations about the next type of programming paradigm, "logic programming."

Logic Programs

In logic programming paradigms, programs are ordered sets of truth statement expressed in a formal language akin to predicate logic. The origin of the paradigm dates back to debates in AI about whether knowledge should be represented in a declarative or procedural manner. Historically the first such language was "planner", and "prolog" constitutes an example of modern logic programming language.

Let's examine an example similar to the one above, written in prolog:

```
area(RADIUS,A) :- A is RADIUS*RADIUS*3.14.
circumference(RADIUS,C) :- C is 2*RADIUS*3.14.

map(_, [], []) :- true.
map(F, [X|Xs], [Y|Ys]) :- call(F, X, Y), map(F, Xs, Ys).

area(2,R).
map(area, [0, 1, 2], R).
map(circumference, [0, 1, 2], R).

% Output:
% R = 12.56.
% R = [0.0, 3.14, 12.56]
% R = [0.0, 6.28, 12.56]
```

In this example, we define in total three logical predicates. The first two define the ways in which two variables are related for that pair to define for example an area: the predicate "area(x,y)" is true if x and y are such that y=x*x*3.14. The map predicate is trickier, we have to break down the definition into two subcases. Given any function, the empty list maps onto the empty list, so we say that the predicate "map" is true whenever its last two arguments are the empty list. Another case in which it is true is when its first argument is a function, the other two are lists, and two things are true: (i) the first elements of the two lists are related by a function application, and (ii) the remainder of the lists themselves satisfy the map predicate. All other cases of "map" are, by default, considered false. Then, when we affirm that "area(2, R)" is true, the language has an undefined variable R, but it can tell us that the only way in which this predicate can be true is if that "R" variable equal 12.56. Similarly for the list: the only way in which "map(area, [0, 1, 2], R)" can be true is if R is a list, and recursively the list can be computed.

This paradigm is much less common in many domains than the imperative one, but it is used in control system, knowledge representation in databases, verification and optimization tasks. The reason I mention this specific paradigm is that it allows for very different representation of certain objects than the other paradigms. For example, a circle as represented in an imperative language such as logo will eventually be a procedure for how to draw a circle: an important property of that definition is that the center of the circle, for example, is left completely

out of the equation. On the other hand, prolog will allow us to define predicates that define the circle as a function of its property of equidistance from the center.

Prolog is, in fact, well suited for turning straightforwardly mathematical definitions into predicates, and as such has proven useful in working with set of geometric objects defined by constraints (Brüderlin, 1985; Franklin et al., 1986). It is not obvious, however, how to integrate the strength of this language into the framework devised in **chapter 5**, and this point is discussed again at the end of that chapter.

Probabilistic Programming Languages

Probabilistic Programming Languages (PPL) are a relatively recent addition to the programming languages paradigms: the underlying framework comes from (Solomonoff, 1964), but effective implementation started with (Koller et al., 1997). The main idea in PPLs is that probabilistic sampling is a first-class citizen of the language. This makes execution of programs non-deterministic, as two executions of the same program may yield different sampling, and in turn different output.

The success of PPLs comes from the advances in the ability to perform backward inference to answer questions like the following: "given an observation, and a certain number of programs, which program is most likely to have generated this observation?". This view has been particularly useful when conjoined with probabilistic models of cognition (Chater et al., 2006; Goodman et al., 2012; Tenenbaum et al., 2011) and the idea of the Bayesian brain: it gives a formal framework for performing complex Bayesian inference over compositionally structured representations, and has application that range from inferring 3D objects from 2D images, inferring internal parameters of complex models from their output alone, or figuring out what has happened for a certain physical system to get to where it is.

For example, assuming we know that a certain linear relation exists between two variables which we measure with a fixed amount of noise, but that every once in a while, the measurements yield a true outlier, independent from the ground truth. There are statistical methods to perform linear regressions that are robust to outliers, but PPL can also be used: we can model the situation with a small program (e.g. at its core, "`f(x) = if random() < threshold then uniform(min, max)`

```
else a*x + b + gaussian(mu, sd)")
```
with many underspecified variables (in our example, threshold, min, max, a, b, mu and sd). Under some assumptions about those variables, we can then let modern reverse inference mechanisms figure out which set of parameters maximizes the likelihood of getting the output. This example is inspired from (Cusumano-Towner et al., 2019); see (Ciccione et al., 2022; Ciccione & Dehaene, 2021) for examples of similar ideas applied to humans' graphicacy abilities.

In fact, in **chapter 5** I introduce the work performed with DreamCoder to perform program induction on the space of geometric programs. In that work, we use exact match do determine whether a program is an acceptable candidate for a shape, mainly for computational reasons. But we could have made the execution of a program non-deterministic, and used repeated execution to approximate an output probability function, and then replace exact match with likelihood under the assumption of a given program: this line of research has been successful for hand-drawn graphs in (Ellis et al., 2018).

Structure of the chapters

In chapter 1, I show that even the detection of an intruder among quadrilaterals distinguishes humans from non-human primates. Leveraging an intruder task with quadrilateral shapes of different regularity, I show that humans of diverse education, age and culture share a sense of geometric complexity: some shapes systematically make trials easier than others. But baboons lack this sense even after adequate extensive training: over a sequence of increasingly hard tasks, we can confirm that baboons can understand the intruder task and generalize across stimuli, but fail to generalize to quadrilateral shapes. With extensive training on the quadrilateral shapes, their general performance can rise to the level of 5-year-old children, but display no evidence of the geometric regularity effect detected in humans.

Artificial neural networks of object recognition fit all baboons' data well, but explaining humans' behavior requires using additional symbolic properties such as the presence of right angles. Symbolic models of the humans' behavior generalize to several related tasks, which validates the robustness of the choice of geometric primitives included in the model. This sharp dissociation suggests that two strategies are

available to encode geometric shapes: both humans and non-human primates share a perceptual strategy, well captured by models of the ventral visual pathway, but only human have access to an exact, symbolic strategy.

In chapter 2, I use neuroimaging techniques to shed light on the neural implementation of the two strategies put forward in **chapter 1**. First, I use a new visual search task to measure the confusion matrix across the quadrilaterals designed in **chapter 1**, and model it with the symbolic model, but not the neural network. I can also show that a data-driven decomposition of the complexity matrix coincides with the symbolic model.

Then, to make a task compatible with both adults and infants, I design a purely passive oddball paradigm: participants are shown the quadrilateral shapes of **chapter 1** centered on a screen with random scaling and rotation, once per second, with the possibility of a deviant every once in a while. In adults, I can decode the brain signal associated with oddballs, and the performance of the decoder replicates the geometric regularity effect. Then, using Representational Similarity Analysis I show that the brain signal first resembles the neural network model, and then the symbolic model, indicating that the two models exist in adults despite the fact that their behavior only reflects the symbolic model. Using source reconstruction, I show that the neural network model corresponds to a bilateral occipital cluster of sources, while the symbolic model is associated with a wide cluster which includes frontal sources and sources in the dorsal pathway.

In two groups of three-month-old infants, I try to replicate this experiment while measuring the EEG activation. Preliminary results are inconclusive as to the intruder detection, the shapes themselves are represented differently from one another in a way that is compatible with the geometric regularity effect.

In chapter 3, I use 3T fMRI in adults and six-year-old children to more accurately localize the areas associated with geometric shape perception across education.

In a category localizer, I find that passive geometric perception, when contrasted to other visual objects, activates a network of areas previously associated with non-linguistic high-level mathematics. Conversely, geometric shapes are under-represented compared to

the other categories in the ventral pathway. This result holds for both age groups, indicating that the core mechanisms at play in geometric shape are already present in six-year-old children.

Then, participants performed a variation of the intruder test from **chapter 1** inside the fMRI. Both age groups performed a simple version of the task, and additionally adults performed a hard version of the task. The behavior from inside the scanner replicates our previous experiments, including an overall ranking of performance across age groups and difficulty. Data from adults in the easy condition yielded clusters with increased activity as a function of the complexity of the shape, and Representational Similarity Analysis detected significant clusters associated with the neural network model, the symbolic model, and both models.

In chapter 4, I focus on a single geometric property, right angles, and compare the behavior of educated adults and baboons in a delayed match-to-sample task involving various angles and distractors.

In trying to replicate results on the categorical perception of right angles in adults, I put forward the fact that several properties are required for right angles to perform differently than neighboring angles: (i) the stimuli need to be displayed long enough, and (ii) no other low-level property can be used to perform the task. Both results coincide with the idea that symbolic properties require attention as put forward in **chapter 1**.

I also present early data collected in baboons; however, data collected so far only correspond to a condition where humans do not display categorical perception of geometric shape. We replicate this result in baboons, but this is inconclusive about the cases in which humans perceive right angles categorically.

In chapter 5, I go beyond a small set of highly controlled quadrilaterals and I set to try to account for *all geometric* shapes produced by humans. To do so, I make a concrete proposition for a generative mental language of geometric shapes inspired by attested human geometric productions.

I fully develop the argument that perceiving a shape means finding the shortest program in this language that generates the shape. And for the language I propose, I show that program induction is in principle

a tractable problem: for that I use the DreamCoder algorithm I helped implement. Then, I use this language to generate shapes of increasingly high complexity and show that humans' performance in a match-to-sample task for a shape correlates with the length of its shortest program, above and beyond many other perceptual features.

To decouple this result from the exact language proposition, I also derive more general additive rules that any alternative languages must obey. In particular, I show that repetition and embedding are essential to capture the compositional nature of geometric shape complexity.

Chapter 1

Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity

Abstract

Among primates, humans are special in their ability to create and manipulate highly elaborate structures of language, mathematics, and music. Here we show that this sensitivity to abstract structure is already present in a much simpler domain: the visual perception of regular geometric shapes such as squares, rectangles, and parallelograms. We asked human subjects to detect an intruder shape among six quadrilaterals. Although the intruder was always defined by an identical amount of displacement of a single vertex, the results revealed a geometric regularity effect: detection was considerably easier when either the base shape or the intruder was a regular figure comprising right angles, parallelism, or symmetry rather than a more irregular shape. This effect was replicated in several tasks and in all human populations tested, including uneducated Himba adults and French kindergartners. Baboons, however, showed no such geometric regularity effect, even after extensive training. Baboon behavior was captured by convolutional neural networks (CNNs), but neither CNNs nor a variational autoencoder captured the human geometric regularity effect. However, a symbolic model, based on exact properties of Euclidean geometry, closely fitted human behavior. Our results indicate that the human propensity for symbolic abstraction permeates even elementary shape perception. They suggest a putative signature of human singularity and provide a challenge for nonsymbolic models of human shape perception.

> [The universe] is written in mathematical language, and its characters are triangles, circles and other geometric figures, without which it is impossible to humanly understand a word.

<div style="text-align: right">The Assayer, Galileo Galilei</div>

This chapter corresponds to an article published in PNAS under the following reference: Sablé-Meyer, M., Fagot, J., Caparos, S., Kerkoerle, T. van, 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

1.1 Introduction

Among primates, humans are unique in their ability to develop formal symbolic systems that capture regularities in the external world, such as the language of mathematics. A great variety of non-exclusive hypotheses have been proposed to account for human singularity, including the emergence of evolved mechanisms for social competence (Herrmann et al., 2007), pedagogy (Csibra & Gergely, 2009), natural language (Berwick & Chomsky, 2016; Hauser et al., 2002), or recursive structures across multiple domains such as language, music and mathematics (Dehaene et al., 2015; Fitch, 2014; Hauser & Watumull, 2017; Penn et al., 2008). To explore these hypotheses, experimental paradigms that afford a direct comparison of human and non-human primate behavior using the exact same methods are the most informative (Beckers et al., 2016; Ferrigno et al., 2020; Malassis et al., 2020; Smith et al., 2004; Wang et al., 2015; Yang, 2013). Here, we present a novel paradigm to investigate the differences between humans and baboons in the domain of geometry, and more specifically, the visual perception of quadrilaterals such a square, a rectangle or a parallelogram. We show that all humans, regardless of culture or education, are sensitive to the presence of geometric regularities such as right angles, parallelism or symmetry, and perform very differently from baboons in an elementary visual perception task.

Prehistorical records suggest that the appreciation of regular geometric shapes is as ancient as humanity itself. Parallel lines, circles, squares and spirals are omnipresent in human art and architecture.

The earliest engravings attributed to Homo sapiens, consisting of a triangular mesh of parallel lines, are estimated to be ~73000 years old (Henshilwood et al., 2018). Even Homo Erectus already drew abstract patterns ~540,000 years ago (J. C. Joordens et al., 2015). Paleoanthropologists do not question the human origins of such drawings because, when given the opportunity to draw, other non-human primates never produce structured figures (Saito et al., 2014). By contrast, the diversity and abstraction of young children's drawings are striking (Goodenough, 1926; Long et al., 2019). Prior research has established that even kindergartners and adults with no formal education from the Amazon already possess sophisticated intuitions for geometry (Dehaene et al., 2006; Izard et al., 2011) forming an intuitive mathematical "language of thought" (Amalric et al., 2017). Those prior results suggest, but do not prove, that humans possess a more symbolic level of understanding of the abstract properties of geometry at the perception level than other primates. Here, our goal was to design a simple empirical test capable of probing this hypothesis.

We reasoned that if humans are spontaneously attuned to the major properties of Euclidean geometry (lines, length, parallelism, perpendicularity, symmetry) and their combinations, then they might exhibit a geometric regularity effect, with a better and faster perception of regular shapes, such as a square, than of irregular ones. This hypothesis is in line with a long tradition in the psychology of perception, pioneered by Wundt, Tichener, then the Gestaltists (Gombrich, 1994), Leeuwenberg's visual grammar (Boselie & Leeuwenberg, 1986; E. L. Leeuwenberg, 1971) and Leyton's generative theory of shape (Leyton, 2003), which posits that the shapes that elicit the most compact internal representations also tend to be judged as most regular or elegant. Several previous experiments, both within and outside the domain of geometry, have shown that whenever regularities are present, humans use them to compress information in working memory and achieve a smaller "minimum description length", thus facilitating memorization, anticipation and outlier detection (Amalric et al., 2017; Feldman, 2000; Mathy & Feldman, 2012; Planton et al., 2020; Shepard et al., 1961).

Crucially, the domain of visual shape perception is simple enough to probe the sensitivity of human and non-human animals to the same mathematical properties. Indeed, a previous study demonstrated that humans could perceive visual patterns with nested symmetries, while

pigeons did not (Westphal-Fitch et al., 2012). Here, we opted for an even simpler intruder test, where a participant must simply find the outlier shape within a set of six, and which has been previously used to explore human intuitions of geometry (Dehaene et al., 2006; Dillon et al., 2019). We used it to test a large number of humans and baboons with the very same stimuli.

1.2   Results

1.2.1   Design of the Geometric Intruder Task

We focused on four geometrical properties of polygons: the presence of parallel lines, equal sides, equal angles, and right angles. Our hypothesis was that the perceived geometric regularity of a shape would be directly related to its number of geometrical properties. On this basis, we selected 11 quadrilaterals ranging from highest regularity (a square) to full irregularity (an arbitrary quadrilateral devoid of any of these properties). The 11 shapes, ordered by predicted regularity, are depicted in **Figure 1.1**A and described in **Table 1.2**. For each such reference shape, four deviant versions were generated by changing the position of the bottom-right vertex by a constant distance, either along the bottom side or along a circle centered on the bottom-left vertex (thus violating either distance or parallelism). All deviants departed from their reference shape by the same amount, and all 11 reference shapes were matched for average distances between vertices (see Supplementary Online Materials). On each trial of the intruder task, we selected one of the 11 possible reference shapes and presented five instances of it, varying in scale and orientation (e.g. 5 rectangles), together with a single deviant (in this case, a non-rectangle with the bottom-right vertex displaced). The location of this outlier was randomized, and six levels of shape rotation and shape scale were pseudo-randomly distributed among the six shapes. The participants' task was to click on the outlier shape, as fast and accurately as possible (**Figure 1.1**B).

1.2.2   Results in Humans

Intruder Task in Educated Adults

In experiment 1, with N = 605 French adults, we observed that error rates in the intruder task varied dramatically with the reference shape,
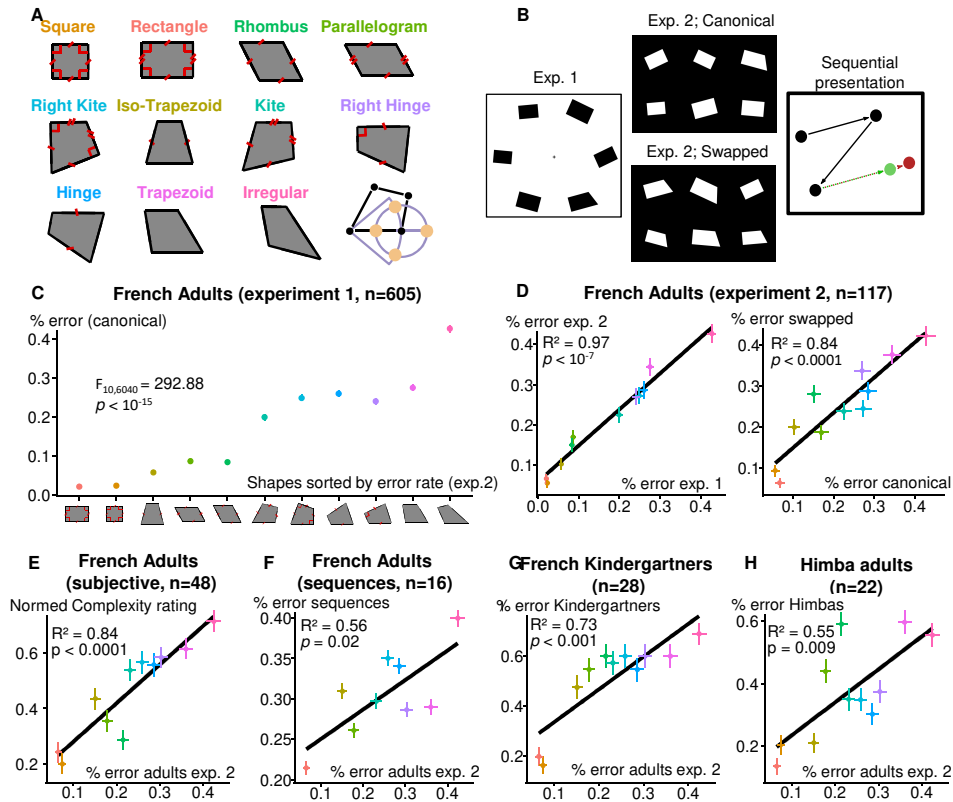
Figure 1.1: **A, stimuli.** We selected 11 quadrilaterals, here ordered according to their number of geometrical regularities (parallelism, equal sides, equal angles or right angles). For each quadrilateral, four deviants were generated by moving the bottom right corner by a fixed distance, thus shortening, lengthening or rotating the bottom side. **B, examples of intruder-task displays.** Left: circular display used in experiment 1. Participants had to tap the intruder. Center: Rectangular display used in experiment 2 and following. In the canonical presentation, five shapes exemplified a fixed quadrilateral, with variations in size and orientation, and the remaining shape was a deviant. In the swapped presentation, those two shapes were swapped. In either case, participants had to tap the intruder. Right: sequential presentation, unfolding from top to bottom and from left to right over the span of 1.8 seconds. Participants had to answer "correct" for properly placed dots (in green), and "incorrect" for deviant dots (example in red). **C, geometric regularity effect in experiment 1:** error rate varied massively with shape regularity in French adults. Shapes are ordered by performance, and each is labeled with a color which is consistent across graphs, including panel A. Error bars represent the standard error pooled over all participants – in this figure it is smaller than dot size. **D-H: Replications of the geometric regularity effect** with: **D**, swapped versus canonical trials in French adults; **E**, subjective judgments of shape complexity on a 1-100 scale; **F**, Sequential presentation of the four corners; **G**, French kindergartners; **H**, uneducated Himba adults from rural Namibia.

from 2% to 40% (**Figure 1.1**C; Univariate Type III Repeated-Measures ANOVA: $F(10, 6040) = 292.88$, $p < 10^{-15}$; explained variance evaluated by the generalized eta squared: $\eta^2_G = .27$). Average performance was well predicted by the total number of geometrical regularities (linear regression on 11 points: $r^2 = .64$, $p = .0031$) and showed a consistent, though imperfect, ordering from regular to irregular (**Figure 1.1**C). Since the regularity of symmetrical figures, such as the iso-trapezoid, was underestimated by our theoretical measure, in subsequent experiments we use the error rate from experiment 2 as an empirical measure of regularity.

By contrast to the major effect of shape, the size, rotation and position of the outlier had significant but only minor effects (size: $F(5, 3020) = 4.46$, $p = .0005$, $\eta^2_G = .005$; rotation: $F(5, 3020) = 21.19$, $p < .0001$, $\eta^2_G = .021$; position: $F(5, 3020) = 4.96$, $p = .0001$, $\eta^2_G = .005$). Response times were tightly correlated with error rates (linear regression: $r^2 = .92$, $p < .0001$) and therefore also exhibited a large geometric regularity effect (**Figure 1.2**).

In experiment 1, the intruder was always a deviant shape, and was therefore more irregular than the reference shape. Thus, participants could have responded by selecting the most irregular among the six shapes on display. To avoid this confound, in experiment 2 and all subsequent experiments, half of the displays were canonical (five instances of one of the 11 reference shapes, plus a single deviant) and half were swapped (five deviants, identical up to a rotation or scale change, plus a single reference shape; see examples in **Figure 1.1**B). As previously, participants were simply asked to click on the shape that differed from the others. In a new group of $N = 117$ French adults, the geometric regularity effect was replicated (differences between shapes: $F(10, 1160) = 70.96$, $p < 10^{-15}$, $\eta^2_G = .25$; correlation with experiment 1: $r^2 = .97$; $p < 10^{-7}$; **Figure 1.1**D), while size, position and rotation effects again had either insignificant or very small effects (size: $F(5, 580) = 2.16$, $p = .056$, $\eta^2_G = .008$; rotation: $F(5, 580) = 9.66$, $p < .0001$, $\eta^2_G = .031$; position: $F(5, 580) = 2.26$, $p = .047$, $\eta^2_G = .008$). Response times also yielded a large geometric regularity effect (correlation with error rate: $r^2 = .95$, $p < .00001$). Error rates were strongly correlated across the two display types ($r^2 = .84$; $p < .0001$; **Figure 1.1**D).
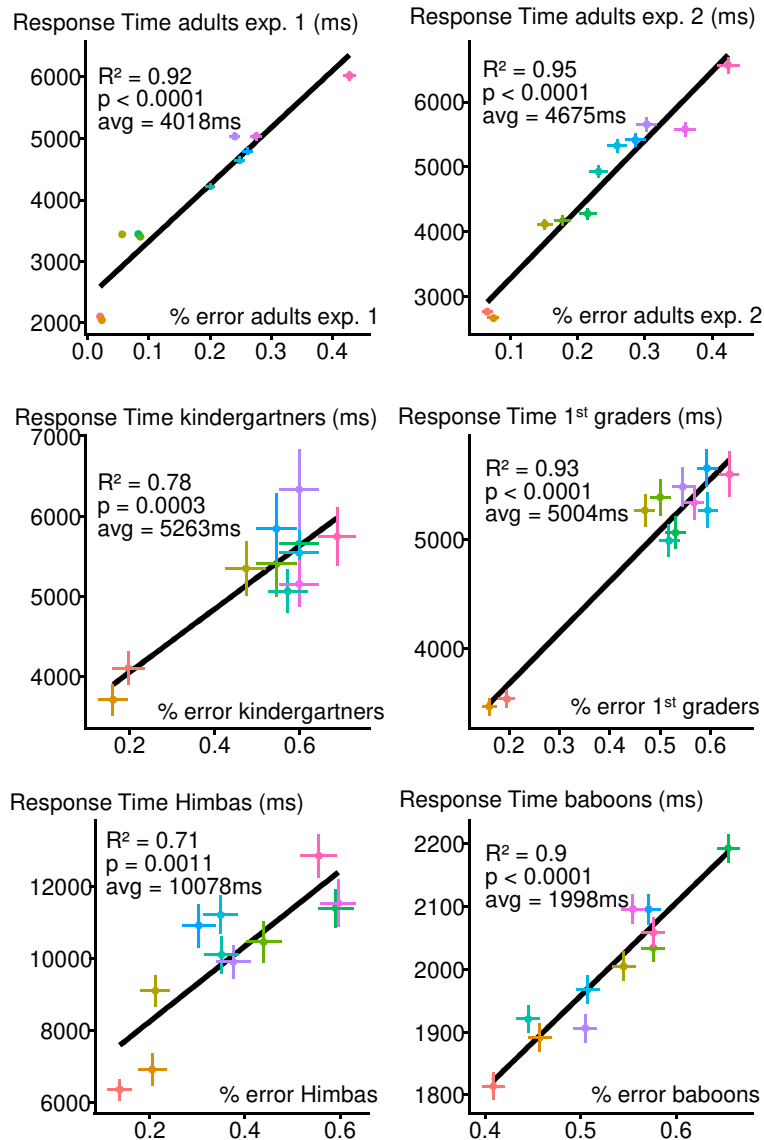
Figure 1.2: Correlation between averaged participants' error rate (x axis) and response times in milliseconds (y axis) across all 11 shapes for each test group. From left to right, from top to bottom: French Adults exp. 1, then exp. 2, then kindergartners and 1st graders, then Himbas, and finally baboons.

Subjective Ratings of Complexity

Three additional experiments investigated the origins of the geometric regularity effect. First, we asked whether geometric regularity was consciously accessible and could therefore be directly reported using subjective ratings. N = 27 French adults rated the subjective complexity and N = 21 rated the subjective regularity of each reference shape on a 1-100 scale. Both subjective ratings correlated tightly with error rates in the intruder task (complexity r^2=.88 and regularity r^2=.76; r^2=.84 after aggregating the two conditions by averaging complexity and 1 – regularity; all p<.0001; **Figure 1.1**E). Since what characterizes complex stimuli at the early visual stages of object recognition is largely thought to be inaccessible to introspection (Pylyshyn, 1999), the finding that humans have correct intuitions that some geometric shapes are simpler than others suggesting that this effect arises at a level of representation beyond early vision.

Visual Search

We further tested this idea by probing whether the search for geometric regularity engages parallel ("pop-out") or serial processes. N = 11 French adults engaged in a classic task of visual search for an outlier in arrays of 6, 12 or 24 shapes. Response times showed that search was always serial, for all 11 shapes, yet with a slope and an error rate that again correlated linearly with geometric regularity (p<.0001, r^2=0.98; **Figure 1.3**; detailed analysis of the effects of number of items and item presence provided in Supplementary Materials). This finding shows that the regularity effect does not arise from an early pre-attentive pop-out, even for the simplest shapes such as square or rectangle. Rather, geometric shape perception involves an attention-dependent stage whose speed increases with geometric regularity.

Sequential Presentation of Shapes

As a further test of the perceptual stage at which the geometric regularity effect arises, we asked whether this effect would still be present if the shapes could not be perceived in one glance, but had to be mentally reconstructed for a sequential display of their vertices. N = 16 French adults participated in an experiment in which the shapes were broken down into a sequence of four dots, one for each vertex location, in a systematic order. By having the sequence unfold over a time span of 1.8

A. Examples of visual search displays

6 shapes

24 shapes

Absent   Present   Absent   Present

B. Search performance as a function of shape type and number

% error in visual search

0.4

0.2

0.0

— 24 shapes
-- 12 shapes
— 6 shapes

% error in visual search

0.4
0.3
0.2
0.1

R² = 0.98
p < 0.0001

0.1 0.2 0.3 0.4
% error exp. 2

C. Response times and slopes of the visual search

Average response time (ms)

6000
4000
2000

24 shapes
12 shapes
6 shapes

Absent
Present

RT slope
(ms/item)

150

100

50

R² = 0.7
p = 0.0014

0.1 0.2 0.3 0.4
% error exp. 2

Figure 1.3: **Visual search paradigm. A, Examples of visual search displays.** In the visual search task, 6, 12 or 24 shapes were randomly positioned inside a circle, and participant had to decide whether all the shapes were identical, irrespective of rotation and scaling, or whether there was one that differed from the others. They gave their binary present/absent response by pressing one of two possible keys on the keyboard. **B, Error rates in visual search task.** Errors rates increased with both the number of shapes and their complexity (geometric irregularity). The latter effect correlated tightly with the average error rate in the intruder task. **C, Search times. Left:** Slope of the visual search as a function of the number of displayed items, the presence or absence of an outlier, and the shape. **Right:** Correlation between the slope of the visual search on present trials and the error rates of the intruder task (exp. 2).

s, thus largely exceeding the time window for integration within the ventral visual recognition system (Forget et al., 2009; Greene, 2016), our goal was to prevent classical bottom-up shape recognition mechanisms, yet still allow subjects to grasp the geometric relationships between the 4 vertices. The experiment was run in small blocks, each with reference shapes. In the first six trials of a given block, the four dots always traced a fixed quadrilateral (e.g. rectangle), with variations in size and orientation. Then, on each subsequent trial, the first 3 dots continued to trace the same quadrilateral (again with variations in size and orientation), but on half of the trials the fourth dot was displaced to one of the four possible deviants shown in **Figure 1.1**A. Participants were asked to indicate if the last dot was correctly or incorrectly located. Even under this sequential condition, the geometric regularity effect was replicated: the error rate still varied dramatically across shapes ($F(8, 120) = 10.1$, $p < 10^{-9}$, $\eta^2_G = .16$) and the effect correlated with the geometric regularity effect observed for static shapes ($r^2 = .56$; $p = .02$; **Figure 1.1**F). Thus, the effect arises from a level of representation where geometric properties can be ascertained even when they are not simultaneously present in the stimulus.

Probing the Influence of Education: Himbas and Young Children

We next investigated the dependence of the effect on age, education and culture. One possibility is that the effect arises from formal education in mathematics, for instance because regular shapes are also familiar, nameable, and taught at school. To address this concern, we turned to human populations with little or no formal schooling. First, we tested French kindergartners (N=28; mean age 5 years 4 months; range 4:11 to 5:10. To shorten the duration of the experiment, children were tested solely with canonical displays. N=156 1st graders were also tested, see supplementary materials and SI Appendix **Figure 1.4** for detailed results). Second, even since those Western children could have been introduced with shape names, we also tested 22 uneducated Himba adults, a pastoral people of northern Namibia whose language contains no words for geometric shapes, receive little or no formal education, and who, unlike French subjects, do not live in a carpentered world (Davidoff et al., 2002).

In both populations, the geometric regularity effect was replicated (**Figure 1.1**G and **Figure 1.1**H; see **Figure 1.5** for a systematic investigation of the significance and effect size of each predictor on
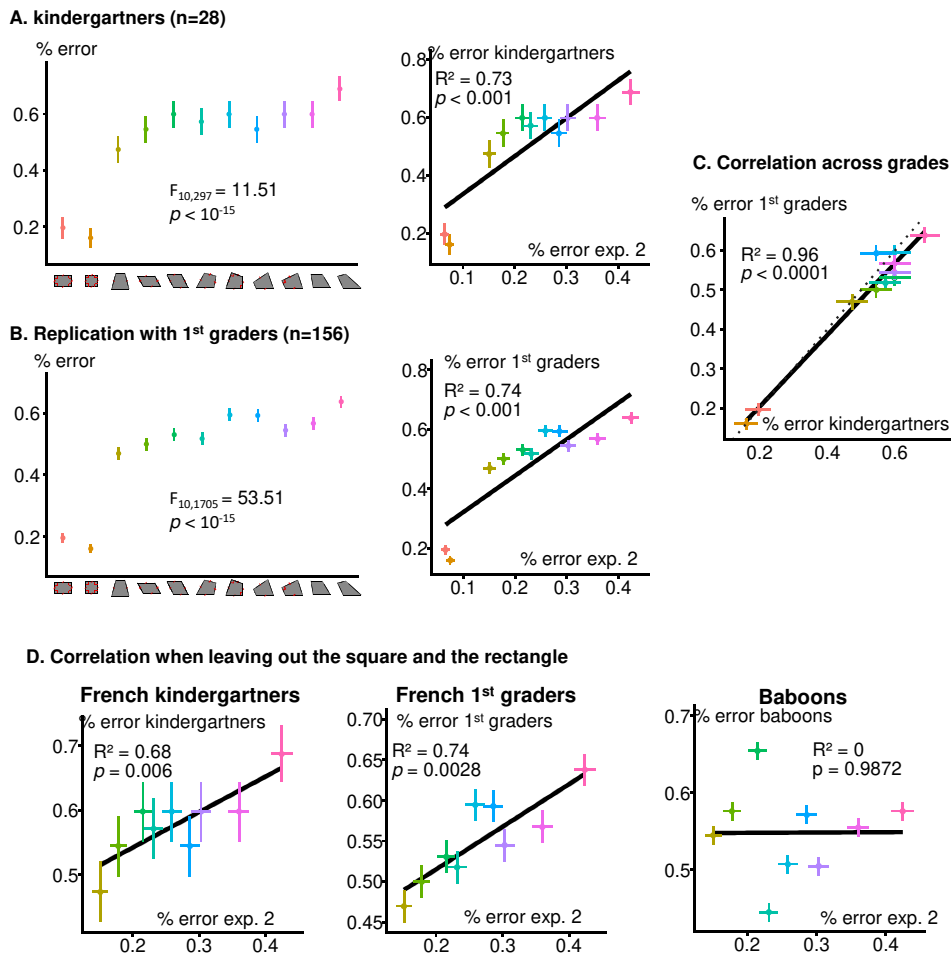
Figure 1.4: **A, Kindergartner study and comparison with adults. Left:** Main effect of quadrilaterals on performance in the intruder task. **Right:** Correlation between French kindergartners and French adults. Colors match the left plot and indicate the shape. **B, 1st graders study and comparison with adults**. **C, Comparison between kindergartners and 1st graders.** The dotted line indicates a slope of 1 while the solid line indicates the best fit (slope = .91, SE = .06). **D, Geometric regularity effects after exclusion of square and rectangle.** Although the data from kindergartners and 1st graders suggested that the square and rectangle shapes were outliers, their performance continued to exhibit a geometric regularity effect and remained correlated with that of French adults even when square and rectangle shapes were excluded from the analysis. In baboons, by contrast, the correlation remained nonsignificant.

each population). In kindergartners, errors rates varied even more dramatically than in educated French adults across the 11 shapes. They remained below 20% for the square and rectangle, ~50% for the iso-trapezoid, and continued to climb up to 60-70% for more irregular quadrilaterals. The correlation of children and French adult performance was strong and remained significant even when excluding the two simplest shapes (square and rectangle; see SI Appendix **Figure 1.4**D). Similarly, the performance of Himba adults varied with geometrical regularity and was correlated with that of both French adults (r^2=0.55) and French kindergartners (r^2=0.59). Both findings converge with previous work (Amalric et al., 2017; Dehaene et al., 2006) to suggest that the geometric regularity effect reflects a universal intuition of geometry present in all humans and largely independent of formal knowledge, language, schooling, and environment.

	Shape			Outlier Pos			Outlier Type			Outlier Scale			Outlier Rotation			Symbolic Model	
	F , p , η²G	p	η²G	F	p	η²G	F	p	η²G	F	p	η²G	F	p	η²G	p	r²
French Adults 1	F(10, 6040) = 292.88	<0.01	0.3	F(5, 3020) = 4.96	<0.01	<0.01	F(3, 1812) = 114.09	<0.01	0.1	F(5, 3020) = 4.46	<0.01	<0.01	F(5, 3020) = 21.19	<0.01	0	<0.01	0.537
French Adults 2	F(10, 1160) = 70.96	<0.01	0.3	F(5, 580) = 2.26	0.05	<0.01	F(3, 348) = 53.60	<0.01	0.1	F(5, 580) = 2.16	0.06	<0.01	F(5, 580) = 9.66	<0.01	0	<0.01	0.591
Himbas	F(10, 210) = 19.61	<0.01	0.4	F(5, 105) = 0.32	0.9	<0.01	F(3, 63) = 10.99	<0.01	0.1	F(5, 105) = 2.07	0.07	0.04	F(5, 105) = 1.77	0.13	0	<0.01	0.351
Preschoolers	F(10, 270) = 14.90	<0.01	0.3	F(5, 135) = 1.92	0.1	0.04	F(3, 81) = 12.03	<0.01	0.2	F(5, 130) = 2.47	0.04	0.05	F(5, 135) = 0.75	0.59	0	<0.01	0.463
1st graders	F(10, 1550) = 76.93	<0.01	0.2	F(5, 775) = 3.51	<0.01	0.01	F(3, 465) = 53.38	<0.01	0.1	F(5, 775) = 9.60	<0.01	0.03	F(5, 775) = 8.38	<0.01	0	<0.01	0.514
baboons	F(10, 100) = 24.68	<0.01	0.4	F(5, 50) = 3.50	<0.01	0.08	F(3, 30) = 102.97	<0.01	0.6	F(5, 50) = 2.98	0.02	0.05	F(5, 50) = 44.82	<0.01	0.4	0.12	0.0568

Figure 1.5: For each tested population, we ran five separate ANOVAs to measure the significance and effect size on performance of five different aspects of the stimuli: geometric shape (11 shapes), position of the outlier (6 positions), type of outlier (4 types of deviants, as defined in **Figure 1.1**), scale of the outlier (6 scale changes), and rotation of the outlier (6 angles). The table reports, for each ANOVA, the p-value and generalized eta-squared value (proportion of variance accounted for). On all human populations, there was a main effect of the shape (i.e. the geometric regularity effect), and a significant but smaller effect of outlier type. Other predictors were either not significant or had extremely small effect size. By contrast, three variables impact baboons' behavior: the shape, the type of outlier, and the rotation of the outlier. The shape effect (different from the human geometric regularity effect) is described in the main text. As for the outlier type and rotation effects, baboons fared better on trials where the deviants were smaller due to an inward displacement of the bottom right vertex, and fared better when the outlier was maximally rotated in one direction or the other.

1.2.3 Can Baboon Pass the Intruder Test?

Next, we investigated whether the effect was also present in a non-human primate species, the guinea baboon (*Papio papio*). Baboons' visual system is largely similar to that of humans, and they perform similarly in some shape recognition tasks (e.g. ref *25*). We capitalized on a large facility where baboons can freely access testing booths with touch screens (Fagot & Bonté, 2010). Twenty-six baboons received individualized training on the intruder task, using a great variety of images and textures (**Figure 1.6**). Complete detail of each subject's learning history and performance is provided in **Figure 1.7** and in Supplementary Materials. A full data set was obtained from 20 animals
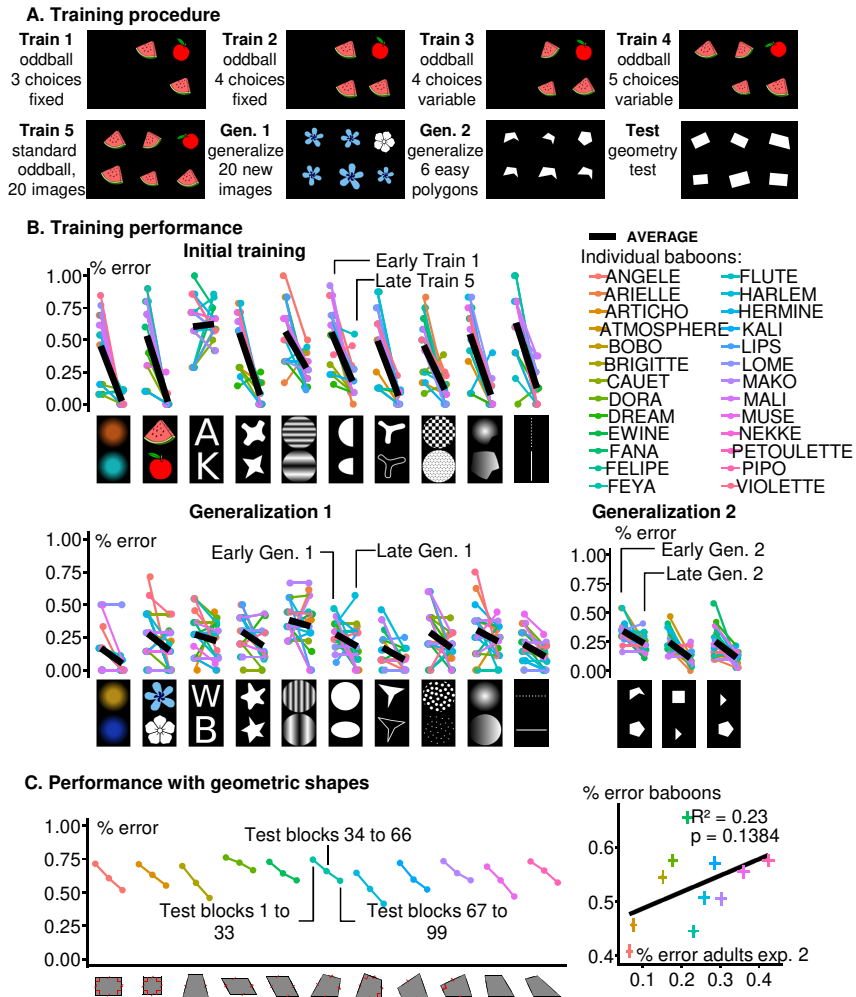
Figure 1.6: **The geometric regularity effect is absent in baboons.** **A**, training procedure. Each animal was trained for thousands of trials on the intruder task, first with a small number of fixed images (n=3, training stage 1), then with a larger number of images (up to 6, training stage 5) and with variations in size and orientation. Mastery of the task was verified through two generalization tests using novel images. Each baboon moved from one stage to the next only when the error rate fell below 20%. **B,** Summary of baboon training performance (first and last blocks of 88 trials each). Each color represents one baboon. Most animals attained criterion on the 10 pairs of shapes used for training (top) and successfully generalized to 10 new pairs of shapes (bottom left) and to 3 pairs of easily distinguishable polygons (bottom right; chance = 83.3% errors with 6 shapes). **C**, performance in the geometric intruder task. Left: average performance for each geometric shape at three stages: the first 33 test blocks, the middle 33 test blocks, and the last 33 test blocks. Each block contained 88 trials, and baboons took at most 99 blocks. Right: correlation between the average error rate in baboons and in French adults taking the same test (experiment 2).
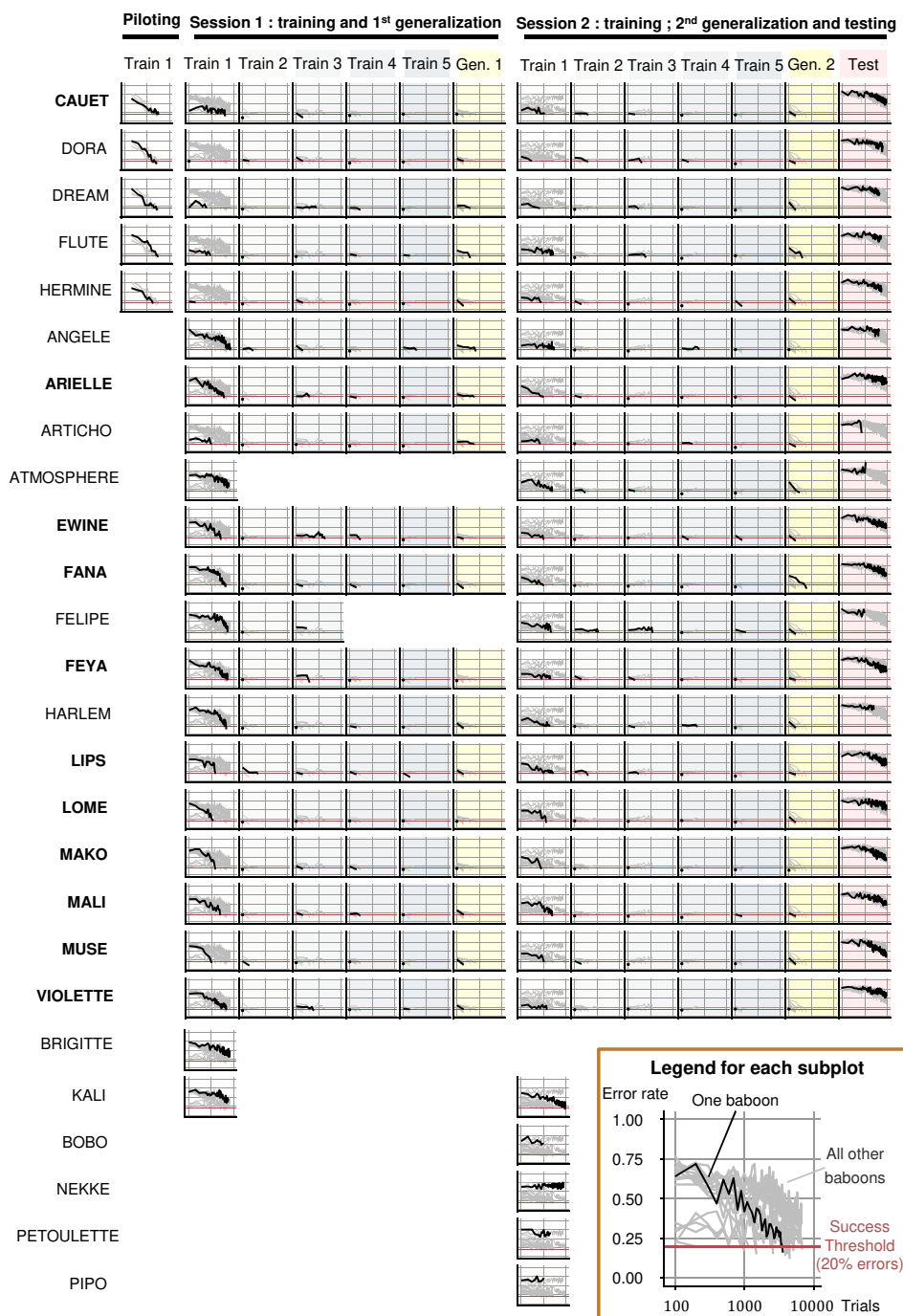
Figure 1.7: Each graph shows the average error rate as a function of the number of trials that the animal took, split according to the different phases of the training and testing (as defined in **Figure 1.6**). Each line corresponds to a baboon: the first 20 lines show all animals that produced data in the final test of geometric figures, and the last 6 rows show all animals that dropped at various stages of training. The x-axis is a logarithmic axis (Log10), so that generalization blocks (which typically contain far fewer trials) can be seen. When a plot is missing, it means that the baboon did not take that particular block. Baboons with names in bold pursued the task until after block 81 and were therefore included in the main analyses.

who completed (1) an initial series of training stages on the intruder task with 10 non-geometric image pairs, progressively increasing in the number of available choices (**Figure 1.6**A; 20 animals reached criterion; average of 5200 trials to criterion, range 1000 – 14500); (2) a first generalization to 10 novel non-geometric image pairs, indicating that they understood the intruder task (only tested in 18 animals; average = 272 trials, range 100 – 700); (3) a second generalization to black-and-white geometric shapes, where a simple non-geometric parameter sufficed to respond (e.g. pick a small triangle amidst large pentagons; average = 220 trials, range 100 – 600); and finally (4) generalization and further retraining with the complete set of quadrilaterals identical to human participants (average 6305 trials, range 704 – 8712).
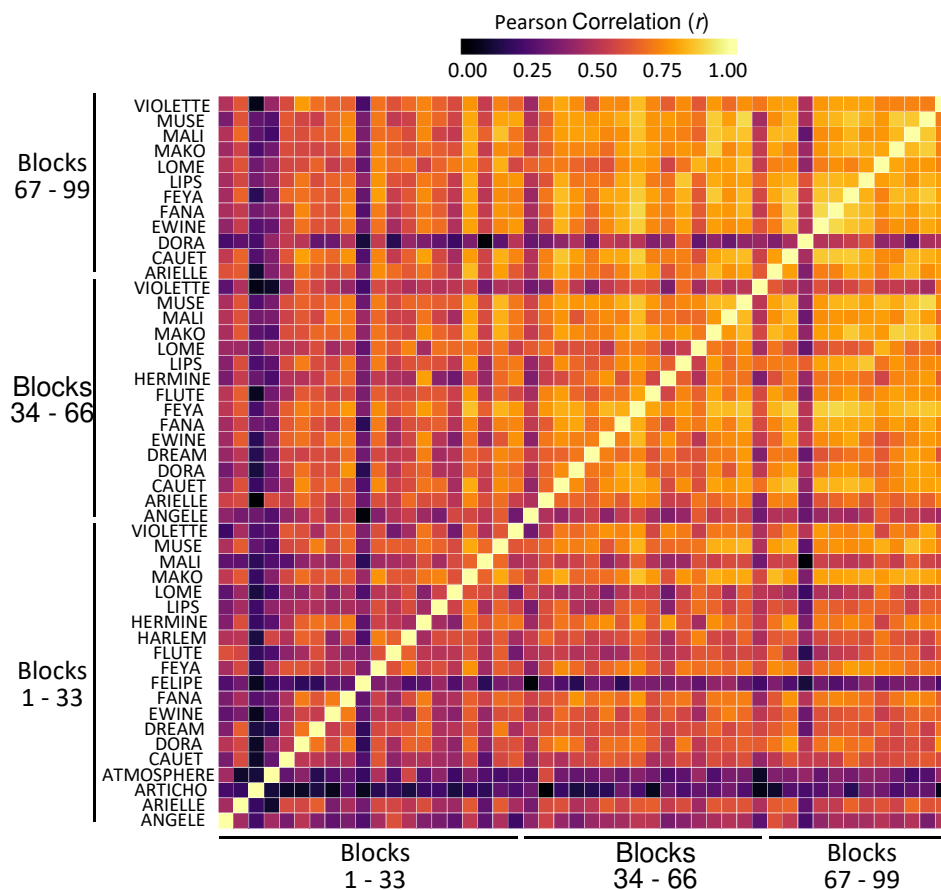
Figure 1.8: Cross-correlation matrix of the performance of each individual baboon over the course of testing, across 44 data points (11 shapes X 4 deviant types).

Twenty of the 26 animals showed a clear understanding of the in-

truder task, because following training with 20 non-geometric images, they showed immediate first-trial generalization to new such images and/or to easily distinguishable polygons (**Figure 1.6**B). However, when presented with the 11 quadrilaterals, baboons' performance collapsed, suggesting that they found all of them equally similar (**Figure 1.6**C). Their performance was close to chance on the first test block (76.2% errors, SE=1%; chance = 83.3%) and slowly progressed on subsequent days. 11 animals continued performing the geometrical task for 8000 trials or more, eventually reaching 53% errors (significant deviation [SD]=6.7%) on blocks 81 to 99. Note that this performance was comparable to that of the kindergartners and 1st graders, who achieved respectively 51% (SD=14%) and 48% errors (SD=16%). Yet even in the latter blocks, for the 11 primates who reached that stage and had therefore received substantial training, no geometric regularity effect was observed. Although error rates differed across the 11 shapes (F(10, 100) = 24.68, p<10^{-14}, .0001, η^2_G=.44), with a consistent ordering across baboons (**Figure 1.8**) and a tight correlation with their RT (See SI Appendix, **Figure 1.2**), they correlated weakly and non-significantly with the geometric regularity effect found in human populations (**Figure 1.6**C). Rather, baboon performance was impacted, at least in part, by visual properties that had little to no impact on human participants, such as outlier rotation and outlier type (see **Figure 1.5**). Thus, baboons performed poorly with quadrilaterals and were insensitive to their geometric regularities.

1.2.4   Models of Human and Baboon Performance.

To shed light on the dissociated performance of humans and baboons, we contrasted two classes of models of the intruder task (**Figure 1.9**). The first class assumes that quadrilaterals are processed by standard image recognition mechanisms in the ventral visual pathway, while the second assumes an additional level of discrete, symbolic processing of non-accidental geometrical properties.

We modeled the ventral visual pathway using CORnet, one of the top-scoring convolutional neural networks (CNN) on brain-score.org, a platform that compares computational models with behavioral and neural observations (Schrimpf et al., 2018) (other CNNs gave identical results; see Supplementary Online Materials). This model was pre-trained to label photographs on ImageNet, a large set of
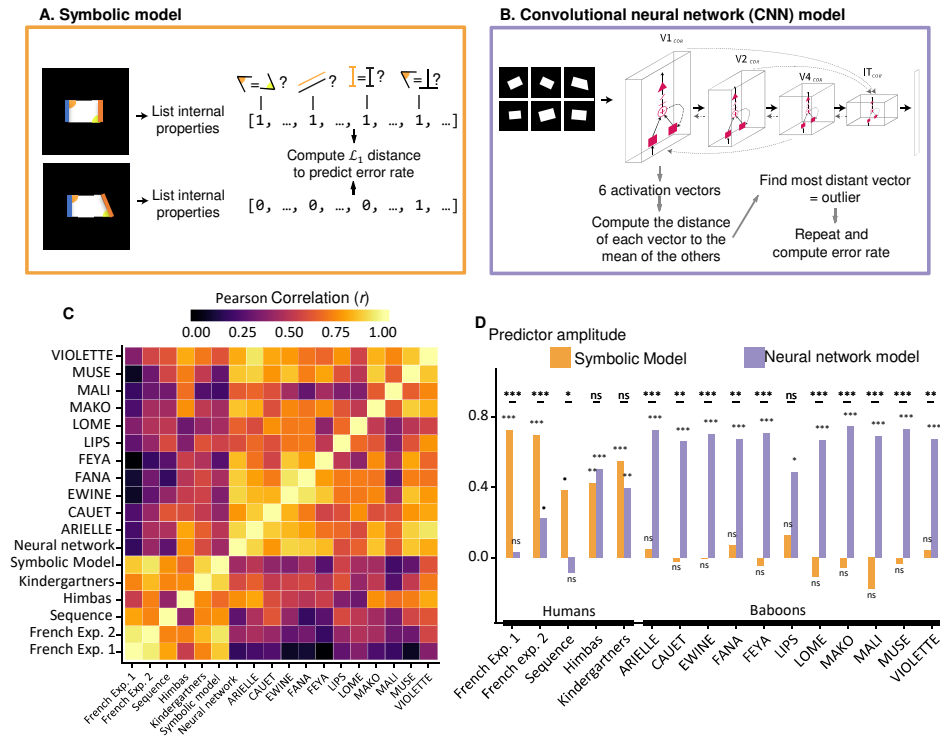
Figure 1.9: **A double dissociation in geometric shape perception. A**, symbolic model. Each shape is coded by a vector of discrete geometric properties (equal angles, parallel sides, equal lengths and right angles; each relationship is assumed to be detected with a tolerance of 12.5%). The distance between the standard and outlier vectors is then used as a predictor of the ease of intruder detection. **B**, neural network model (panel modified from ref. (Kubilius et al., 2019), with permission from the authors). CORnet, a model of the ventral visual pathway for image recognition, is used to encode each of the six shapes of a given trial by an activation vector in inferotemporal cortex (IT). The shape whose vector is the most distant ($\mathcal{L}_2$-norm) from the average of the five others, is taken as the network's intruder response. Predicted error rate is obtained by averaging across hundreds of trials. **C,** Simple correlation matrix across shapes between the performance of individual baboons (names in capitals, top rows), the predictions of the two models (middle rows), and various human groups (bottom rows). Color indicates the correlation coefficient *r*. **D,** Standardized regression weights (beta) in a multiple regression of the data from various human and non-human primate groups across 44 data points (11 shapes X 4 outlier types) using the symbolic and neural-network models as predictors. Stars indicate significance level (•, p<.05; *, p<.01; **, p<.001; ***, p<.0001).

images featuring natural and man-made items. To determine if this model could successfully simulate the outlier task, we fed the network, without retraining, with each of the six images actually presented to the participants on a given trial, collected the corresponding activation vectors in each CNN layer, and defined as the intruder the image whose vector differed most from the mean of the others. When averaging across trials, this process yielded a predicted error rate for each shape, separately for each layer in the model.

r² of the best fit
on French adults
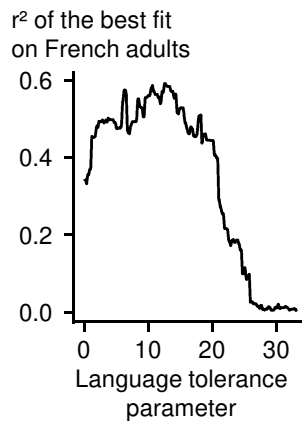
Figure 1.10: Correlation (R²) between the behavioral data of Experiment 2 (with French adults) and the predictions of the symbolic model, as a function of the tolerance threshold for accepting two sides or two angles as approximately equal. Any tolerance threshold between ~3% and ~20% yielded roughly similar fit, indicating that the model is robust to the exact choice of its only free parameter.

A second class of model, capitalizing on the prior demonstration of categorical perception for parallels and perpendicularity (Dillon et al., 2019), assumes that quadrilaterals are mentally encoded as a symbolic list of discrete geometric properties. For each shape, the model loops over all pairs of sides and angles and generates a vector of 0's and 1's for the presence or absence of equal angles, equal sides, parallelism, and right angles (with a tolerance fitted to 12.5%, although this parameter had little impact, See SI Appendix **Figure 1.10**). The difficulty of spotting the intruder is then predicted to be inversely related to the $\mathcal{L}_1$ distance (Manhattan distance) between the symbolic vectors coding for the reference and deviant shapes.

Figure 1.9C shows the matrix of correlation, over the 11 shapes, between the errors rates for each human population, each of the 11 well-trained baboons, and the predictions of the two models. Two squares are apparent. First, all baboons are intercorrelated, and their performance is well predicted by the last layer of the CNN model, putatively

corresponding to ventral inferior temporal cortex (IT; mean across an-
imals: r = .81, SE = .03). However, the CNN model is a poor predic-
tor of human performance (mean across human groups: r = .48, SE =
.10; the two distributions are significantly different: t-test, p = .024)
and reaches significance only for Himbas and kindergartners (p = .005
and p = .048 respectively). Second, conversely, all human groups are
well predicted by the symbolic model (mean r=.84, SE=.05, see **Table
1.1** for a breakdown of the effect of each symbolic property), but that
model is a poor predictor of baboon behavior (mean r = .44, SE = .04;
the two distributions are significantly different: t-test, p < .001).

Table 1.1: To quantify the contribution of each geometric property to our symbolic
model, we ran a mixed-effect linear regression on the data from our French adult ex-
periment 2. The model predicted the error rate of participants on 11 shapes, given the
presence or absence of exact property in each shape, with participants as a random
effect. The intercept corresponds to the predicted error rate for a shape without any
regularity (44%), and each additional property significantly improves the prediction of
the performance of participants. Equal sides had the greatest impact (13% gain overall),
followed by parallelism (10%), symmetry (7%), and finally right angles (3%).

	estimate	std.error	statistic	df	p.value
Intercept	0.44	0.01	29.54	332.51	<10e-8
right-angle	-0.03	0.01	-3.33	1166	<.001
parallels	-0.1	0.01	-11.09	1166	<10e-8
symmetry	-0.07	0.01	-5.94	1166	<10e-8
equal-sides	-0.13	0.02	-8.7	1166	<10e-8

This double dissociation was confirmed by a two-parameter multiple
regression where the predictions of the two models were put in com-
petition to predict 44 data points (11 shapes x 4 deviants) per popula-
tion (**Figure 1.9**D). The three experiments with French adults who re-
ceived formal education were almost exclusively captured by the sym-
bolic regressor, and each baboon's data by the neural-network regres-
sor. Interestingly, uneducated populations (Himba adults and French
kindergartners) showed a significant contribution of both models.

Thus, the modelling suggests that two strategies are available to solve
the intruder task and may coexist in humans (Davidoff et al., 2002;
Rosch, 1973): an early visual capacity, shared with other non-human
primates, to recognize shapes in the ventral visual pathway and use
this code to detect a salient deviation in shape; and a higher-level uni-
versal human capacity to grasp abstract geometric properties. The for-
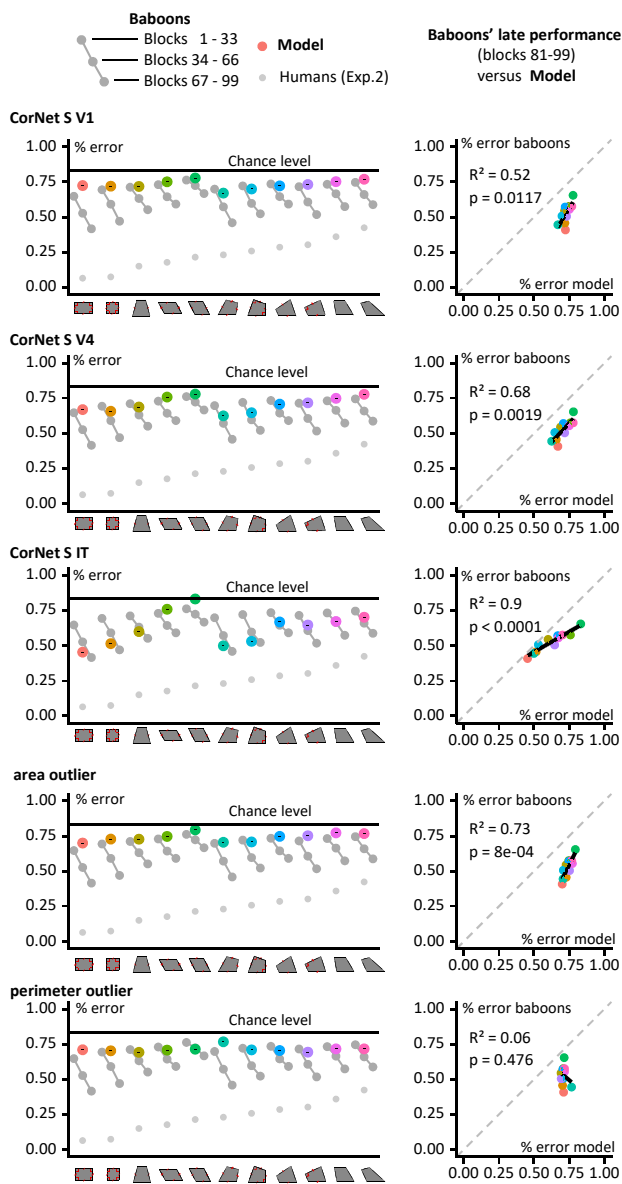
Figure 1.11: Each row displays the prediction from a given model of visual perception, with the predicted error rates across shapes (left; displayed over the data from baboons in dark gray and humans in light gray) and the correlation with the aggregate of baboons' data after the 80th trial (right). The first four rows show the prediction of each major layer of CorNet in order (V1, V4, and IT used throughout this document), followed by a model that picks the shape with area most distant from the average of the other shape's area, and an equivalent model with the perimeter. All reach significant levels at the p < .05 levels except the perimeter, and the R² increase with the layers in CorNet.

mer may exploit a variety of early and late visual cues, since further analysis of the CNN's performance showed some degree of predictability of the baboons' behavior by the V1 layer already, or by the surface area of the stimuli (See SI Appendix, **Figure 1.11**). The abstract strategy, however, appears out of reach of such simple perceptual models (indeed, without further assumptions, the neural networks would have been incapable of passing the sequence version of the task, as humans did).
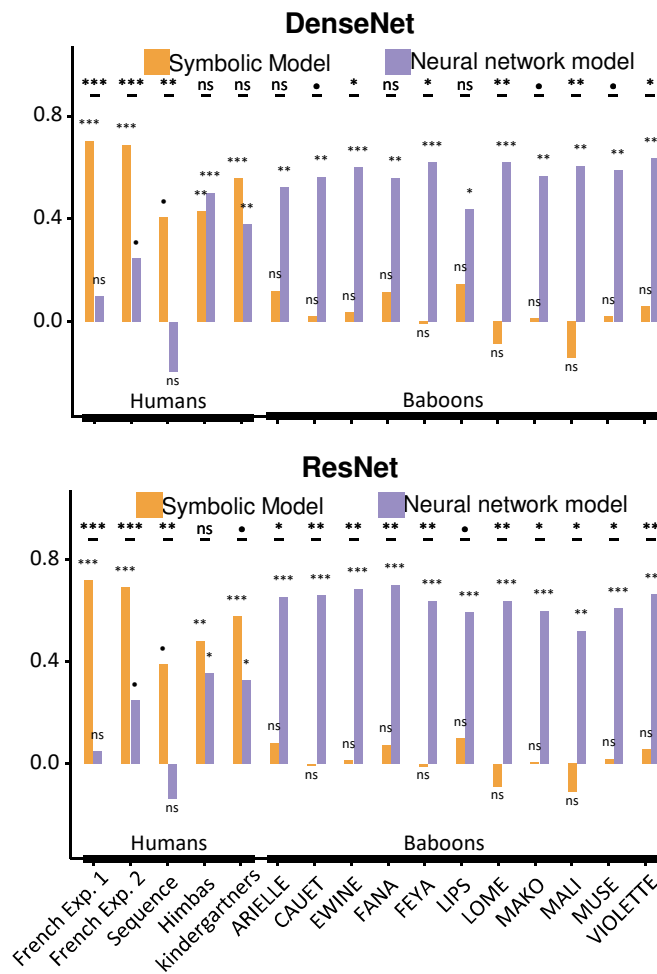
Figure 1.12: The figure shows the standardized regression weights (beta) of a multiple regression of the average performance from various human and non-human primate groups across 44 data points (11 shapes X 4 deviant types), using the symbolic and neural-network models as predictors. Stars indicate significance level (•, p<.05; *, p<.01; **, p<.001; ***, p<.0001). **Left,** using the output of the penultimate layer of densenet196 pretrained on ImageNet. **Right,** using the output of the penultimate layer of resnet101 pretrained on ImageNet.

We verified that several other similar neural networks, such as

A. Network performance across retraining with geometric shapes

B. Predictive power of the retrained networks, as a function of training scheme
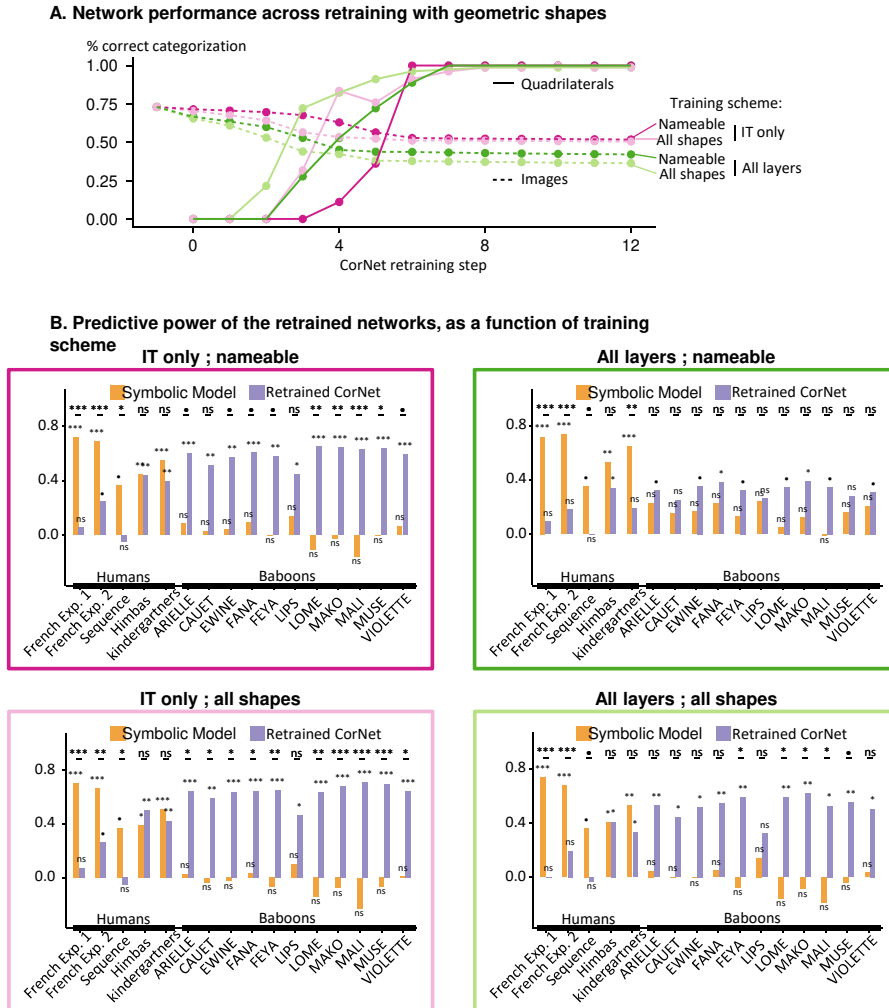
Figure 1.13: A, **Evolution of network performance across different retraining schemes.** We started with CorNet-S trained on ImageNet and retrained it by adding new output (decoder) units for geometric shapes and presenting it with only quadrilaterals for 13 epochs. For each epoch, we tested the network on new unseen views of the quadrilaterals (solid lines) and on images from ImageNet (dashed lines). We studied the effects of 4 different training schemes, defined by (1) retraining either on all 11 shapes (darker colors), or only on a subset of 5 nameable shapes (rectangle, square, rhombus, parallelogram, trapezoid; lighter colors), and (2) either freezing all layers but the penultimate one, corresponding to inferotemporal cortex IT (green), or backpropagating the error through the entire network (pink). B, **Correlation with experimentally observed performance.** Same format as **Figure 1.9** in the main text. The figure shows the standardized regression weights (beta) of a multiple regression of the average performance from various human and non-human primate groups across 44 data points (11 shapes X 4 deviant types), using the symbolic and neural-network models as predictors. Stars indicate significance level ($\bullet$, $p<.05$; *, $p<.01$; **, $p<.001$; ***, $p<.0001$). Each subplot corresponds to a specific training scheme, with color-matching panel A.
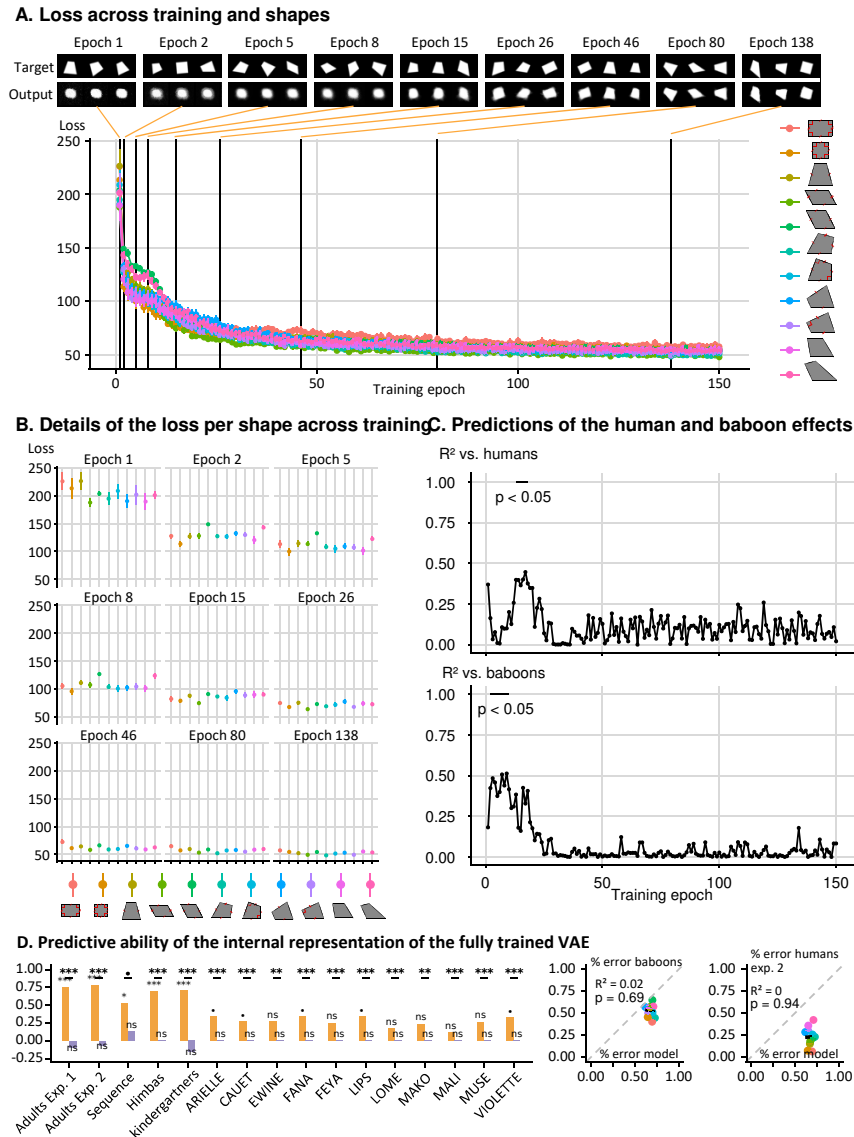
Figure 1.14: PyTorch's off the shelf VAE to produce all of our reference shapes in six possible orientations and scaling. This plot shows the network's loss on the testing dataset across training epoch for each shape – 50% of each shape was set aside for testing and the network was never trained onthese shapes. At the top of the graph, exemplars of the target shape, and the network's output, are produced, to show that the network does reproduces some fine-grained elements of the shapes, and does not just approximate a single shape that would minimize distances for all of our target shapes. **B**, Details of the loss per shape across training. At exponentially spaced epochs, detail of the loss (y axis) for each refence shape (x axis). **C**, Prediction of the human and baboon effect. 11 points pearson's R^2 of the correlation between the loss across shapes and the average error rates for humans (top) and baboons (bottom). **D**, Predictive ability of the internal representation of the fully trained VAE. Left: Standardized regression weights (beta) in a multiple regression of the data from various human and non-human primate groups across 44 data points (11 shapes X 4 outlier types) using the symbolic and VAE models as predictors. Stars indicate significance level ($\bullet$, p<.05; *, p<.01; **, p<.001; ***, p<.0001). Right: detail of the correlation with the behavior from baboons and humans (exp. 2)

DenseNet or ResNet, were similarly unable to fit human behavior (See SI Appendix, **Figure 1.12**). It could be argued that the geometric shape fell too far off the training space to elicit uninterpretable results. However, the model trained to label the ImageNet dataset did attribute to each geometry shape a highly consistent label (mostly "envelopes"; **Table 1.3**). In order to test the effect of the training space, we modified the network with extra output units and trained it to label our reference shapes (See SI Appendix, **Figure 1.13**). Four training strategies were tried, depending on whether we trained the network to label all 11 shapes or just the shapes with names in English; and whether all layers were allowed to change, or just the final layer (see Supplementary Materials). Nevertheless, all four manipulations failed to increase their predictive power of the CNN for any human population, and either worsened the predictive power for the baboon behavior, or left it unchanged. Since CNNs are far from perfect in capturing human behavior, even for natural stimuli(Baker et al., 2018; Geirhos et al., 2019; Ullman et al., 2016), we also tested Variational Auto-Encoders(Kingma & Welling, 2014) (VAE). VAE's architecture enforce the unsupervised learning of a low dimensionality representation of a set of data by jointly learning to encode and decode to/from a bottleneck layer. In that sense a VAE "compresses" information and may therefore be more suited to the task of encoding regular shapes. A classical VAE was successfully trained to encode and decode our reference shapes (**Figure 1.14**A). However, it too did not exhibit the geometric regularity effect. First, its loss function varied very little across the 11 shapes (**Figure 1.14**B). All shapes were learned similarly across training epochs, and the loss did not correlate well with either the human or the baboon behavior (**Figure 1.14**C). Second, using the same methodology as for CNNs, we probed whether the internal compressed representation of the model could be used to spot the outlier; again, it proved to be predictive of neither the humans' or the baboons' behavior (**Figure 1.14**D).

1.3 Discussion

Using the geometric intruder test, regardless of the human populations we tested, we observed a replicable geometric regularity effect: finding an intruder amongst six quadrilaterals is much easier when either the reference or the deviant shape are highly regular. This effect is already present in young children (kindergartners and 1st

graders), and was also replicated in uneducated adults from a remote non-Western population with reduced access to education, suggesting that the effect does not depend on age, culture and education. Additionally, we show that this effect is replicable using different presentation modes (by presenting the entire shape at once, or the four vertices sequentially) and different tasks (intruder, serial search, or subjective complexity rating).

Given this apparent universality in humans, it is noteworthy that the baboons did not share this effect. Their performance was initially quite poor with all quadrilaterals, but even when it later improved to the level of human children and showed significant variations across shapes, it still did not correlate with the geometrical regularity effect. This striking difference occurred even though the baboons clearly understood the demands of the intruder task, having reached a threshold of 80% correct or more on a first set of stimuli (where chance is 16.7% correct) and then generalized to new non-geometrical stimuli. It also cannot come from a lack of motivation: while a few baboons did not complete the training, the twenty on which we collected data spontaneously performed an average of 867 geometrical trials per day (1st quartile 278 trials, median 641 trials, 3rd quartile 1332 trials).

An empiricist could argue that the difference was due to the different environments in which humans and baboons live. The "carpentered world" hypothesis (Segall et al., 1963) proposes that an increased sensibility for right angles and parallel lines arises naturally from a Western style of life in a world full of rectilinear shapes (objects, buildings, books, etc.). Indeed, this was the dominant environment for most of our participants. However, several arguments refute this idea. First and foremost, the effect was present in the Himba people, but not in baboons. Yet the rural settlements of the Himba are quite unlike industrialized societies and their environment is relatively free of rectilinear objects. Conversely, the baboons we tested were not wild animals, but grew up and lived in an environment comprising a mixture of natural objects (trees and rocks) and man-made, rectilinear objects (buildings, doors, computer screen), which was arguably as "carpentered" as the Himbas' (see illustration in Supplementary Materials).

Second, even in a carpentered world, after projection in two dimensions, irregular shapes are arguably more frequent than regular ones on the retina, because the observers are rarely perfectly aligned with their environment for a rectilinear projection to occur. Parallelograms

are also rare in our environment – and yet they figured among the shapes with few errors. Thus, it is not clear how frequency in the environment would explain our result. Finally, we directly tested this empiricist hypothesis by training artificial neural networks with a dataset (ImageNet) that featured many man-made rectilinear image categories, such as envelopes, binders, band-aids or lampshades (labels which they readily applied to our quadrilaterals; see **Table 1.3**). Even more crucially, we retrained them with our geometric shapes (See SI Appendix, **Figure 1.13**). Neither types of training sufficed for the neural networks to predict human behavior.

The dissociated performance of humans and baboons suggests that the intruder task can be solved using two strategies: a perceptual strategy, well captured by current neural-network models of the ventral visual pathway, in which geometric shapes are encoded using the same feature space also used to recognize any image (e.g. faces, objects, etc); and a symbolic strategy, in which geometric shapes are encoded by their discrete non-accidental regularities such as right-angles or parallel sides. The latter strategy seems available to all humans, whether in Paris or in rural Namibia. It is tempting to speculate that it may be available *only* to humans, as suggested by the failure of all the baboons we tested. At the moment, however, this proposal remains tentative, because we only tested a limited number of humans and a single non-human primate species. Baboons also responded much faster than humans (~2s versus 5s or more, ee SI Appendix, **Figure 1.2**), possibly preventing the deployment of a more abstract strategy. Both facets of our proposal will have to be submitted to further tests, for instance by contrasting human infants, who are known to be born with sophisticated symbolic abilities (Dehaene-Lambertz & Spelke, 2015), and chimpanzees, who may lack a logical or hierarchical mode of data analysis (Penn et al., 2008).

The present results converge with prior research, using more complex geometric displays and tasks, which indicated that all humans, even young or uneducated ones, possess intuitions for geometry (Amalric et al., 2017; Dehaene et al., 2006; Izard et al., 2011) and automatically apply a symbolic, language-like formalism to geometric data (Amalric & Dehaene, 2017; Wang et al., 2019). Brain imaging showed that this "language of geometry" rests primarily on dorsal and inferior sectors of prefrontal cortex (Wang et al., 2019). These regions are activated whenever humans reason about mathematical concepts and

recombine them algebraically (Amalric & Dehaene, 2017; Maruyama et al., 2012; Monti et al., 2012). While they are located outside of classical language areas, their surface area is strikingly expanded in the human lineage (Chaplin et al., 2013; Xu et al., 2020), and they are therefore a good candidate for the emergence of novel human capacities in evolution, including symbolic mathematics. Previous work has shown that proto-mathematical core knowledge is present in other non-human primates, such as numerosity in macaque monkeys (Cantlon & Brannon, 2007; Nieder & Dehaene, 2009) or spatial navigation in baboons (Noser & Byrne, 2007). However, what these species may be lacking is a capacity to discretize those representations and recombine them in larger language-like combinatorial expressions such as "four equal sides" (Dehaene et al., 2015; Fitch, 2014; Hauser & Watumull, 2017; Penn et al., 2008), which are needed in order to conceive of a square and draw it. In the future, it would be informative to test whether chimpanzees who received "language training", i.e. learned to use visual tokens to label numbers and objects (Matsuzawa, 1985; Premack, 1988), would show the geometric regularity effect. There are reasons to doubt it, since careful analyses suggest that, unlike young children, chimpanzees do not use these tokens in productive combinations (Yang, 2013).

A parallel issue is, how could the neural networks we tested be modified to eventually pass the geometrical intruder test? Classical convolutional neural networks mimic only part of human visual recognition abilities (Ullman et al., 2016). They roughly correspond to the first, bottom-up pass of invariant visual object recognition (Kubilius et al., 2019), but much more sophisticated recurrent top-down architectures are required to attain human-level performance in slower perceptual decision making tasks (George et al., 2017; Spoerer et al., 2020). It will be interesting to examine if those newer models pass the present test or, as we tentatively suggest, if yet another level of symbolic representation, perhaps based on symbolic tree-based generative models and program inference (Balog et al., 2017; Devlin et al., 2017; Lake et al., 2015), is needed.

In summary, the present results suggest a new putative human cognitive universal: the capacity to perceive the regularity of a geometric shape such as a square. They hint at the exciting possibility that humans differ from other primates in cognitive mechanisms that are much more basic than language comprehension or theory

of mind, and involve a rapid grasp of mathematical regularities in their environment. Those findings also provide a novel challenge for artificial intelligence, as none of the classical neural network models we tested so far could capture human behavior.

1.4 Materials and Methods

1.4.1 Reference Shapes

All experiments relied on a single set of 11 fixed reference shapes, which were all quadrilaterals (**Figure 1.1**A; the coordinates of their vertices are listed in **Table 1.2**). We matched most reference shapes for two parameters. First, the average distance between all pairs of vertices (i.e., the mean of six distances) was the same across the 11 shapes. This ensured that the reference shapes had the same overall size. Second, the bottom edge was of fixed length across 9 of the 11 shapes – this was particularly important for the sequence experiment, where this segment was the last to appear on the screen and was the only one that could contain an outlier. The square and the rhombus were the only exceptions: they were only matched to other shapes on the average of distances. This was necessary because (1) the square had only one degree of freedom, and (2) the rhombus would otherwise have been either too similar to the square or utterly flat.

For some shapes (e.g. rectangle), this set of constraints led to a single choice for the specific shape. For others, we selected a shape that satisfied the constraints while being maximally different from the shapes in other categories. For instance, the specific quadrilateral that we selected for the "irregular" category made it maximally obvious that it did not have equal sides, parallel sides, equal angles or right angles. The full details required to reproduce the shapes are provided in **Table 1.2**.

Table 1.2: For reproducibility we provide here the precise coordinates of the vectors defining the three corners of each shape. With the bottom left vertex at coordinates (0,0), the first six columns define the three vectors required to locate the top-left, top-right and bottom-right vertices. When presented, the reference orientation (0°) of each shape was the one where the top edge was horizontal, around which the random orientations (-25°, -15°, -5°, 5°, 15°, 25°) took place. All bottom-right vectors, except for the square and the rhombus, are matched for length. The "Avg pairs" column gives the average distance between all pairs of points, another metric matched across shapes. The Perimeter and Area columns give respectively the perimeter and area relative to that of the rectangle: for lack of enough degrees of freedom, these properties are not matched across shapes. Neither explain the human behavior (area: p = .32, perimeter: p = .13) or the symbolic model (area: p = .28, perimeter p = .14). See additional discussion of this in the Additional Analysis section.

	topLeft x	topLeft y	topRight x	topRight y	botRight x	botRight y	Avg pairs	Perimeter	Area	Number of properties
rectangle	0	1	1.5	1	1.5	0	1.434	1	1	15
square	0	1.26	1.26	1.26	1.26	0	1.434	1.008	1.059	19
iso-trapezoid	0.365	1.362	1.109	1.362	1.5	0	1.433	1.014	1.019	5
parallelogram	-0.517	0.896	0.983	0.896	1.5	0	1.434	1.014	0.896	7
rhombus	-0.908	0.931	0.392	0.931	1.3	0	1.434	1.04	0.807	9
kite	0.766	1.29	1.77	1.007	1.5	0	1.434	1.017	1.007	5
right-kite	0.529	1.404	1.5	1.038	1.5	0	1.434	1.015	1.038	7
hinge	-0.248	0.533	0.98	1.393	1.5	0	1.434	1.015	0.986	1
right-hinge	-0.296	0.634	1.064	1.268	1.5	0	1.434	1.008	0.984	2
trapezoid	-0.227	1.2	0.724	1.2	1.5	0	1.434	1.02	0.98	1
Irregular	-0.45	1.058	0.227	1.24	1.5	0	1.434	1.025	0.885	0

The constraints that we adopted implied that the shapes were not strictly equalized in other dimensions such as surface or perimeter. Such residual differences might explain why the performance of neural networks and baboons varied slightly across shapes, but crucially they were uncorrelated with shape regularity.

1.4.2 Deviant Shapes

For each reference shape, we generated four deviant shapes by changing the position of the bottom-right vertex. All deviant vertices were equidistant from the correct vertex location. Two deviant vertices were positioned along the bottom edge, either lengthening it or shortening it (see **Figure 1.1**A). The two other deviant positions preserved the correct distance from the bottom left vertex, and thus the length of the bottom edge, but changed its orientation. The distance of the deviant position from the correct position was fixed for all experiments and was common to all shapes. It was computed as a proportion of the (fixed) average distance between all pairs of vertices (55% for the sequence experiment; 30% for all other experiments).

1.4.3 Variations in Orientation and Size

In their default presentation, the shapes were centered on their center of mass, and their top edge was horizontal. We then rotated the six shapes by a random permutation of the following angles: [-25°, -15°, -5°, 5°, 15°, 25°]. We avoided 0° rotation to prevent participants from relying on parallelism with the edges of the computer screen, and we avoided larger angles to side step the fact that some shapes had rotational symmetry (for instance, a 45° rotated square is identical to a -45° rotated square, but the same does not hold for a trapezoid). We also scaled the shapes by a random permutation of the following scaling factors applied to the edge lengths: [0.875, 0.925, 0.975, 1.025, 1.075, 1.125].

1.4.4 Participants and Experimental Procedures

Details of the participants, design, procedure, ethical committee approval and analyses specific to each experiment are presented SI Appendix. All experiments involving French subjects were approved by the ethical committee of Université Paris-Saclay. The experiment involving Himba adults was approved by the ethical committee of Goldsmiths University of London. All subjects or their legal guardians provided informed consent.

Briefly, 612 French adults were recruited for online experiment 1, 117 for online experiment 2, and 48 for on-line subjective ratings. For the sequence and visual search experiments, we tested respectively 16 and 11 participants in individual isolated testing booths. 28 French kindergartners (mean age 64 months; range 59-70 months; 15 boys, 13 girls) from two classrooms were tested individually in their school. Finally, 44 native Himba adults were recruited on-site in small individual villages of Northern Namibia (Southern Africa). All were monolingual native speakers of Otjihimba, a dialect of the Otjiherero language, which does not have vocabulary for most geometric shapes. Out of these, we report data for the 22 participants who did not attend a single year of schooling (for additional analyses of the effect of schooling, see Supplementary Materials).

Baboons (26 Papio papio, 18 females, age range 1.5-23 years, mean age 11 years) were tested at the CNRS primate facility (Rousset-sur-Arc, France). Baboons lived in a 700 m2 outdoor enclosure with access to indoor housing and could, on a voluntary basis, at any time,

69

enter ten Automated Learning Devices for Monkeys equipped with a 19-inch touch screen, a food dispenser, and a radio-frequency identification (RFID) reader that could identify the animals

1.5 Additional Methods and Details of each experiment

1.5.1 Humans

Adults, Experiment 1

Participants. 612 French adults were recruited for an online experiment (395 males, 217 females, age group breakdown: <18 years, 42 subjects; 18-25 years, 127 subjects; 25-60 years, 419 subjects; >60 years, 24 subjects). The experiment was advertised on social media using the lab's social media account. The entire experiment was run on the participant's device and took typically less than 15 minutes. Participants were not compensated for their participation. No personally identifying information was collected in this experiment. This experiment was approved by the ethical committee of Université Paris-Saclay under the reference CER-Paris-Saclay-2019-08.

Procedure & Stimuli. This experiment featured only canonical displays (5 reference shapes and 1 deviant shape). It started with two training pairs of geometric shapes, randomly selected from the 3 we used throughout the generalization 2 task for baboons. There were therefore exactly 2 + 11*4 = 46 trials. The experiment was programmed using the jsPsych framework (de Leeuw, 2015), "a JavaScript library for running behavioral experiments in a web browser." Participants first filled a consent form, then a demographic questionnaire, which collected information regarding their sex, age range, and education level. Then they were presented with the task instructions, and finally a sequence of intruder trials. On each trial, they were asked to click on the outlier, either with the mouse or with a touchscreen if their device had one. In this experiment only, the six shapes were organized in a circle as big as the screen permitted. Upon clicking on a shape, participants received visual (highlighting the selected shape in red if incorrect, in green otherwise, and highlighting the correct shape in green) and auditory feedback (rising or falling

tone). Shapes were shown in solid black on white background.

Statistical analysis: Responses slower than the overall 99th percentile were removed from analysis of this experiment, as well as experiment 2 and the Himbas experiment to match the analyses: during online experiments, some trial took unreasonable durations (e.g. over a minute), strongly suggesting participants taking a break during the experiments. In experiment 1, the 99th percentile thresholding sometimes removed all datapoints from some participants' conditions (e.g. an entire shape); in such case the participant was removed entirely: in total, 7 out of 612 recruited participants were removed.

Adults, Experiment 2

Participants. 117 French adults were recruited for an online experiment (45 males and 72 females; age group breakdown: 18-25, 9 subjects; 25-40, 43 subjects; 40-60, 56 subjects; >60, 9 subjects). The recruitment process and ethical approval were identical to that of the first experiment. Because this experiment was longer, participants were incentivized to participate by being offered to participate in a lottery for a 30€ cash prize that three participants would receive. Should they want to participate to the lottery, participants had to disclose an email address, which was collected separately from the experiment's data and could not be linked to it afterwards. 83 out of 117 participants submitted their email for participation.

Procedure and Stimuli. The procedure and stimuli were identical to that of experiment 1 with the following five differences. (1) Participants saw an additional webpage with information about the lottery. (2) Shapes were displayed in white on a black background. (3) Instead of displaying the shapes along a circle they were displayed in two lines of three items, as shown in **Figure 1.1**B. (4) Participants received 10 training trials with images and another 6 with the easy geometric training shapes, in two consecutive blocks. They had to repeat the training blocks if they performed worse than 80% correct. The training stimuli were identical to those used in the baboon experiment (see **Figure 1.6**) and included random rotation[1] and scaling. (5) Half of the displays

---

[1] Due to a bug in the code of the experiment, the training images were not scaled, though they were properly rotated. This bug only affected the training images. All geometric shapes, in training or testing, were properly scaled. This issue was present exclusively for experiments with humans.

used a standard presentation (5 reference shapes and 1 deviant), and half used a swapped presentation (5 deviant shapes and 1 reference shape), for a total of 88 experimental trials with geometric shapes.

Compared to experiment 1, the changes listed under points 2-5 were introduced in order to anticipate the changes required to replicate the baboon experiment. The displays in **Figure 1.1**B show example stimuli from this version of the experiment. This design was adopted throughout all other experiments. This experiment is available at https://neurospin-data.cea.fr/exp/mathias-sable-meyer/oddball/.

Sequence Experiment

Participants. 19 participants were tested in this experiment. It was run at ENS in Paris, in isolated testing booths. The first three participants were pilots whose results were used to tune the difficulty of the experiment. Subjects were recruited through the RISC mailing list, mean age was 23.1 years old (std = 2.55), 9 women and 11 men, with a mean of 3.44 years of post-bachelor education (std = 1.5). All participants signed an informed consent form and received 15€ for their participation. Due a schedule conflict one participant did not complete one condition ("parallelogram") of the experiment: the missing value was replaced in the ANOVA with that participant's overall average error rate, and left missing from all other analysis.

Procedure. The experiment was organized in 9 mini-blocks, each with a fixed geometric shape. In each mini-block, participants were first shown 6 examples of a given sequence (with random scaling and rotation), and were then presented with sequences that could contain a deviant. For each sequence after the sixth example, after the 4th dot was displayed, they had to press a button to indicate whether that sequence followed the reference sequence or not. Following each answer, they received auditory feedback using an ascending pitch if correct and a descending pitch otherwise and were shown the four dots location, as well as a 5th dot at the correct location for deviant trials. After 150 trials, there was a short pause, and then a new mini-block started, with 6 new examples to start with.

Stimuli. The sequences of dots traced the geometric shapes in a top-left, top-right, bottom-left, bottom-right order. Shapes were presented with a random orientation (with angles now ranging from 0 to 359°)

and random scaling so that they spanned 150 to 225 pixels on the screen, and they were positioned so that the last position would be at one of 9 possible locations on the screen. In this sequential format, we considered it essential that the last two positions were identical for all shapes. We therefore excluded the two shapes for which we could not match the bottom edge, namely the square and the rhombus. Given the greater difficulty of the task in the sequential presentation mode, we had to adjust the distance of the deviant to the correct location. Pilot participants were run in order to estimate the distance required to obtain a success rate of ~75% overall, and the deviant value used for the remaining N=16 participants was 0.55 times the matched average distance of any two points. The presentation order of the blocks was random with a single block for each shape and 150 trials within each block[2], with half of the trials being outliers. The timing of the sequence was as follows: points appeared for 400 ms followed by a 200 ms empty scree. After the participants' response, the screen stayed black for a random duration ranging from 750 ms to 1250 ms.

Subjective Rating

Participants. 48 French adults were recruited for an online experiment (21 Males and 27 females; age group breakdown: 1-18, 1 subject; 18-25, 3 subjects; 25-60, 41 subjects; >60, 3 subjects). The recruitment process and ethical approval were identical to that of the first experiment.

Stimuli. We presented the participants with our 11 quadrilaterals, in the reference orientation and presented as static images with a white shape on a black background.

Procedure. After the consent and the questionnaire, participants were instructed to give a rating for each shape one the page using a scale from 1 to 100, while trying to be as consistent in the rating as possible. Participants were randomly assigned to one of two conditions: either they were asked to give a rating of "complexity" (27 participants) or to give a rating of "regularity" (21 participants). Participants saw a page with shapes from another study not analyzed

---

[2]This was adjusted depending on participants time constraints: min 100, max 170, median 155. Two participants had to stop before the end because of time constraints, one missing one shape and the other two.

here, and then a page with our 11 reference shapes and a slider from 1 to 100 for each shape. They were asked to not transfer the scale from the previous shapes from to the 11 quadrilaterals, but instead to try and use the entire scale again and to be as consistent as possible between the shapes. We merged the data from the two conditions by reversing the scale of the "regularity" condition so that a score of 100 on "regularity" would map on to a score of 1 on "complexity" and conversely.

Visual Search Paradigm

Participants.   11 French adults were recruited (5 Females, 5 males, age range 21 - 35, mean 27.3 years, one did not complete the demographic form). Participants were not compensated for their participation. This experiment was covered by the ethical committee of Université Paris-Saclay under the reference CER-Paris-Saclay-2019-063.

Stimuli.   For each trial, repetitions of a given shape and possibly its deviant were presented in black on light gray (**Figure 1.3**A). Their rotation and scaling were uniformly sampled, similarly to previous experiments, and they were randomly placed inside a gray circle that spanned almost the entire computer screen. The experiment comprised 11 blocks, one per reference shape, each with 24 trials randomly shuffled, using a factorial design with three factors, namely, deviant type (4 possible deviants), numbers of shapes on screen (3 possibilities: 6, 12 or 24) and presence or absence of a deviant shape, for a total of 264 trials. The experiment was programmed using the jsPsych framework and was run online.

Procedure.   When connecting to the shared online URL, participants clicked to start and were prompted with instructions. For each display, they had to press the left arrow key if they thought that one of the shapes differed from the others, and the right arrow key if they thought that all shapes were identical. After pressing one of the arrow keys, the experiment started: the screen displayed a light-gray circle spanning the maximum available area with 15px padding at the top and the bottom, inside which items were placed randomly. After each response, subjects received both auditory and visual feedback, which explicitly indicated the location of the deviant shape if one was present (the deviant was colored green if answered correctly, red otherwise).

The experiment was structured in blocks of similar shapes and lasted about 20 minutes in total.

Analysis and Results. For each shape, each number of displayed item, and each target presence condition, we removed responses whose response time exceeded the mean response time plus three standard deviation. Detailed analyses of the visual search available in the supplementary materials.

Himbas

Participants. 44 native Himba adults were recruited for an experiment taking place on a tablet computer (mean age 24.5 years, minimum 14 years old and maximum 62 years old, 13 Male and 31 Females). The Himba of Northern Namibia (Southern Africa) are a population living a traditional lifestyle in rural settlements, with little exposure to Western society. All the participants were native speakers of (and monolingual in) Otjihimba, a dialect of the Otjiherero language, which does not have vocabulary for most geometric shapes (though they refer to "squares", for example, with a very direct metaphor akin to "a shape with four angles"). Out of the 44 participants, we analyzed data of 22 participants who did not attend a single year of schooling (15 Females, 7 Males, age range 14-62, mean 26); additional analyses of the effect of schooling below. Ethical approval was obtained from the ethics committee of Goldsmiths University of London (REISC_1390, 4 june 2018).

Procedure & Stimuli. The experiment was rigorously identical to experiment 2, but the instructions were given verbally by a translator. Participants were compensated in kind (1Kg of sugar, 1Kg of flour, and 500mg of soap).

A typical testing day with the Himba unfolds as follows. On arrival at a village, we park outside the village boundary. The interpreter speaks to the village chief or his representative if he is absent for more than a day. The chief is informed of the general purpose of our visit and asked if he can inform the village that they may participate in our tasks in return for a small gift of flour, sugar and soap (value ~USD 3). We do not offer money for which, in any case, the remote villagers would have little use. If the chief agrees (there has never been a case when he has not) we set up our equipment. We never approach any individual

Himba, but our translator welcomes them if they ask to take part. Occasionally, people are too busy or reluctant to take part, but normally the only reason for obtaining small samples of participants is the absence of a large part of the population away from the village with their herds. In general, the word gets round and people volunteer, sometimes coming from other nearby villages.

In all cases, participants are told that they can refuse to take part in the study or withdraw at any point. We do not collect the names of the participants. We collect information of gender, estimated age, and reported level of education. We explain the purpose of the study in words that can be understood by the participant. Explanations are translated from English to Otjihimba by the local guide. We obtain oral consent, and inform the participants that they will receive the gift in any case, even if they decide to terminate the task. Although we decided to always terminate a testing session if a participant shows signs of distress, this never happened given the trivial nature of the tasks. Beyond acquiring approval that conforms to our professional Code of Practice, we always bear in mind codes of conduct appropriate for the Himba.

The translator explains the following to each participant:

"You are here to participate in a vision task which is a bit like a game. You do not have to participate if you do not want to and can stop at any time if you feel uncomfortable. The task is not difficult and will last for about 30 to 45 minutes. You will be given instructions and do a short practice first. The task is harmless and does not cause any pain. You can ask us not to use your results after you have participated. At the end of the task, you will receive three presents (flour, sugar, and soap). Before we start, you must confirm that you agree with these things. You can now ask any question if something is unclear. If you do not like the task, you can stop at any time and leave. You will receive the presents anyway."

All these elements (plus some simple explanations about the aim of the test, that is, to study "how we see the world") are also given to the chief when we arrive in a traditional village. We hope and expect that the Himba will be direct and indirect beneficiaries, and that the project will contribute to the national and international database on endangered languages and cultures, and to the preservation of the Himba language and culture. We take seriously the responsibilities and the mutualities of benefit that accrue from cross-cultural research with re-

mote peoples, and we believe that we can demonstrate that we have actively furthered remote peoples' interests in our previous research. Issues of identity, belonging and exclusion are currently highly prominent and our project contributes to inter-cultural understanding in a non-trivial way. The intellectual property rights of the Himba in their language and culture is explicitly respected.

Kindergartners

Participants. 28 French kindergartners (mean age 64 months; range 59-70 months; 15 boys, 13 girls) from two classrooms were tested individually in their school, by groups of two, in a quiet room. Each participant was accompanied by one experimenter. They were not compensated for their participation. This experiment was approved by the ethical committee of Université Paris-Saclay under the reference CER-Paris-Saclay-2019-08 after a specific amendment was submitted. Parents were contacted and had to give their consent beforehand. The participants gave oral consent on the day of the experiment.

Procedure & Stimuli. The experiment was identical to experiment 2 except for the fact that we removed the swapped trials to make the experiment shorter.

First Graders

Participants. 156 French first participated in this study. Parents were sent letters beforehand, and could request that children not participate in the project. Participants were tested individually on tables in a quiet room in their school. The data collection was part of the Bien Joué project, approved by the ethical committee of Université Paris-Saclay under the reference CER-Paris-Saclay-2019-042-A1.

Procedure & Stimuli. The experiment was completely identical to the kindergartners' experiment.

1.5.2 Baboons

General Setup.

Participants were 26 Guinea baboons (Papio papio, 18 females, age range 1.5-23 years, mean age 11 years) from the CNRS primate facil-

ity (Rousset-sur-Arc, France). Baboons lived in a 700 m2 outdoor enclosure with access to indoor housing and had a permanent access to ten Automated Learning Devices for Monkeys 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 participate in the research at will, without having to be captured, as the test programs can 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).

Training Scheme

The baboon experiment required several steps of training to ensure that, stimuli set aside, the primates understood the intruder task and could generalize rapidly to new stimuli from different domains. Because we were not sure about the outcome of each of the steps, the entire experiment presented in **Figure 1.6**A was run over three different batches of about one week: a pilot mid-October 2018, a first test of generalization late November 2018, and the test with the quadrilaterals in May 2019.

In the first pilot batch, we tested only 6 primates (Cauet, Dora, Dream, Flute, Hermine and Articho, although the latter animal was not interested in the task and stopped early on). We attempted to start training with displays containing 6 shapes with one intruder. While all baboons except Articho succeeded after 2000 to 3200 trials, the low reinforcement level (chance at one in six) made the early exploration of the task unrewarding and we feared baboons might become disinterested before starting to grasp the task. Therefore, for the two other batches of training, we introduced progressive learning steps with only 3, then 4, 5 and ultimately 6 shapes on display for each trial (see **Figure 1.6**).

In the second batch, we tested all available primates (22 animals) following the structure of **Figure 1.6**A up to and including generalization

1, i.e., the first generalization task. Each primate automatically moved to the next step whenever the error rate fell under 20%. Out of 22 baboons, 18 learned the task to the criterion up to stage 5 and progressed to the generalization task. Out of these 18, all generalized successfully: the percentage of errors was significantly better than chance on the first block with 10 novel images in both presentation modes (binomial test against chance, separately for each baboon: all p's < .001). With further training, all animals again reached the 20% error threshold. Out of the remaining four, three did not reach the end of the first training task at all, and one reached the second training task and stopped. On **Figure 1.6**B, the data reported in the "initial training" and "generalization 1" plots are taken from this batch of data.

The third and final batch tested all available primates (25 animals), following the structure of **Figure 1.6**A. All animals were restarted from the first training task and followed the entire training scheme, only skipping generalization 1 and going straight to generalization 2, then on to the main test. Out of 25 baboons, 20 baboons reached generalization 2. Testing for significant generalization on only 6 different trials could not be done for each animal individually, but we verified that performance was better than chance when grouping the 20 animals together (binomial test, 42 errors in 120 trials, chance at 83.3%, p < .0001). After further training on those stimuli, all of them successfully reached 20% error threshold on generalization 2 and moved on to the test task where they stayed either until they reached 100 blocks of 88 trials (11 primates) or until they stopped performing the task.

Among the 5 baboons who did not participate to the final test, 4 never reached the 20% error threshold on the first training task (three of them stopped being interested in the task early on, one stayed at chance for more than 7500 trials but kept trying). Finally, one primate progressed very slowly over 8800 trials in the first training task, reached the 20% error threshold on block 88 (after having performed 5700 trials in session one and reaching 54% errors), and stopped performing the task. The data reported in **Figure 1.6**B ("generalization 2") and **Figure 1.6**C are taken from this batch of data, i.e. from the 20 primates that reached generalization 2. For reference, **Figure 1.7** shows the evolution of performance over successive training stages for each of those 20 animals (first 20 rows), and the performance for the remaining 6 animals who could not be successfully trained (last 6 rows).

Method

The stimuli were identical to those used with French adults in the second version of the intruder task, except (i) the experiment itself was reprogrammed using custom software specific to the baboon lab, and (ii) baboons received a drop of dry wheat for every correct response. Incorrect responses were followed by a 3-sec time-out indicated by a green screen.

Additional Analyses

To evaluate the heterogeneity across primates, **Figure 1.8** presents the cross-correlation matrix of the error rates of the 20 baboons that reached the testing task, separately for early (first 33 blocks), middle (blocs 34 to 66) and late (blocs 67 to 99) parts of the experiment. Of note, baboons were free to take different numbers of blocks – this explains why there are fewer primates in the "late" category. Within a category, all primates are comparable in that they performed the same number of blocks. We can see that as baboons progressed in their training (and fewer remain), their behavior became increasingly consistent across animals.

1.5.3   Models

Definition of the Symbolic Model

The symbolic model assumes that participants extract the discrete geometric properties of shapes while abstracting away from superficial changes in size, location, orientation and display type (static or sequential). As a result, the model predicts that outlier detection difficulty should depend only on the symbolic distance between the lists of features of the standard and outlier shapes. The more geometric properties a shape has, the more properties a deviant might break, therefore the easier it should be to detect. Because the distance is computed pairwise, this model does currently not account for any difference between canonical and swapped conditions, although a penalty could easily be added.

This model has a single free parameter: a perceptual threshold θ below which the model fails to discriminate lengths or angles and therefore considers them equal. The model considers that two lengths are equal by looking at their ratio: two lengths l_1 and l_2, with $l_1 > l_2$, are consid-

ered equal whenever $\frac{l_1}{l_2} - 1 < \theta$. For simplicity, the same threshold is used for angles: two angles are considered equal whenever they differ by less than $\theta \times \frac{\pi}{2}$

For any given quadrilateral, and for a given threshold, the model computes a vector of bits of length 22, representing the following properties: (i) 6 bits, one per pair of edges, coding whether their lengths are equal or different, (ii) 6 bits, one per pair of edges, coding whether their directions are parallel or not, (iii) 6 bits, one per pair of angles, coding whether their angles are equal or not, and (iv) 4 bits, one per angle, coding whether the angles are right angles or not.

For all reference shapes and all deviants, the model computes the distance between the shapes by counting the number of symbolic properties on which the two shapes differ, and returns a list of 11x4 distances.

The threshold θ was fitted by maximizing the r^2 fit between the symbolic model and the behavioral data of French adults, Exp. 2. For the figures and the analyses, we used the value of 12.5%, but a good fit (r^2 = .37) was already obtained with θ=0, and any value between 3% and 20% yielded similar r^2 values (**Figure 1.10**), indicating that our results do not hinge on a particular choice of behavioral tolerance threshold but rather on any reasonable ability to detect similarity lengths and angles.

Definition of the Neural Network Model and its Variants

We used the CORnet neural network, variant S, whose architecture is schematically depicted in **Figure 1.9**B. We used the weights made available by the authors of (Kubilius et al., 2018) after training on the ImageNet-1000 dataset, where the task of the network was to assign each image of the dataset a label among 1000 possible categories . We did not modify the network or weights, but simply retrieved the activity of units in the internal layers (roughly matching brain areas V1, V2, V4 and IT). To simulate a behavioral trial, we fed the six shapes separately to the network, and retrieved the six vector outputs of the penultimate layer, corresponding to inferotemporal cortex (IT) and which yielded the best performance (**Figure 1.13** shows the predictions when other layers are used). We considered the vector most distant from the average of the others to be the outlying shape, and repeated this process 10000 times to approximate the error rate of the network. We also report the performance obtained from layers

V1, V2 and V4, as well as that obtained by simply picking the outlier on dimensions such as the perimeter or area. The same procedure was repeated using two other top-scoring networks of brain-score.org: DenseNet and ResNet (see **Figure 1.12**).

Variational Auto-Encoder (VAE) Model

For the VAE, we used PyTorch (Paszke et al., 2019)'s off-the-shelf implementation of the canonical model (Kingma & Welling, 2014) (ReLUs and the adam optimizer replaced of sigmoids and adagrad, as recommended by PyTorch's implementation to make the network converge faster.) For each of the 11 reference shape, we generated 6 rotated times 6 scaled images of size 24x24. These 36 images were randomly split in a training set and a testing set, both of size 18. The VAE was then trained over the course of 150 epochs to minimize the loss on the training set, with an evaluation on the testing set at each epoch (**Figure 1.14**A shows the loss on the testing set across epochs for each shape). This gave us access to the VAE's performance across the course of learning for each shape (details in **Figure 1.14**B) and we correlated the performance for each shape with the behavior of both humans (exp.2) and baboons in **Figure 1.14**C. To make the comparison with CNNs more straightforward, for each of our shapes (references and deviants), we extracted the output of the innermost layers of the fully trained VAE, the latent mean and the latent standard deviation layer, from which we replicated the methodology using with the CNNs in order to simulating behavioral outlier detection. The results are summarized in **Figure 1.14**D: overall, the output of the innermost layers varied very little across shapes, and those variations did not capture the variance of either any of the human population, or any of the baboons.

1.5.4   Additional Analyses, Results and Discussions

Detailed Analysis of the Visual Search Experiment

The error rates and mean response times of the visual search experiment were entered into an ANOVA with shapes (11-level factor), number of items (as a numerical factor in 6, 12 or 24), target presence (present or absent), and their interaction, and participant as the random factor. For error rates, there was a significant effect of shape ($F(10,100) = 16.15$, $p < .0001$), of number of items ($F(1,10) = 20.33$, $p = .0011$), of target presence ($F(1,10) = 31.45$, $p = .0002$), of shape and

target presence (F(10, 100) = 2.03, p = .0375), but little interaction between number of items and target presence (F(1,10) = 4.91, p = .0509), no significant interaction between shape and number of items (F(10,100) = 0.87, p = .564), nor a three-way interaction between shape, number of items and target presence (F(10,100) = .42, p = .93). For response times, there was a significant effect of shape (F(10,100) = 9.89, p < .0001), of number of items (F(1,10) = 26.70, p = .0004), of target presence (F(1,10) = 29.58, p = .0003), of the interaction between shape and number of items (F(10,100) = 3.71, p = .0003), but no significant interaction between number of items and target presence (F(1,10) = 3.78, p = .080), no significant interaction between shape and target presence (F(10, 100) = 0.90, p = .54) nor a three-way interaction between shape, number of items and target presence (F(10,100) = .52, p = .88).

The error rates closely followed the classical geometric regularity effect observed in the intruder task, as there was a significant correlation between the mean error rates in visual search and the French adults error rates in the intruder task (experiment 2), both overall (R^2=.98, p < 0.0001, **Figure 1.3**B) and regardless of the number of items on the screen (6 items, R^2 = .86, p < .0001, 12 items R^2 = .93, p < .0001, 24 shapes R^2 = .96, p < .0001). The mean RTs also followed a geometric regularity effect overall (R^2 = 0.88, p < 0.0001) and for each number of items (6 items, R^2 = .90, p < .0001, 12 items R^2 = .89, p < .0001, 24 shapes R^2 = .85, p < .0001).

To test for the seriality of visual search, the mean response time within each subject was entered in separate ANOVAs for each shape, with number of items (a numerical factor equal to 6, 12 or 24), target presence (present or absent), and their interaction as factors, and participants as a random factor. All shapes elicited a serial visual search (all *p* < 0.05 for the effect of the number of items; **Figure 1.3**C).

For each shape and participants, we computed the slope of the visual search for both present and absent condition by fitting a linear model on the median of the response times per item number. We then tested whether the slope of the visual search in the "absent" condition was twice the slope of the "present" condition, as expected from serial search (Wolfe, 1998). For each shape, we used a paired t-test to compare, across subjects, the distribution of slopes in the absent condition and the distribution of twice the slope in the present condition. None of those differences except for one shape were significant at the .05

level (right-kite: p = .044; all other shapes p > 0.05). Additionally, the best fit of a linear model across subjects that predicts the slope, as computed above, when the item is absent from the slope when it is present had a significant (p = .0003) coefficient of 1.66, SE = .30, not significantly different from 2 (p = .29).

Finally, the slope of the visual search exhibited a geometric regularity effect: it correlated with the error rates observed in experiment 2, both overall (R^2 = .70, p = .0013), and when the target was present (R^2 = .60, p = .0047; **Figure 1.3**C) and absent (R^2 = .68, p = .0019).

Role of Feedback in Human and Baboons

It could be argued that, in the intruder task, human subjects were treated differently from baboons because on error trials, the visual feedback the correct responses was highlighted in green (surrounded with a green square for training images, filled in green for geometric shapes), thus giving an additional indication about the task. To examine whether this made any difference in humans, we analyzed the data from each participant's very first trial with a given shape, before they received any feedback. In both experiments 1 and 2, such analysis produced results that were indistinguishable from the results of the full dataset analysis. The error rates were strongly correlated with those of the full dataset (exp. 1: r^2 = .99, p < .0001; exp. 2: r^2 = .94, p < .0001); the best fit of a linear regression "full data $\sim \beta_0 + \beta_1$ * first_trial" had an intercept β_0 not significantly different from 0 and a slope β_1 not significantly different from 1 (all p's > .1), suggesting that little or no learning took place in human participants over the course of the 88 trials.

Retraining of the Neural Networks with Geometric Shapes

A possible reason for the failure of neural networks to mimic human data could be that the geometric shapes differed from the network's training data (colored photographs). Perhaps our stimuli ended up on the extremities of the feature hyperspace, thus leading to inconsistent or chaotic behavior of the network. Here we present several arguments that mitigate this possibility.

First, the labels that were attributed to the shapes were highly consistent and suggested that the network did recognize them. **Table 1.3** provides details of the labels given by CorNet without retraining.

The network overwhelmingly categorized the shapes as "envelopes," and its next choices were mostly "Band-Aids" or "binders", with a few interesting deviations (e.g. trapezoids were classified as "lamp-shades"). This result was replicated almost perfectly with DenseNet, while ResNet primarily categorized the shapes as envelopes, followed by noisier categories.

Table 1.3: For each shape, columns show the first five top predictions and the associated average confidence level, for CorNet trained on ImageNet. Each shape was presented in 36 slightly different variants (6 rotations X 6 scaling factors). We averaged these predictions for each shape, and put in each column the prediction whose average associated confidence level was the highest, and the corresponding average confidence level.

	label1	label2	label3	label4	label5
Rectangle	envelope, 71.68%	band aid, 7.22%	band aid, 2.02%	spatula, 2.28%	letter opener, 1.61%
Square	envelope, 74.62%	envelope, 33.07%	switch, 2.11%	book jacket, 1.66%	face powder, 1.57%
iso-trapezoid	envelope, 62.05%	envelope, 37.8%	lampshade, 6.98%	lampshade, 4.51%	binder, 2.04%
Parallelogram	envelope, 64.91%	band aid, 7.64%	cleaver, 3.75%	binder, 2.7%	table lamp, 2.57%
Rhombus	envelope, 58.4%	band aid, 9%	letter opener, 4.61%	book jacket, 2.8%	wing, 3.96%
Kite	envelope, 68.63%	band aid, 10.18%	binder, 2.04%	carton, 1.64%	switch, 1.68%
right-kite	envelope, 74.28%	band aid, 7.51%	face powder, 2.25%	binder, 1.8%	binder, 1.89%
Hinge	envelope, 77.49%	band aid, 3.91%	table lamp, 2.08%	band aid, 1.58%	cleaver, 1.95%
right-hinge	envelope, 77.09%	band aid, 4.69%	letter opener, 2.38%	binder, 1.71%	table lamp, 1.55%
Trapezoid	envelope, 72.57%	band aid, 8.41%	binder, 2.62%	face powder, 2.46%	face powder, 1.63%
Irregular	envelope, 59.55%	envelope, 28.07%	binder, 3.3%	carton, 2.6%	table lamp, 2.02%

Second, the three convolutional neural networks we tested were highly consistent in the error rates that they predicted; and, as showing in **Figure 1.9**C, **Figure 1.9**D and **Figure 1.12**, these predictions were not random, but tightly correlated with baboon behavior.

Third, we examined how CorNet would perform if it received additional training with geometric shapes (similar perhaps to a young child being exposed to geometric shapes and toys). Our results are summarized in **Figure 1.13**. We trained different versions of CorNet to categorize either all of our 11 shapes ("All shapes"), or the subset of 5 shapes that have a common name in English ("Nameable shapes", i.e. square, rectangle, rhombus, parallelogram and trapezoid).

Our goal was to keep the properties that made the original network successful in image recognition, but also familiarize it with our shape space. We proceeded as follows: (i) we added either 5 or 11 output unit to the output (decoder) units; those were fully connected to the previous layer and randomly initialized, while keeping the rest of the network intact; (ii) we trained the network to categorize solely our shapes (solid white on black images, one shape per image, same rotation and scaling factors as for behavioral experiments), and allowed the backpropagation to modify either the entire network ("All layers"),

or only the last main group of layers ("IT only"), with training on 80% of the images per shape and validation on the remaining 20% (plotted on **Figure 1.13**A); the learning optimizer was Adam with a learning rate of 1.0E-6; (iii) we checked, for each training step, the performance of the updated network on the original dataset, ImageNet. After sufficient training, all conditions lead to perfect categorization of all geometric shapes, including on the validation set of shapes. Meanwhile, performance on ImageNet remained high, with a higher loss when the entire network was allowed to change in order to accommodate the new geometric shapes (**Figure 1.13**A); (iv) finally, using our multiple-regression methodology, we compared the predictive power of each of the four types of retrained network with that of our symbolic model.

The results appear in **Figure 1.13**B. None of the four training schemes significantly improved the predictive power of the neural network model on human participants. As for baboons, the various training conditions either did not change anything or worsened the predictive power.

Possible Effect of Non-matched Visual Properties

We matched our 11 shapes on several important size variables (see the section on "Stimuli" above). However, those constraints imposed that we could not match them for other visual properties. In particular, the shapes were not strictly equalized in area and perimeter (see **Table 1.2**). Given the random scaling we added to each of the six shapes, choosing the outlier based on area or perimeter could not give rise to the high level of performance observed in humans. Furthermore, although the error rate predicted by such strategies varied across shapes, regressions indicated it could not explain the geometric regularity effect observed in humans (all $p > .05$). In **Figure 1.11** we show the predictions and correlation with baboons: the area-based strategy significantly correlates with the observed behavior in baboons ($p < .0001$) while the perimeter-based strategy does not ($p = .5$). Both strategies elicit more errors than the baboons, indicating that these strategies do not suffice to explain the baboons' behavior.

Possible Effect of Education in Himba Participants

The Himba population we sampled was heterogeneous in its formal education background. Out of the 44 participants we tested, 22 never attended school (those subjects are reported in the main text), and the 22 others ranged from 1 to 8 years of school, with otherwise comparable general demographic information.

This variability provided an opportunity to test for the effect of the number of years of schooling on the geometric regularity effect. The error rates were entered into an ANOVA with geometric regularity (a numerical factor determined by the error rate in French subjects in experiment 2), years of schooling (a numerical factor ranging from 0 to 8), their interaction and participants as random factors. There was a significant effect of shapes ($F(1, 42) = 229.21$, $p < .0001$) but no significant effect of the years of schooling ($F(1, 42) = 0.05$, $p = .82$) and no significant interaction ($F(1,42) = .65$, $p = 0.43$). This negative finding does not exclude that, with more participants, an effect of education would be observed. However, this additional analysis confirms the universality of the geometric regularity effect.

Possible Impact of a "Carpentered World"

Figure 1.15: Rectilinear testing booth used in the CNRS primate facility of Rousset-sur-Arc, France

The Western environment has been called a "carpentered world" (Segall et al., 1963), where vision is bombarded with many rectilinear objects (e.g. buildings, tables, books, etc.). Could such a difference in the statistics of the environment explain the geometric regularity effect? We believe that this is unlikely for several reasons explained in the discussion part of the main text. The main reason is that we replicated the effect in the Himba, but failed to observed it in baboons. The rural settlements of the Himba are quite unlike industrialized societies and their environment is relatively free of rectilinear objects (for photographs, see e.g. https://en.wikipedia.org/wiki/Himba_people).

Conversely, the baboons were not wild animals, but grew up and lived in an environment comprised of both natural objects (trees, rocks) and man-made, rectilinear objects (buildings, doors, testing booths, computer screens... see inset picture). Arguably, the baboon's environment is equally or even more "carpentered" than the Himbas, see **Figure 1.15**.

1.6 Addendum Post-publication

I have been made aware of an article by Zekun Sun and Chaz Firestone (Sun & Firestone, 2021) that was published during the reviewing process of the previous chapter as an article. The content of that article merits a special discussion here: the authors generate shapes of increasing complexity by parametrically modulating their internal skeletons, and observes that (i) geometric complexity impacts visual search, even when low-level features are matched, (ii) geometric complexity drives exploration and engagement, and (iii) geometric complexity is cognitively penetrable.

Future work using the quadrilateral shapes presented in **chapter 1** should contrast our symbolic model with a model which leverages shape skeleton – though the notion of which skeleton is appropriate itself is not completely obvious (Firestone & Scholl, 2014). If the stimuli used cannot separate the two models, it could be useful to generate additional shape which would score low on one metric and high on the other to better understand the cases in which the two models differ. It is useful to observe that while shape skeleton is appropriate for a very wide number of visual objects, including contour of real-work objects, our metric is only defined for geometric shapes, see **chapter 5** for a longer discussion of this observation.

Chapter 2

MEG and EEG evidence for symbolic and non-symbolic neural mechanisms of geometric shape perception in adults and infants

Abstract

How does the human brain encode highly structured visual objects such as geometric shapes? We test the previously proposed hypothesis that there are two strategies mentally representing geometric shapes: a perceptual bottom strategy which can be mimicked by a neural network model of object recognition, and a symbolic strategy which represents shapes using a list of their exact geometric properties such as parallel sides and right angles. We exposed participants to streams of geometric shapes interleaved with occasional oddballs, organized by blocks of geometric complexity. We find that the extent to which we can decode the oddballs from the brain signal matches how much fewer symbolic properties they have compared to dominant shape of the block. Crucially, we show that the neural representation of the shapes is highly structured in both space and time, and goes through two distinct phases: first, it is organized as predicted by the neural network of object recognition in the occipital areas, and then it is organized as predicted by a symbolic model in a broad dorso-frontal brain network. In infants, we provide preliminary evidence of different processing associated with different shapes using a similar task in EEG, but fall short of unequivocal conclusions; we discuss possible ways forward.

89

> There are a number of cognitive systems which seem to have quite distinct and specific properties. These systems provide the basis for certain cognitive capacities [...] .The language faculty is one of these cognitive systems. There are others. For example, our capacity to organize visual space, or to deal with abstract properties of the number system, [...]
>
> _____
>
> Noam Chomsky, quoted in (Rieber, 2013)

Our experiments in the previous chapter shed light on the possible co-existence of two strategies for dealing with the mental representation of geometric shapes. In the present chapter, I test this possibility directly by looking at the spatio-temporal dynamic of a shape perception task with M/EEG, in adults and infants.

2.1   Introduction

In the previous chapter, we posited that two models coexist when perceiving quadrilateral shapes, even for our educated French adult groups. The first model would be shared with non-human primates, and corresponds to a perceptual strategy, well captured by current neural network models of the ventral visual pathway. The second model would rely on geometric properties such as right angles and parallelisms, while the second would fall in the. However, in that population the data was very strongly in favor of the symbolic model – could the behavior only reflect part of the underlying mental process? If so, we expect that using neuroimaging techniques we can separate, both in space and in time the two unfolding mental processes.

Using the same stimuli as the previous experiment, we devised a new behavioral experiment based on distances *between* shapes, as opposed to our outlier-detection task so far which could only give us within-shape information. This gives us a notion of distance between different shapes, by looking at an aggregate of the response time and the error rate. We then model this confusion matrix using exactly the same models as those in the previous chapter, serving as a confirmation that they generalize beyond the exact task they were designed for. We can also perform Multidimensional scaling (MDS) to get a measure the most import dimensions at play when performing the intruder task with geometric shapes, and see whether those dimensions map onto elements of our models.

Then we could use these models to explore the spatio-temporal dynamics of the perception of geometric shapes in MEG. Inside a MEG recording device, we presented participants with long streams of our quadrilateral shapes arranged in blocks with occasional intruders, following an oddball paradigm. Then we tried to (i) see whether the intruder created bigger surprisal signal during regular shape blocks than irregular ones, and (ii) account for the brain signal in terms of both visual and symbolic models of shapes, both temporally and spatially using source reconstruction.

Finally, we try to the question left open in the previous chapter of the influence of the statistics of the environment (and, to a lesser extent, education) by testing three-to-four-month-old infants using a paradigm very similar to the one used in MEG: an oddball paradigm, with EEG recordings of the subjects. If the signal for different shapes followed systematic patterns that we can account for using our symbolic model, it would strongly strengthen the argument in favor of a universally available mechanism for exact geometric properties in humans. However, the data so far is on the fence about this question, and the results presented here, while informative, are inconclusive about that question.

2.2 Method

2.2.1 Participants

Adults, Behavioral

In total, 342 participants took part in this online experiment. The experiment was advertised on Twitter, starting from the 22nd of March, 2022. People interested to participate could simply click on the provided link, read and accept a written consent, in which they declared not to be legally minor. Participants were informed they could withdraw from the experiment at any moment by simply quitting the webpage. The procedure and the consent were approved by the local ethical committee (reference: CER-Paris-Saclay-2019-063). Data collection for the purpose of the study was stopped soon after, on the 8th of April, 2022. 330 participants met our criteria for data analysis, i.e. answering all questions proposed and reporting an age above the age of 18th. Analyzed participants' demographics were as follows: ages 20 to 84, 1st quartile 37, median 53, 3rd quartile 65; genders 142 females,

177 males, 6 non-binary, 2 others, and 3 preferred not to answer. Participants dominantly reported being from the US (228 participants), followed by Egypt (15), Canada (12), Italy and Great Britain (6 each), and a long distribution of other countries.

Adults, MEG

Twenty healthy French adults (13 females; 21-42 years old, mean: 24.9 years old, SD: 8.1 years old) participated in the MEG study. All participants had normal hearing, normal or corrected-to-normal vision, and no neurological deficit. All adults provided informed consent, and adult participants were compensated for their participation. For all but one participant, we had access to anatomical recordings in 3T MRI, either from prior, unrelated experiments in the lab, or because the MEG session was immediately followed by a recording: analyses that require source reconstruction are performed on nineteen subjects.

Infants

We collected data from a total of 43 three-to-four-month-old infants, with normal pregnancy and birth (GA > 38 weeks, Apgar scores ≥ 7/8 at 1/5 min, birthweight > 2.5 kg, cranial perimeter ≥ 33.0 cm), tested in the lab. The protocol was approved by the regional ethical committee for biomedical research (Comité de Protection des Personnes Region Centre Ouest 1, EudraCT/ID RCB: 2017-A00513-50), and the study was carried out according with relevant guidelines and regulations. The experiment was run during two dissociated periods with slightly different methodological choices: during the first period from July 22nd, 2020 to October 21st, 2020, we collected data from 23 participants (9 females; 14 males; average age 105 days, sd=17 days), out of which 3 were considered impossible to analyze due to excessive movement or very noisy data. The second group was comprised of 20 new participants collected between December 9th, 2021 and February 11th, 2022, and follows a similar distribution of age and gender. Out of this group, a single infant was considered impossible to analyze due to a very high number of epochs categorized as "bad" by the preprocessing pipeline. The testing age was chosen to be the youngest at which we could reasonably assume that vision was sufficiently developed. Caregivers gave their written informed consent before starting the experiment.

2.2.2 Materials
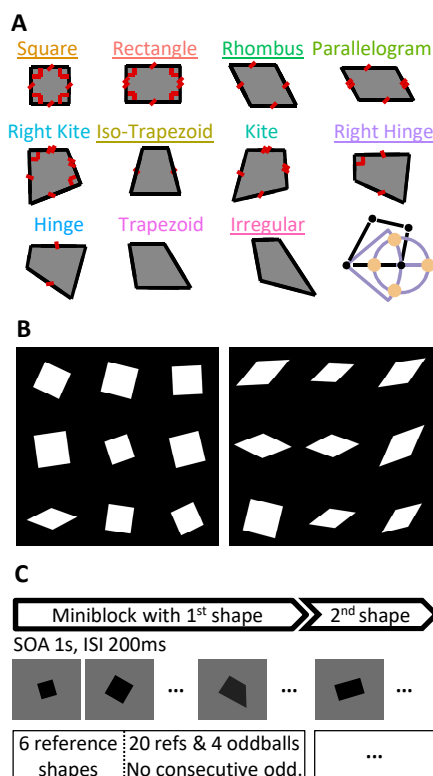
Figure 2.1: **A. Stimuli** The eleven shapes used throughout the experiments in this chapter, as well as the general template used to generate deviant version of each shape (bottom-right). **B. Behavioral Task** Two examples of intruder test using the same shapes, either in the canonical condition or in the swapped condition. C. **Passive presentation task.** General structure of the passive presentation tasks.

Geometric shapes were generated following the procedure described in (Sablé-Meyer, Fagot, et al., 2021) (**Chapter 1**). Deviants, when used, were generated following (Sablé-Meyer, Fagot, et al., 2021) by displacing the bottom right corner by a constant distance in four possible positions: that distance was a fraction of the average of all pairs of points, which was standardized across shapes and was .3 which was used in (Sablé-Meyer, Fagot, et al., 2021) for the hard condition.

Behavior

In the behavioral task, no deviants were used. Instead, participants had to detect one of the eleven shapes amongst eight copies of one of the other shapes. Shapes were presented in a 3x3 grid, in pure white on pure black. Each shape was randomly scaled differently by shuffling possible scaling factors [0.85, 0.88, 0.92, 0.95, 0.98, 1.02, 1.05,

1.08, 1.12, 1.15], and randomly rotated similarly using [-25°, -19.4°, -13.8°, -8.3°, -2.7°, 2.7°, 8.3°, 13.8°, 19.4°, 25°]. Upon answer, auditory feedback was provided in the form of upward/downward going tunes, and coloring of the shapes (green for the intruder, possibly red for the chosen shape if different from the correct response). Participants took 110 trials, one for each reference/intruder pair of shapes, with no two identical reference shapes used in consecutive trials, and the outlier of a trial being different from the reference shape of the previous trial. Two examples of trials are shown in **Figure 2.1**.

M/EEG

In the MEG and EEG tasks, we relied on purely passive shape perception, as the experiment needed to be suitable for three-month-old infants. Shapes were presented centered on the screen, one shape every second, with shapes remaining onscreen for 800ms and a centered fixation cross present between shapes for 200ms. Additionally, to keep babies attending visually, during their 800ms presentation, the shapes slowly increased in size: in total, a scale factor of 1.2 was applied over the course of 800ms, with linear interpolation of the shape size during the duration of the presentation. The size change was intended to capture infants' attention, and 1Hz presentation was a trade-off between the number of trials per participants we could collect and a minimum duration required for babies (Yu & Smith, 2016).

Shapes were presented in blocks, following an oddball paradigm. Within a block, all shapes were identical up to scaling (randomly sampled in [0.875, 0.925, 0.975, 1.025, 1.075, 1.125]) and rotation (sampled in [-25°, -15°, -5°, 5°, 15°, 25°]), except for a few oddballs which were deviant version of the reference shape.

In MEG in adults, blocks were 30 shapes long, with exactly four oddballs that replaced any shape after the first six and never two oddballs in a row, for a final 13.3% of presented oddballs. A run was made of 11 blocks, one per shape in random order, and participants were recruited to attend 8 runs (although a few were stopped one run earlier).

In EEG in babies, the exact parameters changed between two experimental groups. In both groups, only a subset of the shapes were used to shorten the experiment duration: square, rhombus, rectangle, isosceles-trapezoid, right-hinge and irregular (underlined on **Figure 2.1.A**). In the first group, blocks were 60 shapes long, with 8 reference

shapes to start with and then 40 references and 12 oddballs shuffled so that no two oddballs are in a row, for a final 20% of presented oddballs. In the second group, we radically shortened the blocks down to 15 (with some randomness, each block's length was uniform in [13, 17]) and no oddball were presented.

2.2.3 Procedure

Adults, Behavior

The experimental procedure started with instructions, followed with a series of questions on demographic aspects (device used: mouse or touchscreen, country of origin, gender, age, highest degree obtained) and on subjective self-evaluation assessments, with answers on a Likert scale from 1 to 10: current skills in mathematics; current skills in first language. Then, participants performed the task. The instruction were as follows: "The game is very simple: you will see sets of shapes on your screen. Apart from small rotation and scaling differences, they will be identical, except for one intruder. Your task is to answer as fast and accurately as you can about the location of the intruder by clicking on it. The difficulty will vary, but you always have to answer."

Adults, MEG

After inclusion by the lab's recruiting team, participants were prepared for the MEG with ECG and EOGs captors, as well as Head Position Indicator coils, which were digitalized to track the head position throughout the experiment. Then we explained participants that the task was a replication of an experiment with babies, and therefore was purely passive: they would be presented shapes and were instructed to pay attention to each shape, while moving, blinking and saccading as little as possible. Then they sat in the MEG and we checked the head position, ECG/EOG and MEG signal. From that point onward, we never opened the MEG door again to avoid having to reset the signal and allow for cross-run decoding and generalization. Then participants took typically 8 runs consecutively, with small breaks between runs to rest their eyes. At the end of the experiment, participants took the intruder test from **Chapter 1**, and finally we spent some time debriefing with participants the goal of the experiment.

Infants

After the caregivers arrived and formal requirements were met, an elastic EEG net soaked in salted water was put over the subject's head; then both the caregiver and the subject moved to the experiment room, where the net was connected to the recording apparatus and last check of the data quality were performed. Subjects were tested in a sound-proof Faraday cage equipped with a computer screen and loudspeakers. Infants were held by a caregiver, sitting on their lap, and their position was chosen to guarantee personal comfort while ensuring good-quality data acquisition, notably avoid excessive movement, and ensure easy gaze to the computer screen. Stimuli were displayed centered on the screen, once per second. To maximize engagement, each stimulus was synchronized with an auditory stimulus, identical for all shapes. Breaks were taken whenever necessary, using soothing music and colorful stimuli on the screen, in a few cases the experimenter entered the room with puppets to distract the subject for a bit. The experiment ended as soon as infants became restless.

In the first group of participants, another experiment was intermixed with the geometric shape, for a colleague's independent project. That project displayed visual mixture of faces and houses at a rate of 1 per second as well. When participants started becoming agitated with either type of stimuli, we switched to the other type of stimuli, often capturing their attention again. The face/house stimuli are not analyzed in the present work, and this particular procedure was not replicated in the second group of participants.

2.2.4 Data preprocessing

MEG preprocessing

MEG Data. The preprocessing of the data was performed using `MNE-BIDS-Pipeline`, a streamlined implementation of the core ideas presented in (Jas et al., 2018) and leveraging BIDS specifications (Niso et al., 2018; Pernet et al., 2019). The pipeline performs automatic bad channel detection (both noisy and flat), then apply Maxwell filtering and Signal Space Separation on the raw data (Taulu & Kajola, 2005). Then the data is filtered between .1 Hz and 40Hz, and resampled to 250 Hz. Then it is epoched for each shape, staring 100ms before stimulus onset and stopping 1100ms after, and the relevant metadata for each epoch is recovered from the stimulation procedure at this step.

Artifacts in the data (e.g. blinks, and heartbeats) are repaired with signal-space projection (Uusitalo & Ilmoniemi, 1997), and threshold derived with "autoreject global" (Jas et al., 2017). For source reconstruction, some preprocessing steps are performed by fmriprep (see below). Then, sources are positioned using the "oct5" spacing with 1026 sources per hemisphere, and we use the e(xact)LORETA method (following recommendations from (Jatoi et al., 2014; Pascual-Marqui et al., 2018)) using empty-room recordings performed right before or right after the actual experiment to estimate the noise covariance matrix.

Fmriprep's MRI Anatomical Preprocessing. T1-weighted (T1w) images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008). The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow. Brain tissue segmentation of cerebrospinal fluid, white-matter and gray-matter was performed on the brain-extracted T1w using fast (Y. Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (Dale et al., 1999) and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017)

Infants

Raw data was recorded ad 500Hz using EEG nets comprised 128 sensors, continuously throughout the entire experiment. The data was preprocessed using a recent version of the APICE pipeline (Fló et al., 2022), with a pass-band filter of .2-40Hz, epoching of [-150ms, 1150ms] window around each shape presentation, and otherwise default parameters for artefact rejections, bad channels interpolation, and bad epoch tagging.

2.3 Results

2.3.1 Adults, Behavior
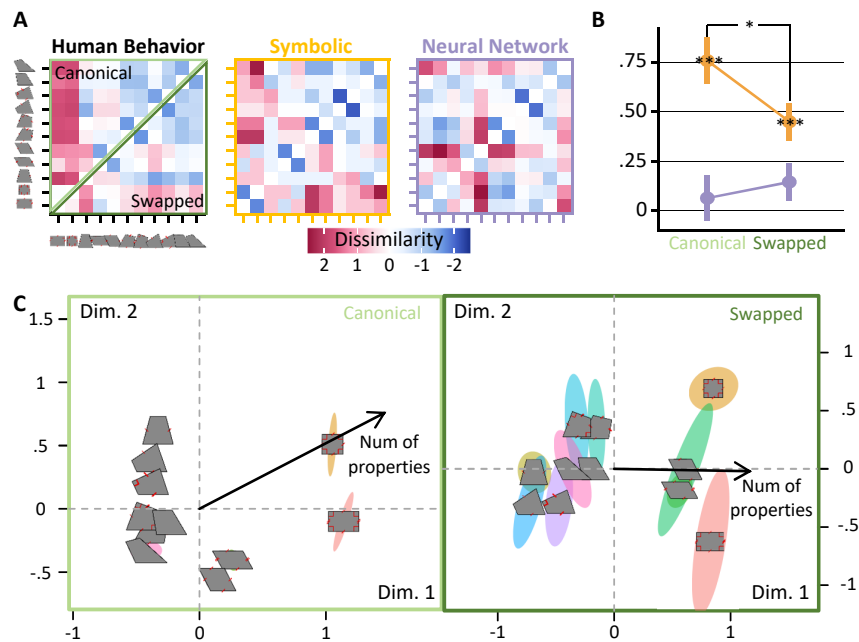
Figure 2.2: **A.** Confusion matrix from the human behavior, as well as the two models used to account for it. Data was centered and scaled to be directly comparable, units are therefore z-scores of the associated dataset. Note that while the two models are symmetrical, the human behavior is not (see the canonical vs. swapped conditions, indicated in different shades of green on this plot), and is therefore modeled differently afterward. **B.** Coefficients in a GLM predicting humans' confusion matrix with our models, either in the canonical or in the swapped condition. In both cases, the symbolic model is significantly different from 0 and significantly better than the IT model. The symbolic model significantly worsens in the swapped condition when compared to the canonical condition, but while the IT model appear to improve it does not reach significance. **C.** MDS stemming from the confusion of the canonical (left) and swapped (right) conditions. Two principal dimensions are identified, and our 11 shapes are positioned in that space, together with confidence ellipses obtained by resampling (500 folds, 95%-confidence interval ellipses). The black arrows represent the best fit of an additional dimension, the "symbolic model" prediction, with both data-derived principal dimensions, and mostly lines up with the first dimension.

Results

Participants were much better than chance overall, and for each reference shape (Error Rates [ER]: chance level is 8/9=.88; overall ER=.109±.004; min=.02±.003, max=.186±.01; "±" indicates standard error of the mean). The differences across shapes significantly impacted both error rates and response times, see **Table 2.1**, similarly the intruder had a significant effect, and whether the trial had a hard shape within easier shapes or the other way around also had a significant effect. Both reference, intruder and interaction come out as having significant effects on both RT and ER (Type-II Anovas on (g)lms, all *ps* <.001).

Table 2.1: Statistics of the ANOVA across subjects associated with the three relevant predictors, for two dependent measures, with the statistics and the effect size associated with each component. "canonical" refers to trials where the intruder was a simpler shape than the reference one.

DV	Predictor	η^2_G	Statistics
er	reference	0.19	$F_{10,3290}$=114.34; p<.001
	intruder	0.109	$F_{10,3290}$=65.37; p<.001
	canonical	0.035	$F_{1,329}$=79.48; p<.001
rt	reference	0.064	$F_{10,3290}$=28.24; p<.001
	intruder	0.037	$F_{10,3290}$=16.27; p<.001
	canonical	0.013	$F_{1,329}$=15.76; p<.001

Thanks to the contrastive nature of the search, we can also measure a dissimilarity metric based on t the visual search. Because we only had one trial per participant and condition, we didn't do any summary statistics at the level of participants: in order to build a confusion matrix across shapes, we therefore used $d = \frac{avg_{ER}}{avg_{RT}}$ as a dissimilarity measure (average Error Rates divided by average Response Times.) Thanks to the high number of participants, no norming was required at the participant level (e.g. using the rank of the response time, or rescaling the response times on a shared scale). As confirmations we performed all the analyses with such harmonization strategies and observed only negligible differences and do not report them.

Then, we predicted the behavioral confusion matrix using two models, straightforwardly adapted from the previous chapter. On the one hand,

we computed a geometric-property distance matrix. For each shape we computed a feature vector as indicated in (Sablé-Meyer, Fagot, et al., 2021); then the distance between two shapes is simply the distance between two feature vectors, otherwise said the number of symbolic properties by which two shapes differ. On the other hand, we used CORnet to compute a visual feature vector: we created various rendering for each shape (using various rotation and scaling parameters), and then computed the internal representations associated with each exemplar of a shape in the successive layers of CORnet. Then we preformed representational similarity analysis (RSA) on those internal representation, using `rsatoolbox`[1] to compute cross-validated mahalanobis distance between all possible pairs of shapes. The two models are not significantly correlated across all 55 possible pairs (linear model, $F_{1,53}$= 2.1, p=.15), but their correlation coefficient sits at .19 so further analyses will try to perform GLMs instead of separate correlations.

Note that both of these matrices are symmetrical, i.e. expect the distance from shape A to shape B to be the same as that from shape B to shape A. But this is not true at all in general, as evidenced by a vast literature *using* search asymmetries to understand visual attention, basic features, or novelty (Treisman & Gormican, 1988; Treisman & Souther, 1985; Wolfe, 2001). For that reason, we performed separated analyses for both conditions. In both cases, the symbolic model's beta was significantly above zero, and vastly dominated the visual model which was not significantly different from zero. Interestingly, the symbolic model's effect was significantly different between the canonical and the swapped condition: it decreased. The visual model did not significantly change.

If we replace the human confusion metric with simply $\frac{1}{RT}$, we can compute a confusion matrix for each participant, instead of averaging over participants, and repeat the same analysis with a mixed-effect model, to get much more sensitive estimation of the respective coefficients for both models. The overall effects and effect sizes are virtually the same, but now all the estimations have much narrower error bars and while

---

[1]I couldn't find a reference citation for this toolbox, but as of writing its documentation states, "The rsatoolbox is developed through a community effort by the labs of Nikolaus Kriegeskorte, Jörn Diedrichsen, Marieke Mur and Ian Charest. It was conceived during the RSA retreat 2019 in Blue Mountains, Ontario. The toolbox replaces the 2013 matlab version the toolbox of rsatoolbox previously at ilogue/rsatoolbox and reflects many of the new methodological developements."

effect sizes so more tests reach significance threshold. In particular, in this more fine-grained model there is a significant effect of the visual model in both conditions, and this effect significantly increases when the effect of the symbolic model decreases.

Finally, we performed MDS on the behavior's confusion matrix to estimate from the data the most impactful dimensions that explain participants' behavior. Once again, we performed these analyses separately on the canonical and swapped conditions as they elicit different behaviors. To estimate the variability of the MDS, we performed a resampling analysis by repeatedly sampling with repetitions a list of subjects from our data, performing MDS again, aligning the new MDS with the full MDS using Procrustes analysis to remove meaningless differences such as rotation, translation and dilation, and finally deriving variability by repeating this process 500 times. Our implementation of this method follows that of (Borg et al., 2018; Borg & Groenen, 2005; De Leeuw & Mair, 2009; Mair et al., 2022). **Figure 2.2** shows the shapes projected on the principal dimensions for both canonical and swapped conditions, as well as 95% confidence intervals (ellipses). Note that ellipses in the swapped conditions are bigger because the data itself is more variable: swapped conditions are typically harder, leading to longer response times, but crucially also an increased variability in response times across participants (the variability scales with the response time, a case of heteroscedasticity.) What is more, since each shape is associated with a number of symbolic properties, we can also project the vector of symbolic properties onto our data-driven dimensions. If our model is much poorer than the inherent dimensions stemming from the data, its projection should be "short" (i.e., close to the origin), and its orientation random. However, what we can see is that in both conditions, our model's norm is greater than 1 and aligns significantly with our first dimension. Using resampling again, we can see that (i) in both conditions, our model's projection significantly correlates with dimension 1, as intuited from the graph (Student test on the distribution of slopes for the 1st dimension, canonical t=1768.4, p<.0001, swapped t=236.2, p<.0001), (ii) in both conditions, our model's projection also correlates with dimension 2 (canonical t=329.4, p<.0001; swapped t=16.5, p<.0001, this can't be seen from the graph but indeed about a fourth only of our sampling folds had a negative slope), and finally (iii) the correlation with dimension 1 is always greater than with dimension 2 (Paired student test; canonical t=269.8, p<.0001; swapped t=42.6, p<.0001).

Discussion

First, these new results indicate that our model for shape complexity, presented in the previous chapter, is valid not only within shapes, but also across our different quadrilaterals: they systematically differ in complexity, and we can model this difference using a model that worked successfully for the intruder detection task.

Furthermore, we can see that the main dimensions that are required to account for the data sort our shapes according to a main axis that resembles our theory-motivated model. This result suggests that not only our model accounts some of the behavior, but it characterizes what dominates the decision process when picking the intruder, an observation we couldn't do based solely on model comparison with other models such as neural network models.

Finally, this provides us with a new way to look at our shape: the similarities and differences between the different shapes. This is an important step for brain imaging analyses, because we can now use this the newly defined confusion matrix to explore the spatio-temporal aspects of shape processing that MEG and fMRI provide.

2.3.2   Aduls, MEG
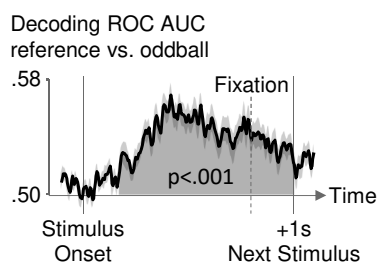
Decoders

Figure 2.3: **Overall Decoding Performance.** Performance (ROC AUC, chance=.5) of a linear decoder classifying reference vs. oddball shapes from the magnetometer of the MEG signal; averaged over all participants, ribbon indicates std. error across participants, grayed area indicates a significant cluster obtained by non-parametric temporal clustering, p<.001.

First, we wanted to confirm that despite the absence of an explicit task, participants were paying attention to the visual presentation and detecting outliers. For this we trained a decoder on the magnetometers of the MEG signal to discriminate reference shapes versus oddballs, for
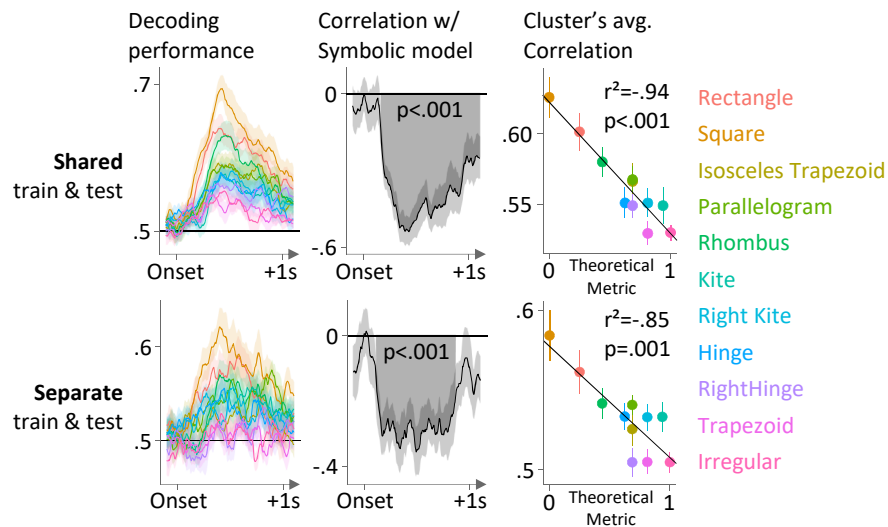
Figure 2.4: Decoding Performance. Performance of a decoder trained to classify reference vs. oddball from the magnetometer of the MEG signal. First column: performance for each shape indicated with colors. Second column: correlation across shape, for each time step, with a simple symbolic model of shape complexity: to keep the usual ordering, the data is correlates with 1 minus the number of symbolic properties rescaled from 0 to 1; a value of 0 indicated that the shape has the maximum number of symbolic properties in the set, and 1 the minimum. Right column: visualization of the correlation across the main significant cluster's average data. In the top row, row a single decoder was trained with data from all shapes, while in the bottom row one decoder was trained for each shape.

each participant and each time step. To avoid confounds due to the signals' autocorrelation, we train and test our decoders on separate runs systematically, using six folds (we used three splits, [1, 3, 5, 7] vs. [2, 4, 6, 8]; [1, 2, 5, 6] vs [3, 4, 7, 8] and [1, 2, 3, 4] vs [5, 6, 7, 8], for each split we train on one side and tested on the other; then reciprocally, and we average performance over all folds). Furthermore, we only included reference shapes in positions that could in principle have been deviant, to avoid the decoder picking on "early" versus "late" presented shapes.

Overall, this yields a very significant, slow, sustained effect visible in **Figure 2.3**: from 176 ms after stimulus presentation onward, the decoder's performance are significantly better than chance, with a peak performance obtained at 420 ms after stimulus onset, a performance of .57 (SE=.008), and an overall very significant cluster (p<.001; non-parametric temporal cluster-level paired t-test, (Maris & Oostenveld, 2007) as implemented by MNE). Note that the cluster detection method is not suited to accurately delimit onset and offset of clusters (Sassenhagen & Draschkow, 2019); however, it does indicate a strong temporal cluster which is very visible in the data in **Figure**

2.3.

Having trained such a decoder, we can separately test it for each shape. Given that participants, in an intruder task, find it harder to detect intruders within irregular shapes than within e.g. squares, we expect this to be reflected in the MEG signal: the signal for the square and its deviants should be different enough for the decoder to pick them apart, but the same should not hold true for the less regular shapes. Indeed, this is what we observe (see **Figure 2.4**) when we test the decoder[2]. The square and rectangle yield very high performance, and the performance slowly degrade as we test less regular shapes. Nevertheless, each shape is decoded better than chance at the p<.05, even the most irregular ones: the training, agnostic to the nature of the shape, picked regularities that are appropriate for all shapes, to different extents.

For each time point, we can compute the correlation between the 11 decoders' performance, and a metric derived from the number of symbolic properties: one minus the number of symbolic properties, rescaled from 0 to 1. This metric was chosen to keep simpler shapes on the left of the x-axis, in accordance with other figures correlating data with the error rates. This correlation is negative because shapes with more symbolic properties are easier to decode, and it remains significantly negative in a sustained way, yielding a long, very significant temporal cluster. To visualize the effect, we can also average the data across the entire significant cluster for each shape and participant, and plot a simple linear regression, visible in the rightmost column of **Figure 2.4**. This is mostly for visualization purpose and does not constitute an independent finding from the temporal cluster (otherwise this would constitute double-dipping).

A related but crucially independent analysis is to train **separate** decoders for the outlier of each shape separately. We report this analysis in the bottom row of **Figure 2.4**. The first things to observe is that the performance are much noisier: this is expected from the fact that the decoders are trained on less than a tenth of the data in this analysis. Again, the performance for detecting intruders in simple shapes exceed those of detecting them in complex shapes, and a gradient of complexity emerges. This yields a significant temporal cluster on the

---

[2]For this analysis and the next one, the epoched data was smoothed with a sliding window of 100ms for readability, but all analyses were replicated on unsmoothed data and identical results hold, with clusters broken up because of higher variability in the data.

correlation with the symbolic model, comparable in timing with the previous analyses. This decoder, unlike the previous one, fails to decode better than chance for four shapes: the isosceles trapezoid, the right hinge, the trapezoid and the irregular shape.

This last point deserves a remark: in principle, a decoder could pick up "oddball signal signatures" that are shape-independent, but expressed more or less strongly depending on the shape in the block. Conversely, they may latch on signatures of the oddball shape that are different for different shapes. What these two decoders suggests is that the first situation is at play in our experiment: indeed, the first decoder manages to perform better than chance even on the most irregular shapes, indicating that what its decision criterion, which was computed in a shape-agnostic way, is useful even for irregular shapes. On the other hand, when we split the training across shapes, it seems the data restricted to the irregular shapes is not enough to accurately find a detection criterion for oddball in the irregular shapes case, which is reflected in the poor performance of the second decoder for those shapes. Overall poorer performance can reflect the fact that each decoder is trained on $\frac{1}{11}^{th}$ of the data, but the decodability of the most irregular shape in one decoder and not the other shouldn't be impacted by this imbalance.

Representational Similarity Analysis

In order to test more specific theories of the mental representation of shapes, we performed Representational Similarity Analysis (RSA) of the data. For each pair of shapes and for each time point, we computed the cross-validated squared Mahalanobis distance (short: cross-nobis distance; here cross-validated over runs) following (Walther et al., 2016) to get a metric of how similar the brain activity evoked by each pair of shapes is. Intuitively, this can be seen as a more sophisticated way to measure the extent to which topographies from two different shapes are correlated over time.

For each time point, this gives us a dissimilarity matrix which we can model using the two models we adapted for the behavior data. After scaling all dissimilarity matrices, we can find the coefficients that minimize the whitened correlation between the data-derived matrix and a weighted sum of the two models: this gives us a temporal dynamic of the mental representation of shapes according to our two models. We report those coefficients in **Figure 2.5.A**, together with temporal

A
Coefficients of our two models of interest in a
GLM predicting crossnobis-derived confusion matrices

Neural Network

Symbolic

p<.001

p=.025

p=.043

Time

Stimulus Onset — 200 ms — 400 ms — 600 ms — Fixation Cross — Next Stimulus

B

Neural Network

Average coef. in GLM

Symbolic

0ms<t<284ms
p<.05

196ms<t<324ms
p<.05
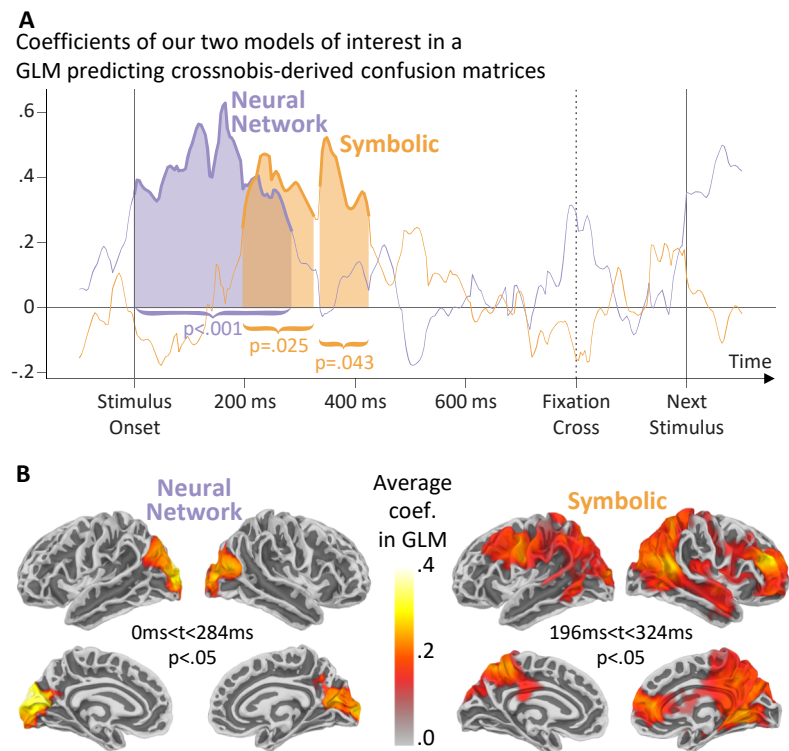
Figure 2.5: A. RSA in sensor space and temporal clusters. Time series of the coefficients that minimize the whitened correlation between the data-derived crossnobis dissimilarity across shapes and the dissimilarity derived from both a Neural Network model and a symbolic model described in **Figure 2.2**. Shaded areas indicate clusters derived with non-parametric permutation tests significant at the p<.05 level. **B. RSA in source space and spatial clusters.** Significant spatial clusters associated with each cluster detected in **A.** after performing a searchlight RSA in source space across shapes and modeling the resulting dissimilarity matrix as previously.

clusters derived by non-parametric permutation tests. There is a first significant cluster, from onset to about 280ms, for the neural network predictor: recall that the cluster method isn't meant to find precise onset and offset; in any case because the experiment was organized by blocks of identical shapes, so it is not entirely surprising that the onset is so early. No smoothing was applied here. There are also two clusters for the symbolic model, unfortunately separated unless thresholds are changed or soothing is applied even though they may look like they are the same cluster. These clusters start around 200ms and last until about 400ms.

The plot in **Figure 2.5.A** suggests that the correlation between the two models impacts the estimation of the coefficients, as the estimates appear negatively correlated over time; to make sure that this does not drive the effect we replicated this analysis but instead computed separately the whitened correlation distance between each model and the data and computed clusters on this distance. The results are essentially identical, with an early neural network cluster from onset to about 300ms, p<.05, and a single symbolic cluster that spans the two clusters identified with the first method, again with p<.05, see **Figure 2.6.A**

Instead of building clusters and statistics over participants, we also performed bootstrap over theories: 1000 times, we shuffled the dissimilarity matrix of the two models, and performed the regression with the brain data. Then, for each shuffling we averaged the coefficients across subjects, and for each time point we counted how many shufflings were above the predictions from the real model in order to derive a non-parametric *p*-value. This does not in and of itself provide clusters or correct for multiple comparison across time points, but it yields significant time points that fall within the clusters identifies previously and supports the same conclusions, see **Figure 2.6.B**

Finally, it's worth pointing out that very similar results were achieved by deriving the dissimilarity matrix from the data not with the crossnobis distance, but with other distances. More striking, similar results were obtained by deriving the similarity matrix by the confusion matrix of decoders, see **Figure 2.6.C**. To do this we trained independent decoders to discriminate all possible pair of shapes from the data, once again using cross-validation across runs, and used the generalizing score as a confusion value: when the score is close to chance, the shapes are hard to discriminate and have a low
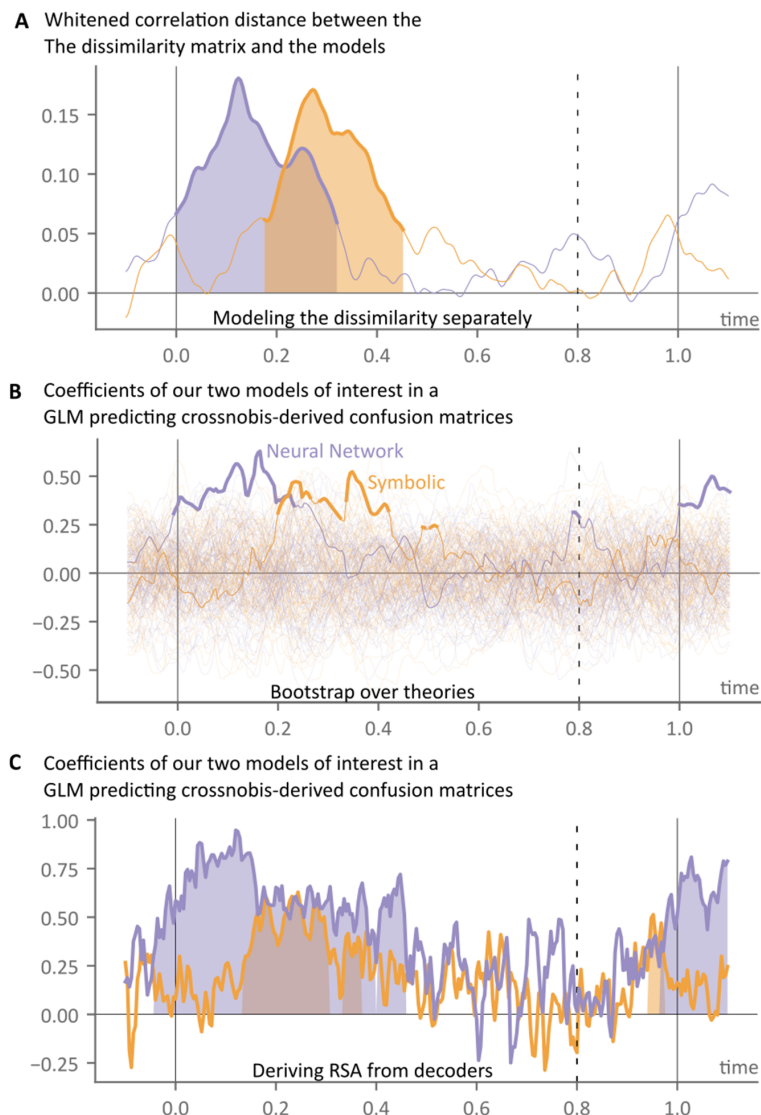
Figure 2.6: **A. Modeling the dissimilarity separately**; because the two models are not perfectly orthogonal, we can see that the estimation of coefficients makes the time series slightly negatively correlated in **Figure 2.5.A**; to confirm that this unintended correlation does not drive the effect, we also compute the whitened correlation distance between the models and the data separately and report it in this figure. **B. Bootstrap over theories** 1000 times we shuffled theories and performed the RSA analysis again. Each thin line represents a possible shuffled outcome averaged over subjects, the main colored lines indicate the reference, non-shuffled outcome, and is in bold when the unshuffled RSA analysis is over .95 shuffled analyses. **C. Using decoders to derive the RSA** Replication of **Figure 2.5.A** but where the RSA from the MEG data is derived using the confusion matrix of many one-versus-one decoders.

dissimilarity score; conversely shapes that are easy to tease apart for the decoder are considered to be highly dissimilar. This differs from the previous method because instead of the whole brain signal to be used equally, the decoder might find localized difference to use when discriminating between shape, and the location may be different for different shapes. Still, the results are comparable to those obtained with the crossnobis distance, but differ in the following ways: the neural network cluster starts earlier, around 50ms before the onset, and ends later at around 430ms, its coefficients reach as high as .9 indicating an excellent fit with the data. The symbolic cluster starts earlier around 130ms, and also ends earlier around 360ms.

The difference observed when deriving dissimilarity matrices from the decoder might reflect its ability to weight sensors differently, instead of correlating them all equally. As seen with the source-space RSA analysis, the two models are associated with spatially separated areas, and are presumably therefore associated with different sensors. At first sight the results are close to the straightforward RSA results, with a strong, early visual effect, which then reduces while the symbolic model picks up for about 250ms. However, with this analysis the visual model remains significant during this period, and even long after. One possibility is that as the shapes are perpetually "zooming in" on the screen, the visual areas are continuously re-stimulated – though less and less over time. The signal from the visual areas dampens when compared to the general signal over time, but the decoder might still be able to pick it up while it gets overshadowed by the symbolic model otherwise.

Can we find the source of these effects spatially on the cortex? To answer this, on possibility would have been to perform a similar RSA analysis as above, using searchlights on the source space. Instead, as we had access to high resolution anatomical MRI for 19 out of all participants, we opted for a more precise method and turned to source reconstruction. For each participant, each shape and each run, we performed source reconstruction over 1026 sources spaced on each hemisphere. Then, for each source, we took the set of sources less than 2cm away on the cortex's surface, and computed the dissimilarity between shapes, cross-validated over runs for the crossnobis distance. This gives us, in each participant's space, each source, and each time point, a dissimilarity matrix across shapes.

Then, for each source, each time point and each subject, we modeled

the dissimilarity matrix as a linear combination of our two models, by minimizing the whitened correlation distance. This gives use two spatio-temporal dataset per subject, one for each model. For each model, we can then project the spatio-temporal information into a shared, average subject cortical space. Computing spatio-temporal cluster on this data was very computationally expensive and prone to spurious results as the clusters ended up spanning over the entire duration of the epoch and moving all over the brain. To make the results more stable and interpretable, we removed the time dimension by averaging over the periods that were part of significant clusters in the strictly temporal analysis: this gave us one time period for the neural network, and two periods for the symbolic model, in which to search for significant spatial clusters.

The results are displayed **Figure 2.5**: for the neural network model's cluster, and for the first symbolic model's cluster, we found significant spatial cluster in source space at the p<.05 level. Note that despite the use of information from a previous analysis, this does not constitute double-diping, as evidenced by the fact that the second symbolic cluster did not yield a significant spatial cluster in source space. Instead, this allows us to answer the question: when a certain effect emerges, is spatially stable and well defined across subjects? And we find that it is indeed the case. The neural network model's two clusters correspond to bilateral occipital areas, while the symbolic model's two cluster span a much broader area, spanning over the dorsal pathways bilaterally, with a hotspot anterior on the middle temporal gyrus and another one on the right hemisphere around the middle frontal gyrus.

Discussion

The decoding results offer an interesting replication of the explicit intruder task of **chapter 1** in a purely passive shape-perception context. Since sitting in the MEG room for almost an hour looking at shapes without a task isn't engaging, it is quite likely that the structure of the oddball paradigm became apparent to subjects and they sought the oddballs proactively; nevertheless, this result indicates that our previous findings on visual search and intruder tasks were not restricted to the experimental conditions.

The fact that a decoder trained blindly on all shapes has better than chance performance even on the most irregular shapes indicates that it's possible to find a single dimension over which reference shapes
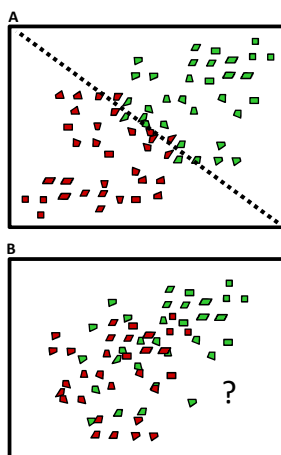
Figure 2.7: Illustration of how shapes (reference in green, deviants in red) can (in **A**) and cannot (in **B**) be organized around the decision boundary. In **A**, a shape and its deviant are organized symmetrically around an axis, which can be used as a shared decision boundary to discriminate oddballs across shapes. Given all shapes, finding the boundary is easy, but given only the most irregular shapes, finding it is hard. In **B**, shapes are organized directly along the geometric regularity axis. As a consequence, the deviants of the square are "more regular" than the reference of the irregular shape. There is no shared boundary for all shapes, but for each shape it should be possible to find the boundary – though it would still be harder for irregular shapes, as the translation along the axis becomes smaller. Our decoding results suggest an organization similar to **A.**

and oddballs differ. The existence of a gradient in performance as a function of geometric regularities suggests that the more irregular the shape, the closer it is to the cutoff boundary, making it harder to disentangle from an outlier: however, this suggests that the decision axis is not simply the number of geometric regularities of the shape presented, because deviants from the square feature more regularity than reference irregular shape. Therefore, the decoder must be able to find a dimension that correlates with the **change** in geometric regularities between the oddball and the surrounding shapes.

However, a decoder trained separately on the different shapes is not able to perform better than chance for the most irregular shapes, indicating that even though the boundary exists in the data, looking only at irregular does not allow for finding a decision boundary, possibly because more data would be required to find the boundary when the points to separate are close together. The overall gradient of geometric complexity remains, however, because the most regular shapes are well decoded and form a gradient. We illustrate what this result entail for the organization of geometric shapes in **Figure 2.7**.

The results from the RSA support the hypothesis of a two-strategy system of geometric shape perception which we posited in **chapter 1**: a

perceptual system, shared with other primates and possibly more generally, and a symbolic system specific to humans. We show that the dynamic of the brain signal across shapes starts by resembling the predictions made by the neural network model, which we have used as a proxy for bottom-up perceptual processing of the visual information. The resemblance kicks in as early as the stimulus onset, presumably helped by the block nature of the design, and lasts about 300ms. But at around 200ms, the MEG signal looks like an equal mixture of the two strategies, and from there onward the neural network models keeps going down while the symbolic model remains predictive, until around 450ms. The two strategies therefore appear to be separated in time, and follow a visual-symbolic order. Thanks to source reconstruction, we can also separate these two strategies in space: the visual strategy corresponds to a well localized, bilateral occipital cluster, while the symbolic strategy recruits a much broader network, encompassing the dorsal stream as well as frontal areas, again bilaterally.

This finding also provides theories for why human and baboons (and possibly other non-human primates) behave systematically differently. The two strategies build upon completely separated brain networks, both in time and space, and two origins for the behavioral divergence emerge. On the one hand, as baboons often try to answer much faster than humans typically do, they may use as much evidence as they have accumulated when they start answering but haven't waited enough for the symbolic elements to be computed, and therefore they only exploit the visual strategy. Another, more likely possibility is that crucial elements of the symbolic network are simply not present in baboons: either an ability for the detection of geometric properties, or an inhibitory effect on the early response, or other top-down action on the occipital cortex.

2.3.3 Infants

The results presented here are preliminary and require more analysis, and possibly more subjects, to be fully informative with regard to the questions at hand. Nonetheless, they are worth presenting as they help build a more comprehensive picture of the universality of perception of geometric shapes in humans.
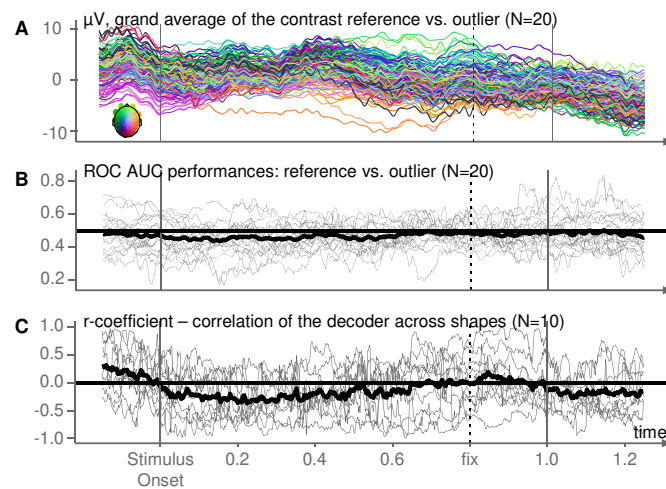
Figure 2.8: Decoding results in the first group. **A.** Average ERPs across subjects of the contrast reference vs. oddball. No significant spatio-temporal cluster is present in this data. No significant spatio-temporal cluster was found with non-parametric permutation tests. **B.** Performance of a decoder discriminating reference versus oddball stimuli. Each gray line is a participant (N=20), and the black line is the average performance. No significant spatial cluster was found for non-parametric permutation tests. **C.** Correlation between the performance of the decoder tested separately on each shape, and the behavioral gradient of complexity experimentally measured in **chapter 1**. Only participants with data for enough shapes are analyzed here: we show the data for subjects who have seen at least 5 different shapes (N=10), and similar analysis have been conducted for other threshold with identical results.

First Group Only

In trying to replicate the MEG results obtained in adults, we first trained an identical decoder to discriminate between reference and oddballs, with two folds to measure generalization score while trying to maximize the training data available to the decoder. This decoder does not perform better than chance as no significant time cluster was found using non-parametric permutation tests over all subjects: the performance for each subject, as well as the average, are shown in **Figure 2.8.B**.

Since the decoder still fluctuates around chance, we also tried to test it separately for each shape, to see if the performance were different across all shapes as we have shown to be the case in adults. Because not all subjects have seen all 6 shapes, we tried restricting the data to cases where participants have seen at least K shapes for K in [2,3,4,5,6]; however, in all cases, the correlation across shape for each time point did not yield a significant cluster: **Figure 2.8.C** shows the results for K=5 which leads to 10 subjects. Often decoders are provided with data only minimally pre-processed, however we also tried providing the decoder with reference averaged, baselined ([-150, 0]ms window) and

z-scored data, the results were very similar.

We also tried to see if we could detect a difference in the evoked response potential (ERP) from reference and oddball stimuli by looking at the grand average of the within-participant differences, without using a decoder. That average is presented in **Figure 2.8.A**, and there are no significant spatio-temporal cluster. For this analysis too, we tried to baseline, reference average and z-score the data, but the results are identical.

This could be due to a number of reasons, including theoretical ones; however, two methodological considerations are worth looking into. First, the blocks were 60 shapes long, once per second, plus delays introduced by occasional breaks: there are many babies for whom we don't have data for all shapes, making correlations even more hazardous. Second, we often introduced breaks in the experiment, either by switching to the face stimuli or with the soothing visual and music distractors: this introduces breaks in the oddball paradigms, that could lead to oddballs being, in fact, rather acceptable when compared to other possible future stimuli. Both these design shortcomings are due to an overly optimistic design, and lead to the second iteration of the design described in the procedure section: deviants were removed, blocks were much shorter, and we didn't alternate between shapes and faces anymore.

Both Groups

With both groups, it is not possible to analyze oddball anymore as they were removed in the second group. However, we can still correlate ERPs with our theoretical proposition of geometrical regularities for the shape, and we can perform RSA analysis across different shapes.

Preliminary results from both groups are presented in **Figure 2.9**. For the analyses in this section, a few additional preprocessing steps were performed: (i) the data was baselined during the [-150,0]ms window, (ii) the data was reference-averaged, and (iii) each epoch was centered and scaled, to avoid some participants dominating the data in some analyses.

First, we performed a linear regression of the average ERP associated with each geometric shape with the theoretical prediction of complexity, i.e. the number of symbolic properties, for each sensor and each subject. The time courses of the intercept and the coefficient are pro-
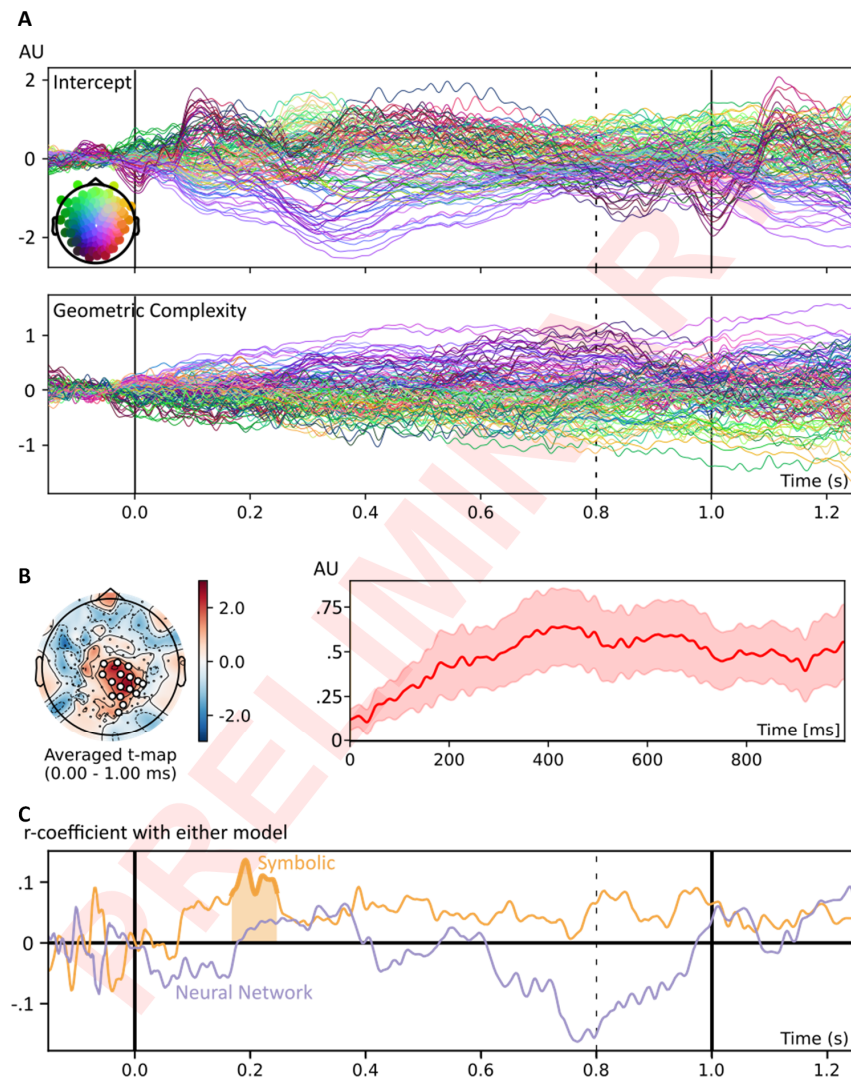
Figure 2.9: **Preliminary results with EEG in infants. A.** Grand average of the intercept (top) and the beta (bottom) in a linear regression of the geometric complexity of the shapes for each sensor, time point and subject. **B.** Significant spatio-temporal cluster (p=.039) identified across babies for the geometric complexity regressor. Its topography is on the left, and the time course on the right. Note that this cluster spans the entire duration of the epoch (0–1s). **C.** Time courses and significant temporal cluster (p<.05) in the modeling of the dissimilarity matrix with our neural network and symbolic model.

115

vided in **Figure 2.9.A**. The intercept corresponds loosely to the grand average of the average across all stimuli together: it's a cleaned version of the average signal of what "seeing any shape" elicit in participants. While the time courses for the geometric complexity seems to mainly diverge, distancing from the baseline over time without much structure, some effects were still systematic across subjects to be picked up by a non-parametric spatio-temporal clustering analysis (p-value to enter a cluster: p<.05; number of bootstrapping folds 2^{10}, search window: [0,1]s, final p-value of the identified cluster: p=.039), presented in **Figure 2.9.B**. They correspond to a group of central electrodes, and despite the apparent systematic divergence over time, (i) that divergence was systematic enough across subjects for the cluster to be significant, and (ii) a small return to baseline can be observed from 500ms to 800ms, indicating that the drift is not simply due to accumulating noise. A very similar cluster can be found when looking only at the first group of infants (p<.049) but it does not come out as significant in the second group alone.

Finally, we replicated the RSA performed in MEG in adults: for each pair of shapes, each time point and each subject, we computed a dissimilarity measure (for this analysis we used the "correlation" distance: one minus the correlation of the topographies). Then for each subject we measured the correlation between this dissimilarity measure and the two models: the neural network model and the symbolic model. Average across subjects of the time course are provided in **Figure 2.9.C**, together with significant temporal clusters (non-parametric permutation tests, one tailed, over the [100, 500]ms window, to try to replicate the clusters found in adults). The two time courses are much noisier than in adults in MEG, and a significant cluster for the symbolic model barely comes out as significant, between 168ms and 246ms, p=.046; no cluster found for the neural network model.

Discussion

The results in infants are not really conclusive yet, but the results presented here are promising: there seem to be an effect of the shape presented on the ERP, and the geometric complexity seem to indeed correlate with some electrodes over time. Additionally, the RSA performed provide a hint that the shapes are represented in a way that significantly correlates with the symbolic model during a period compatible

with expectations from the adults MEG data.

However, each of these results suffers from instability. First, the tests that can be performed separately on either group don't replicate straightforwardly. Second, while we find a reasonable temporal cluster for the symbolic model, it is a bit surprising that we don't find one for the neural network model, which we might expect to be part of bottom-up perception and therefore more precocious. Finally, while the choice of dissimilarity measure for RSA in adults' MEG data did not really impact the result beyond few milliseconds shifts in the resulting significant clusters, the same does not hold true for the EEG data in babies, where choosing metric impacts greatly the resulting time courses.

More generally, the two groups of infants present slightly different overall responses (e.g. the grand average of all shapes), despite the stimuli being identical. For example, without applying a baseline it is possible to use the fixation cross to identify a supposedly constant visual response across shapes: doing so yield close but no-overlapping set of electrodes between the two groups. Current analyses do not model these differences, but it appears possible that the changes introduced between the two experiments impacted the responses more than anticipated. The absence of oddballs, in particular, makes the task introspectively different for adults, since after a few presentations they know what shapes are coming next for a consistent period of about 15 shapes, and unless paying special attention they lose interest.

In ongoing work, we are reusing analyses performed in adults' MEG by using a decoder to construct the dissimilarity matrix instead of the current correlation measure, to make the RSA more sensitive. Future work will most likely require more data to stabilize the current results. In particular, better understanding the differences between the two groups should help adjudicate between the two versions of the experiments, in order to maximize the informativeness of future data collection.

2.4 General Discussion

The evidence presented with these new experiments support our theoretical claims from **chapter 1**: in humans, geometric shapes are per-

ceived using a symbolic strategy, which relies on exact properties and coexists with an underlying perceptual strategy. The behavioral data is both a replication of the predictive power of the symbolic model versus the perceptual model in explicit, intentional decision-making; furthermore, using multi-dimensional scaling, we get purely data-derived dominant characteristics of the behavior, which we can then confirm to map onto our theoretically motivated symbolic model.

The MEG data makes it clear that deviants of regular shapes are more easily detected than those of irregular shapes, even in the absence of a task. This is yet another replication of the main effect I shed light on in **chapter 1**, perhaps unsurprising given its effect size, but certainly welcome. More importantly, the data lets us explore the spatio-temporal dynamic of the phenomenon, by contrasting two confusion matrices: one stemming from perceptual properties of the shapes, and the other from symbolic properties. We confirm the hypothesis that in humans, at least two strategies coexist for tacking tasks involving recognizing geometric shapes, and additionally we separate them both temporally (the perceptual precedes the symbolic) and spatially (mostly occipital versus dorso-frontal).

In trying to replicate these analyses, there was a hope to determine whether infants already display symbolic processing of the geometric shapes, to further understand how universal geometric shape perception is in humans. The results we get, however, are on the fence about this particular question, and additional analyses, and possibly additional data, will be required to adjudicate. Still, results so far are promising: different shapes do seem to systematically produce different ERPs, and we find a significant cluster when we analyze those ERPs as a function of geometric shape regularity for a cluster of central electrodes. Furthermore, the RSA suggests that the mental representation of shapes is compatible with the symbolic model during a period comparable to what was identified in adults, although the same result does not hold true for the neural network model.

Infants can be habituated to simple geometric figures (Schwartz et al., 1979) and even abstract over global rotation changes. (Schwartz et al., 1979) provide evidence that infants are more attentive to local angles composing a shape than to its global orientation, and that they might use local information about angles to encode the entire shape but do not seem to process higher-order relationships between segments. This is a limiting factor for experiments involving quadrilater-

als; however, in the case of our shapes even processing local angles alone should already give rise to the gradient of geometric complexity observed in adults.

On the other hand, using preferential looking, (Dillon et al., 2020) shows that 7-month-old infants detect changes in shape using triangles, with a very fast presentation rhythm. But that study finds that the dominant factor appears to be the relative lengths rather than the angles, at least when abstracting over rotation and scale is required. An important argument put forward for this difference is that angles, as a relation between directions, might be computed more slowly than lengths, pushing its computation beyond the presentation threshold of that experiment. Could the same phenomenon occur with our geometric shapes, where relations between relations need to be computed, making the presentation time too fast for infants?

2.5 Conclusion

In this work, we decompose the temporal, and to a lesser extent the spatial, dynamic unfolding of simple geometric shape perception in humans. Using a set of minimally different quadrilateral shapes, with a varying number of geometric properties, we identify a double dissociation when perceiving geometric shapes which we had posited from earlier modeling of the behavior.

This provides more insight as to why non-human primates may lack the symbolic strategy. Indeed, the strategy associated with the neural network, which was a good model of all baboons, correlates with an early occipital response, which contains areas that have close functional homologous across non-human primates (Orban et al., 2004; Tootell et al., 1996). On the other hand, the strategy associated with symbolic processing spans a very broad network of the brain, including strong dorsal and frontal bilateral activations. The specificity of source reconstruction appears to limit accurate localization of the effects. However, the areas identified make sense: while the historical view is that the prefrontal cortex is disproportionately large in humans when compared to other non-human primates (Deacon, 1998; McBride et al., 1999) is challenged from a methodological standpoint (Semendeferi et al., 2002; Sherwood & Smaers, 2013), there is no doubt that it is at the center of many neuroscientific inquiries when contrasting humans versus other animals.

The temporal dissociation is important as well: baboons tested in **chapter 1** tend to answer faster than humans, a general strategy they employ to maximize reward through the number of trials if not their success. Given that some perceptual strategy is enough to do marginally better than chance – though not by much – baboons may reply as soon as the visual information is available to reply, which would over-emphasize the reliance on the visual strategy that is available sooner. Assuming visual timing consistent with humans, this wouldn't quite account for the data, as the average response time of the baboons averaged at 1600ms, therefore arriving much later than the peak of the predictiveness of the neural network strategy at around 180ms.

In summary, the present results strengthen the findings that there are two strategies for geometric shape perception and sheds light on the neural mechanisms at the root of this dissociation. They hint at plausible accounts of what drives the divergence between human and non-human primates while not fully explaining it. The next chapter will focus on a more accurate localization of the effect discovered here, using fMRI in both adults and children. Future work should explore recordings of the brain activity of primates undergoing a similar oddball paradigm

Chapter 3

Geometric shape perception activates brain regions associated to mathematical cognition: an fMRI study

Abstract

Is geometry closer to mathematics, language, or general vision? In this work, we provide elements to answer this question by using functional MRI in adults and children to look at the brain networks involved in two tasks involving geometric shapes. We find that passively viewing geometric shapes, when compared to other visual categories, over-activates areas typically associated with number cognition and mathematical cognition, while under-activating the ventral stream of visual information processing. We also find that a vast brain network increases in activity together with the difficulty of an intruder detection task using geometric shapes, but only in adults and when the difficulty is moderate; furthermore, we find that the organization of the mental representation of the different shapes can be separately modeled using a symbolic model and a neural network model of object recognition, with areas associated with either separately, and to both.

> "I.Def.1: A point is that which has no part
> I.Def.2: A line is breadthless length.
> I.Df.3: The ends of a line are points."

<div align="right">Euclid's Elements</div>

3.1 Introduction

No specie other than humans has gained access to abstract, mathematical concepts. The origins of this ability are much debated, but recent work (Amalric & Dehaene, 2016) has shown a dissociation of high-level mathematics from typical language networks, suggesting a distinct set of neuronal populations that can deal with the concepts required for mathematics.

Is mathematics a byproduct of the capacity for language – a hypothesis introspectively disfavored by mathematicians and physicists? This hypothesis has received indirect empirical support in several ways: notably because language plays a significant role in the development of spatial reasoning (Pyers et al., 2010), and because deaf individuals who did not learn a conventional language have a harder time manipulating exact numbers (Hyde et al., 2011; Spaepen et al., 2011, 2013).

On the other hand, recent neuroimaging studies find a systematic dissociation between language networks and math networks (Amalric & Dehaene, 2016, 2017). In the specific case of geometry, in addition to the results presented in **chapter 1** which suggest that natural language is not the main constituent of the geometric regularity effect, it has been shown that even infants and adults without formal Western education possess intuitions about number, space, and geometry (Dehaene et al., 2006; Pica et al., 2004). Furthermore, infants have been shown to be able to solve geometric tasks without being able to provide a justification for their decisions, either through language or gesture (Calero et al., 2019)

A possible reconciliation of this tension could come from language as a mechanism to integrate different "core knowledge" together, an idea proposed in (E. S. Spelke, 2003). Under that view, core knowledge domains would provide the essential building blocks from which humans could construct more and more complex representation using language as construction tools, building up through analogies and inductive processes (Hofstadter & Sander, 2013; Lakoff & Núñez, 2000)

that rely on the ability to represent different objects with a shared medium: natural language.

Using both an fMRI version of our intruder task introduced in **chapter 1**, with quadrilateral shapes, and new geometric stimuli in a standard functional localizer, we set out to investigate whether simple geometric shapes, either passively perceived or involved in an active discrimination task, activate networks associated primarily to either mathematics, language, or vision. Furthermore, we tried to understand whether this association changed with education by comparing adults and kindergartners.

3.2 Method

3.2.1 Participants

Twenty healthy French adults (9 females; 19-37 years old, mean: 24.6 years old, SD: 5.2 years old) and 25 French first graders (13 females; all were 6 years old) participated in the study. Three children quit the experiment before the formal task began because they did not like the scanning noise or lying in the confined space. All participants had normal hearing, normal or corrected-to-normal vision, and no neurological deficit. All adults and guardians of children provided informed consent, and adult participants were compensated for their participation.

3.2.2 Materials

Category Localizer

Grayscale images of eight categories were presented to the participants (**Figure 3.1**). The categories included faces, houses, tools, single-digit formula, French words, Chinese characters, single geometric shapes, and three geometric shapes displayed in a row. Each category had 20 exemplars. All faces, 16 houses, and 18 tools have been used in previous localizer experiments (Zhan et al., 2018).

For face stimuli, front-view neutral faces (20 identities, 10 males) were selected from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman). The stimuli were aligned by the eyes and the iris distances. A circular mask was applied to exclude the hair and clothing below the neck. House and tool stimuli were royalty-free
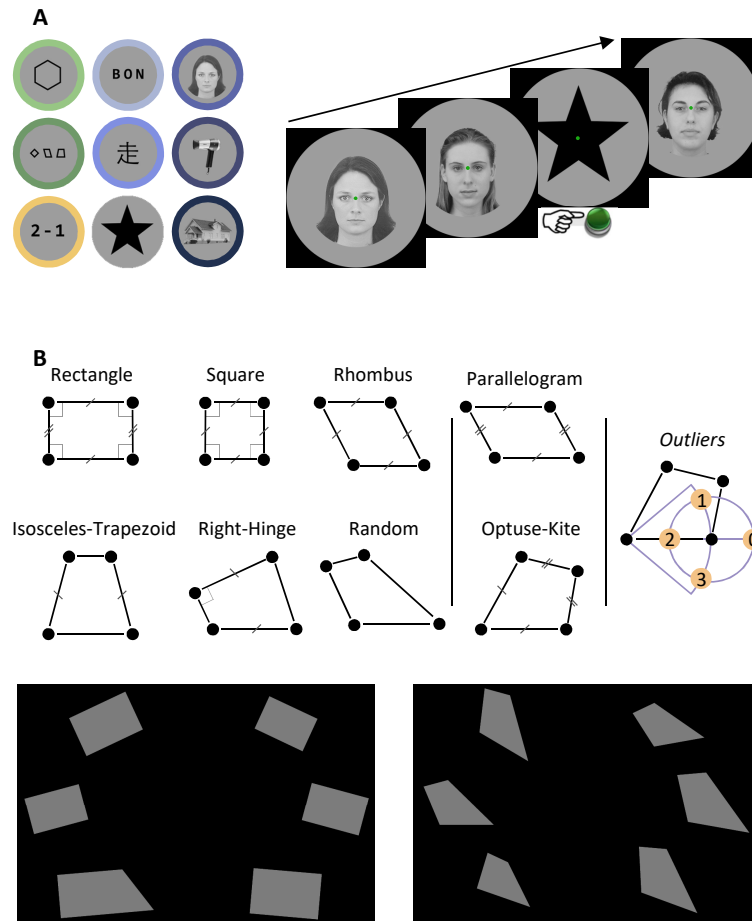
Figure 3.1: **A. Stimuli and method of the category localizer.** On the left are the 8 categories used in the localizer, as well as the start participants had to detect. An example of a mini-block is shown on the right, with faces and a start inserted to be detected. **B. Stimuli and method of the geometry task.** On the top we display the 8 possible shapes used in the geometry task, as well as an example of the way outliers were devised. Below we show to trial examples, in both cases the outlier shape is on the lower-left position (participants had to answer "left"), one with the "rectangle" shape and one with the "random" shape.

images obtained from the internet. House stimuli were photos of 2 to 3-story residence houses. Tool stimuli were photos of daily hand-held tools. Half of the images were horizontally, so that there were 10 images in a position graspable for the left and right hand respectively.

For French word stimuli, 3-letter French words were selected which were known to first graders and had high occurrence frequencies (range=7-2146, mean=302, SD=505 occurrences per million based on Lexique, http://www.lexique.org/). Chinese characters were selected from the school text book of Chinese first graders. The word frequency (range=11-1945, mean=326, SD=451 occurrences per million; (Cai & Brysbaert, 2010)) was not significantly different from French words used here (t(38)=0.2, p=0.87).

Single-digit formula stimuli were 3-character simple operations in the form of "x±y" with x greater than y, x ranging from 2 to 5, and y from 1 to 4. Single shapes consisted in a single, centered outline of a geometrical shape, either circles, quadrilaterals or triangles, and were comparable with faces/houses/tools/Chinese characters which also displayed single objects. A row of shapes consisted of three different shapes side by side whose total width, size, and line width were adjusted to match other stimuli size, and were supposed to be comparable with 3-letter French words and 3-character single-digit operations.

To match the appearance of the monospaced font in (Vinckier et al., 2007), the monospaced font Consolas was used for the French words and numbers, with identical font weight 900. The font for Chinese characters were Heiti, which looks similar to Consolas. Random font size was used to achieve similar variability as with the other categories. The stimuli were embedded in a gray circle (RGB color=157, 157, 157, radius=155 pixels), on the screen with a black background. Within the gray circle, the mean luminance and contrast of the 8 stimuli categories were controlled not to be significantly different from each other (luminance: F(7,152)=0.6, p=0.749; contrast: F(7,152)=1.2, p=0.317, see **Figure 3.2**).

Geometry Task

Geometric shapes were generated following the procedure described in (Sablé-Meyer, Fagot, et al., 2021). To fit the experiment in the time constraints, a subset of shapes was used: the square, rectangle, isosceles trapezoid, rhombus, right hinge and random for the easy condition,
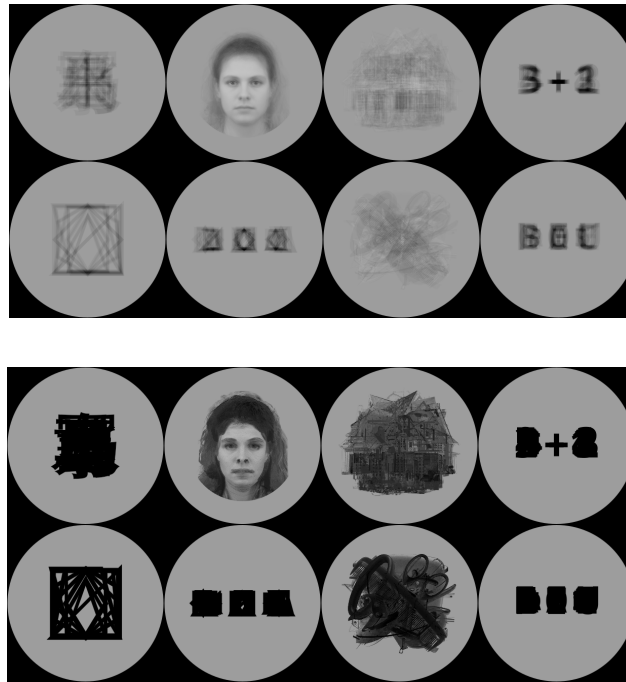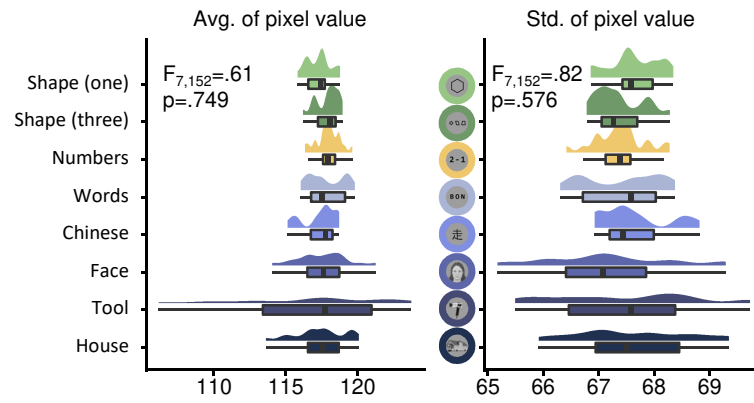
Figure 3.2: A. Controls on the luminance and contrasts of the localizer stimuli. Top: average distribution of means and std pixel values across categories in the localizer. Bottom: average stimuli in each category, and superposed stimuli in each category.

with additionally the parallelogram and the obtuse kite for the hard condition (See **Figure 3.1**). Following (Sablé-Meyer, Fagot, et al., 2021) deviants were generated by displacing the bottom right corner by a constant distance in four possible positions: that distance was a fraction of the average of all pairs of points, which was standardized across shapes, and was either .45 for the easy condition, or .3 which was used in (Sablé-Meyer, Fagot, et al., 2021) for the hard condition.

In each trial, six gray-on-black (shape color rgb values: (127,127,127)), and shapes were displayed in two semicircles. The positions were determined by positioning the three leftmost (resp. rightmost) shapes on the left side (resp. right side) of a circle of radius 120px, at angles 0, pi/2 and pi, and then shifting them 100px to the left (resp. right). The rotation and scaling of each shape were randomized so that no two shapes had the same scaling or rotation factor, and values were sampled in [0.875, 0.925, 0.975, 1.025, 1.075, 1.125] for scaling and [-25°, -15°, -5°, 5°, 15°, 25°] for rotations, avoiding 0° to prevent excessive alignment of specific shapes to screen borders. One of the shapes was an outlier, whose position was sampled randomly and uniformly in all six possible positions, but so that no two consecutive trials featured outliers in the same position. Outliers were sampled in the four possible types of outliers, so that all outliers occurred as often but no two consecutive trials featured identical outlier type.

3.2.3 Experimental Design and Procedure

Category Localizer

The eight categories were presented in blocks of 6s, fully randomized for block presentation order, with the restriction that there were no consecutive blocks from the same category. Stimuli presentation order within a block were also randomized. Each stimulus was presented for 1s, with no interval in between (Dehaene-Lambertz et al., 2018). The inter-block interval was jittered for 4, 6, or 8s (mean = 6s). Each of the eight block types was presented for twice within each run. Participants were asked to keep fixating on a green central fixation dot (radius=8pixels, RGB color=26, 167, 19, always shown on the screen), and to press the right-hand button when a star symbol was presented. The star spanned roughly the same visual angle as the eight categories, and were randomly presented once in one of the two blocks per category, between the 3rd to the 6th stimuli within that block. Three hundred

milliseconds of 650 Hz Beep sound was provided as the audio feedback associated with the button press. A 6s fixation period was included at both the beginning and the end of the run. Each run lasted for 3min 24s, and participants performed three runs of this task in the fMRI session.

Geometry Task

The shapes were presented in blocks, fully randomized for block presentation order with no two consecutive blocks with the same type of shape. Runs of the easy condition always lasted 3m40s, and blocks of the hard condition always lasted 6m58. In the easy condition, blocks of 5 identical base shapes were used in a row, with 2s of presentation, 2s to answer. In the hard condition, shapes were flashed for 200 ms, with 4s to answer, and a 4s mini-pause mid-block. In both conditions there was a 4s, 6s or 8s delay between blocks. A central green fixation cross was always on display, and it turned bold 600 ms before a block would start. After each answer auditory feedback was provided.

3.2.4   MRI Acquisition

MRI images were acquired on a 3T Siemens Prisma scanner. Functional EPI images were scanned covering whole brain with following parameters: TR=1.81s, TE=30.4ms, field of view=864x864, flip angle=71 degrees, resolution=2x2x2mm, multiband factor=3, phase encoding direction: posterior to anterior. Anatomical T1 weighted images were acquired using following parameters: TR=2.3s, TE=2.98ms, field of view=240x256, flip angle=9 degrees, resolution=1x1x1mm. In addition, spin-echo field maps were collected. Single-band reference images were also collected before each functional run. Each fMRI session lasted for around 50 min for children including (by order) runs of a task not discussed in the present article, 3 Category localizer runs, T1 collection, and 2 easy geometry tasks. For adults, the session lasted for around 1h 20min with the same runs as for children, as well as 2 additional hard geometry task.

3.2.5   Imaging Data Preprocessing

The following section was automatically generated by fMRIPrep. Results included in this manuscript come from preprocessing

performed using fMRIPrep 20.0.5 (Esteban, Blair, et al., 2018; Esteban, Markiewicz, et al., 2018), which is based on Nipype 1.4.2 (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2018).

Anatomical Data Preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the `ants-BrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9, Zhang et al., 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Volume-based spatial normalization to two standard spaces (MNI152NLin6Asym, MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model (Evans et al., 2012) and ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009).

Functional Data Preprocessing

For each of the 10 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Susceptibility distortion correction (SDC) was omitted. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated be-

129

fore any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected using `3dTshift` from AFNI 20160207 (Cox & Hyde, 1997). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time-series were resampled into several standard spaces, correspondingly generating the following spatially-normalized, pre-processed BOLD runs: MNI152NLin6Asym, MNI152NLin2009cAsym. First, a reference volume and its skull-stripped version were gener-ated using a custom methodology of fMRIPrep. Several confounding time-series were calculated based on the preprocessed BOLD: frame-wise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise cor-rection (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tComp-Cor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calcu-lated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal

derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.6.2 (Abraham et al., 2014), mostly within the functional processing workflow.

3.3   Results

For all runs, analyses were performed using nilearn version 0.9.0 (Abraham et al., 2014). GLM were fitted with polynomial drift orders (up to degree 5), high pass at .01 Hz, smoothing at 4mm, nilearn's implementation of the SPM Hemodynamic Response Function, a second order autoregressive noise model, and signal was scaled to percent of signal change (time series are shifted to zero mean value and scaled to percent signal change). No slice time correction was applied as this step was included in the the preprocessing pipeline.

3.3.1   Category Localizer

Results

In the category localizer, the GLM design matrix featured a predictor for each stimulus in the 8 categories coded as a 1s event, plus a predictor for the star, 1s event, and a predictor for button presses as 200 ms events, as well as a constant plus five polynomial drifts. From fmriprep we additionally used three translation predictors and three rotation predictors, a cerebrospinal fluid and a white-matter predictor. Finally, we added five predictors associated with the highest percentile of variance in the data, following nilearn's implementation of CompCor (Behzadi et al., 2007).

Clusters were identified by thresholding contrasts of the second level

131
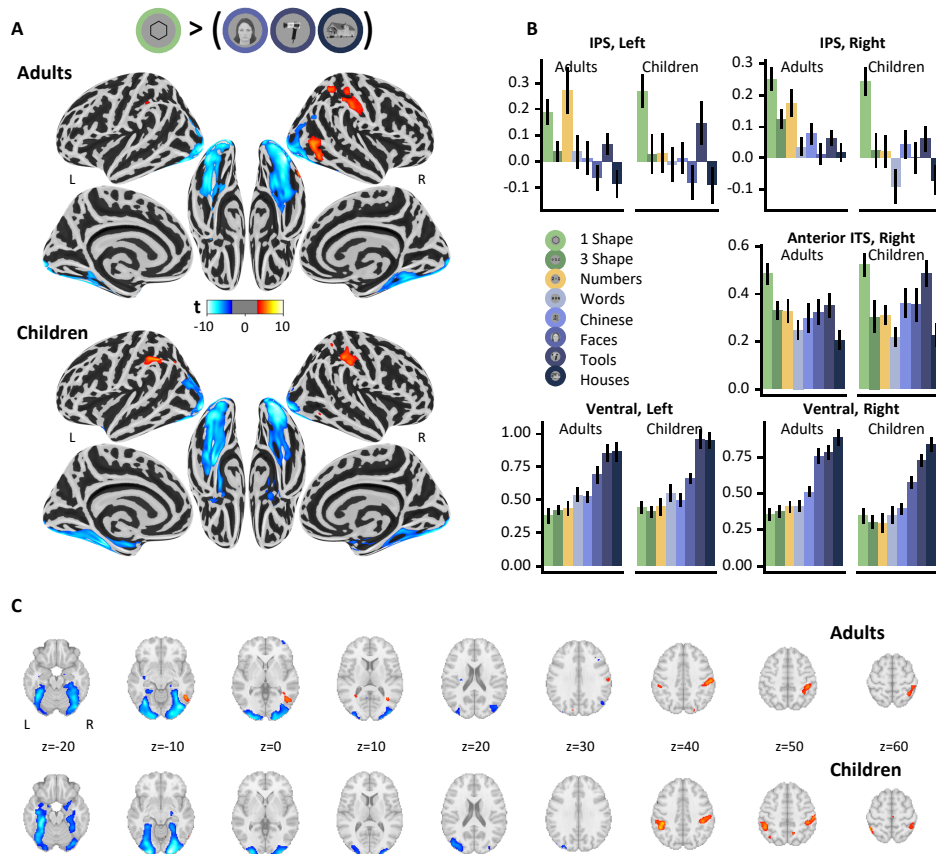
Figure 3.3: First level contrast: shape > single element. **A Thresholded contrast map projected on surface average.** For both age groups, we display the second level, 8mm smoothed contrast map associated with the "shape alone" > 1/3 "face + house + object", thresholded so that p<.0001 uncorrected. **B. GLM coefficients per cluster** for each cluster in the adult group (see Table 1), we display the coefficients associated with the regression for both age groups, with error bars corresponding to standard error of the mean. Colors indicate the category, and the number correspond to **A** and **Table 3.1**. Note that while in **Table 3.1**, significant clusters were selected and evaluated within age group, i.e. "cluster 1" is slightly different in both groups, in this plot the clusters from the adults were used. **C. Thresholded contrast map as in A**, with equidistant spaces on the z-axis in MNI space.

(smoothed at 8mm following (Mikl et al., 2008), p<.0001 uncorrected). The statistical significance of clusters was then assessed by bootstrap: for 10k iterations, the sign of the first level contrast map of each participant was randomly swapped and the weight of the biggest cluster for this particular shuffling was recorded (where the weight is the sum of the t-values of the cluster). From this we derived p-values of each cluster from the real data by comparing its weight to the distribution of random clusters.

First, we study the contrast "shape alone > $\frac{1}{3}\times$(face + house + objects)", where we compare the processing of single shapes with the processing of other categories featuring a single item. **Figure 3.3** shows the thresholded t-map for both age groups, and in **Table 3.1** we report significant clusters identified by bootstrap. For both age groups, the two most dominant clusters are symmetrical ventral pathways. Both reach significance through bootstrapping, and yield negative statistics, indicating that shapes are dominated by the other categories. There are also symmetrical, positive clusters in both age groups inside the Intra Parietal Sulcus (IPS): in the right hemisphere it reaches significance in bootstrap for both age groups, but only reaches significance for 6 years old in the left hemisphere. Finally, a cluster was found in both age groups in the right Inferior Temporal Sulcus, although it reached significance only for the adult population. Interestingly, in the left hemisphere, the IPS cluster was also strongly activated by our number stimuli in adults (but not in 6 years old), a feature partly observed in the right hemisphere IPS as well, but not in the more ventral cluster (see Figure 2B). The contrast "Single shape + Three shapes > $\frac{1}{2}\times$(face + house + tool + text)" yields very similar results.

Table 3.1: **Location, strength and significance of clusters** for both age groups for the contrast shape alone > $\frac{1}{3}$(face + house + object). Locations are given in MNI coordinates. We report all 5 clusters identified as significant in either group. Lists of clusters from each age group didn't perfectly overlap but there was always a straightforward mapping from one to the other. P-values derived by bootstrap. Cluster ID corresponds to the numbers used in **Figure 3.3**, with subpeaks and associated t-values identified within each cluster.

Cluster	Sub peak	MNI (x, y, z)	Peak t	Size (mm^3)	Weight	p-value
1	a	(-50.5, -34.5, 41.5)	4.56	424	220	.517
2	a	(35.5, -48.5, 55.5)	6.68	7112	4110	.0044
	b	(59.5, -22.5, 39.5)	6.11			
	c	(43.5, -34.5, 47.5)	6.05			
	d	(51.5, -26.5, 41.5)	5.94			
3	a	(51.5, -56.5, -6.5)	8.38	3840	2370	.013
	b	(45.5, -64.5, 3.5)	6.8			
	c	(37.5, -58.5, 7.5)	5.87			

133

Cluster	Sub peak	MNI (x, y, z)	Peak t	Size (mm^3)	Weight	p-value
	d	(39.5, -48.5, -0.5)	4.8			
4	a	(-24.5, -104.5, -4.5)	-8.99	38224	-25500	<.0001
	b	(-34.5, -68.5, -10.5)	-8.66			
	c	(-38.5, -54.5, -18.5)	-7.71			
	d	(-40.5, -74.5, -14.5)	-7.41			
5	a	(41.5, -76.5, -16.5)	-9.53	55056	-37600	<.0001
	b	(33.5, -60.5, -14.5)	-8.26			
	c	(17.5, -106.5, -4.5)	-8.05			
	d	(23.5, -90.5, -10.5)	-7.82			
1	a	(-48.5, -48.5, 43.5)	7.86	8232	5060	.005
	b	(-44.5, -38.5, 41.5)	6.81			
	c	(-48.5, -46.5, 57.5)	6.44			
2	a	(43.5, -34.5, 45.5)	6.35	7104	4100	.0078
	b	(55.5, -24.5, 45.5)	6.16			
	c	(47.5, -26.5, 37.5)	5.29			
	d	(41.5, -44.5, 59.5)	4.52			
3	b	(55.5, -64.5, -4.5)	4.63	472	245	.529
4	a	(-26.5, -98.5, -8.5)	-9.66	49888	-32600	<.0001
	b	(-34.5, -36.5, -22.5)	-9.01			
	c	(-40.5, -68.5, -16.5)	-8.27			
	d	(-32.5, -78.5, -12.5)	-8.06			
5	a	(21.5, -96.5, -8.5)	-10.6	36120	-24600	<.0001
	b	(27.5, -84.5, -14.5)	-8.69			
	c	(23.5, -46.5, -12.5)	-6.41			
	d	(35.5, -44.5, -22.5)	-6.11			

Considering now the contrast "Three shapes > words", we find three left lateralized, word-dominant significant clusters in adults, significant with bootstrap. All three are fully absent in 6 years old. We find again a right-lateralized IPS cluster in adults, at a similar location as with the previous contrast, but it is much smaller (peak t-value 4.63, size 112mm3, p=.84) and it is absent in 6 years old.

To understand better the ventral pathway situation, we computed four additional contrasts by pitting each category of face/house/tool/words against the three others. We could check that in almost all of the resulting clusters, there was a positive activation for shapes, but it was vastly dominated by the other visual categories (see **Figure 3.4**).

Discussion

Taken together, these results show that geometric shapes activate the typical ventral visual pathways less than other common visual categories (but activate them nonetheless), while activating a bilateral inferior parietal area more than the other categories. This result is true across age groups, although two details are different: in adults, the parietal activation is asymmetric, with a stronger activation on
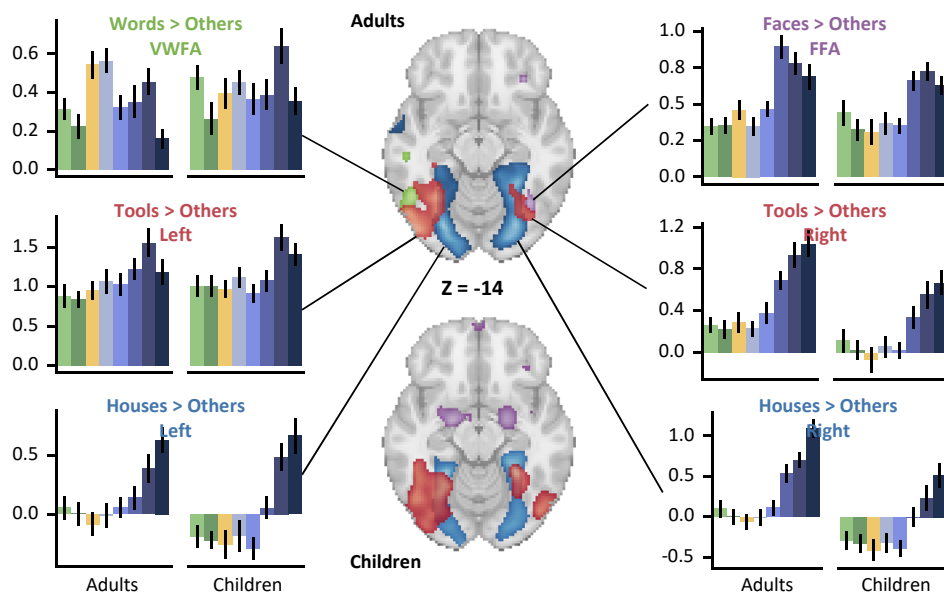
Figure 3.4: Inspecting a slice at z=-14, with areas positively activated with the four contrasts obtained by comparing face, house, tool or text against the three other categories, in both adults and six years old, thresholded (uncorrected) at p<.0001, together with the average coefficients of the first level model at the peak voxel of each category, in both hemispheres when applicable.

the right, while in children the activation is mostly symmetrical. Conversely, in adults there is a significant cluster on the right inferior temporal gyrus which is absent in children.

The results in adults coincide well with the findings in (Amalric & Dehaene, 2016) which found bilateral clusters at identical locations for a math-vs-non-math language-based task, while providing evidence that elements of this results hold already in six-year-old children.

More generally, our results fit with a broader literature on the neuroimaging of mathematics, although geometry was rarely the center of attention. On the one hand, the ventral area we identify has been associated with visually presented numbers (Hermes et al., 2017; Shum et al., 2013) and has been referred to as "visual number form areas" (VNFAs). On the other hand, intraparietal regions are active during many number-related tasks, including number-processing/calculation (Dehaene et al., 1999). These findings lend support to the hypothesis exposed in (Lakoff & Núñez, 2000) that mathematics build upon many different domains including numbers processing and geometric concepts, amongst others.

One caveat of this result, which we can address with our other task,

is that the runs were not task-less: participants were instructed to actively wait for a star to appear to press a button, and a star is a particular case of geometric shape so each shape might lead to a more complex decision process than faces or houses. Given the block nature of the task, and the fact that the star was visually very different from the geometric stimuli (very different contrast and luminance), it is unlikely that this would drive the effect, but as the areas identified coincide with areas identified in "multiple-demand" systems (Duncan, 2010) and problem-solving tasks (Fedorenko et al., 2013), the increase in activation could stem from participants paying more attention during the geometric shape blocks.

3.3.2 Geometry

Results
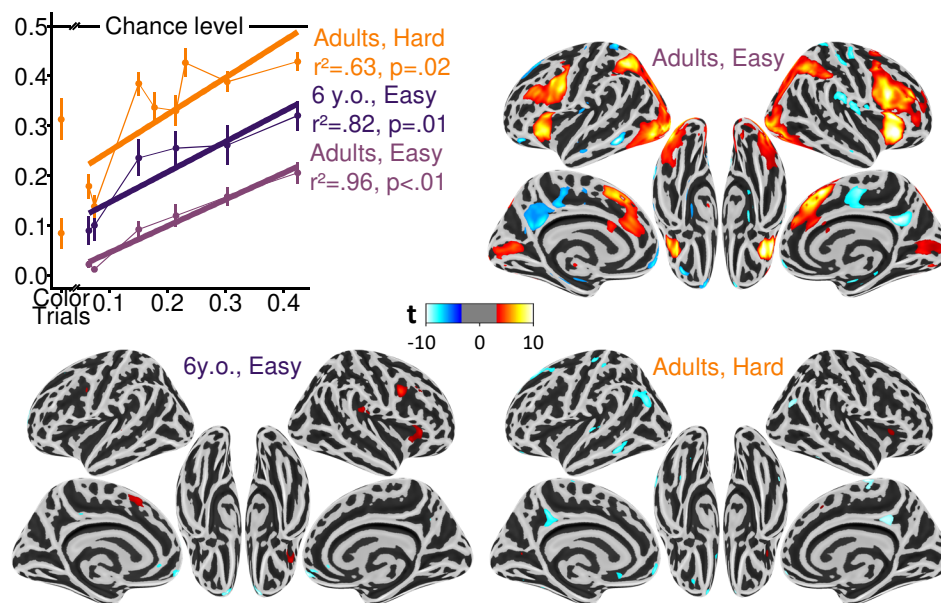
Figure 3.5: **Top-left**: correlation between the behavior of participants inside the fMRI and the averaged results for the same shapes in our online experiment of **chapter 1**, together with the r² and the p-value associated with the regression. **Other**: Linear, centered contrast of the geometric shapes, thresholded at the p<.0001, uncorrected. Top-right: adults, easy condition. Bottom-right: adults, hard condition. Bottom left: 6 years old, easy condition.

Behavior.

Table 3.2: ANOVAs of participants' behavior inside the MRI. For both age groups, for both tasks, we measured the effect of the shape on the error rates using an ANOVA. We report both the statistics and the generalized eta-square, an approximation of the r-squared for ANOVAs. There is always a significant effect, and its effect size ranges from .25 to .48, with its lower value for 6 years old indicating a higher level of noise in their behavior.

		Statistic	η^2_G	*p*-value
Adults	Easy	$F_{5,95}=29.91$	0.481	<.001
6 years old	Easy	$F_{5,95}=20.01$	0.256	<.001
Adults	Hard	$F_{7,133}=21.34$	0.469	<.001

First, we analyzed the behavioral responses inside the MRI. For both age groups and difficulty levels, there was a significant effect of the shape on the error rates, which came out as significant in three separate ANOVAs (see **Table 3.2**). What is more, within the "easy" difficulty level, we ran an ANOVA with shapes as within-participant factors, and the age group as a between-participant factor: there were significant effects of both factors (shapes: $F_{5,190}=44.73$, $\eta^2_G=.308$, p<.001; age group $F_{1,38}=13.64$, $\eta^2_G=.183$, p<.001) but no interaction between the two ($F_{5,190}=1.30$, $\eta^2_G=.013$,p=.27).

This can also be visualized and analyzed by computing the correlation of the behavior inside the MRI with the data obtained in a previous study using a similar paradigm, in N=117 adults tested online. The three independent correlations are displayed in **Figure 3.5** on the top left, together with the associated statistics: each correlation is always significant and the slopes of the regressions are virtually identical and not significantly different across age groups and difficulty. Additionally, the ordering follows the expectations with the easy condition yielding better performance than the hard condition, and the adults having better performance than the six-year-old participants. Note that the chance levels inside the scanner and outside the scanner are very different (one-in-two inside the fMRI, one-in-six for the online experiment).

Linear Contrast.   To see which brain areas had an increased activity associated with an increase in geometric shape regularity, we computed a linear contrast of the betas associated with each shape using the weights of the participants' behavior inside the scanner. This way,

positive t-values indicate areas in which the activity increases with the error rate, and therefore the complexity of the shape, and negative t-values indicate areas where the activity decreases with the error rate. Maps of these contrasts, thresholded at the p<.0001 level uncorrected, are provided in **Figure 3.5**, and information about significant clusters after bootstrap analysis are provided in **Table 3.3**.

Table 3.3: Location, size and significance of clusters for both age groups for the centered linear contrast of the geometric shapes. Locations are given in MNI coordinates. We report all clusters identified as significant at the p<.05 level in any condition, *p*-values are computed with non-parametric bootstrap like in **Table 3.1**

Condition	Coords (x, y, z)	Peak t-value	Size (mm^3)	p-value
Adults, easy	(-34.5, -96.5, 1.5)	12.1	181408	<.0001
	(29.5, 31.5, -0.5)	11.2	111984	<.001
	(-38.5, 5.5, 33.5)	9.65	38448	.0011
	(-38.5, -74.5, -54.5)	7.72	10320	.0176
	(13.5, -14.5, 15.5)	5.51	7784	.0354
	(37.5, -12.5, 25.5)	-6.9	10168	.0198
	(17.5, -40.5, 81.5)	-6.77	45776	.0011
	(5.5, -50.5, 27.5)	-6.64	23080	.0043
	(-14.5, 73.5, 7.5)	-6.07	14216	.0116
	(-18.5, 33.5, 39.5)	-5.69	6224	.0475
	(-46.5, -78.5, 49.5)	-5.53	8464	.0309
Adults, Hard	(45.5, -76.5, -38.5)	-6.5	2136	.0298
	(-20.5, 33.5, 55.5)	-6.31	3408	.0152
	(-48.5, -70.5, 33.5)	-5.55	2720	.0221
	(-6.5, -40.5, 41.5)	-5.28	2080	.0352
6 y.o., Easy	(-2.5, 63.5, -14.5)	-6.15	1848	.0305

There are several very significant clusters in the adults: bilateral middle occipital, bilateral superior parietal, bilateral BA 45 (the occipital and dorsal cluster appear merged in the table as they overlap) are positively activated; right lateralized BA7, right primary somatosensory are negatively activated. Activation in the hard condition, and in children, are much weaker, and there is little overlap: the only region that can be considered to be in common, but doesn't survive bootstrap tests, is a negatively correlated, bilateral BA10, very frontal.

Representational Similarity Analysis. The results presented here are preliminary, and require additional (ongoing) analyses. To better understand the mental representation of the quadrilaterals, we also performed a searchlight-based Representational Similarity Analysis (RSA) analysis of the data. For each voxel, participant and run, we could select nearby voxels in a sphere of radius 3 voxels, and compute the dissimilarity matrix across shapes to see, for each brain region, which
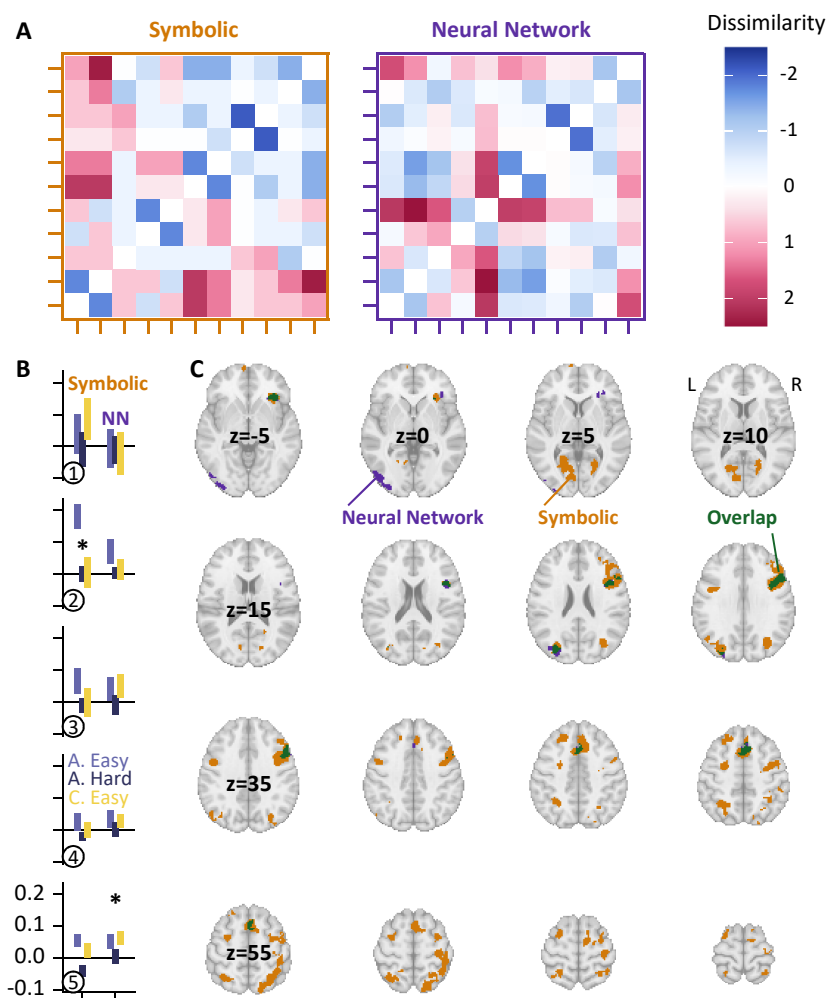
Figure 3.6: Searchlight-based Representational Similarity Analysis of the data. **A.** Confusion matrices used to model the data: the symbolic one is derived from the geometric properties feature vectors, and the other from RSA analysis of the internal representation of the shapes by CorNET, a neural network of object recognition. **B.** Average whitened correlation distance for each ROI identified with the category localizer, for each model (x-axis) and each condition and population (color). The bars represent the standard errors of the mean, with the average not shown in the middle, and significant conditions after multiple-comparison correction are indicated with a star. **C.** Significant clusters identified in adults in the easy condition for equally spaced slices along the z-axis, see **Table 3.4** for the list of the clusters' location and properties. Purple voxels belong to a significant cluster for the neural network model, orange voxels belong to a significant cluster for the symbolic model, and green voxels belong to a significant cluster for both models independently.

shapes were more alike and which were more different. We computed the cross-validated Mahalanobis distance (short: crossnobis distance; cross-validated over runs) following (Walther et al., 2016) for each participant, in both populations and in both the easy and the hard condition.

Then, we separately computed the whitened correlation of each voxel's dissimilarity matrix with two models. The first model is the symbolic model derived by saying that two shapes are as dissimilar as they have different symbolic regularities: it is a centered and normed transformation of the Manhattan distance between the feature vectors of two shapes. The second model is derived from a neural network by computed the crossnobis distance between the internal representation of our shape in the IT layer of CorNET, a neural network of object recognition scoring high in the brain-score metric (Kubilius et al., 2018; Schrimpf et al., 2018). This time the cross validation is performed over random batches of shapes with different scaling and rotations, using the same parameters as the experiment. The dissimilarity matrix derived from these two strategies are shown in **Figure 3.6.A**.

First, we tried to see whether the areas identified in the category localizer were associated with either model in the RSA. For each cluster identified in that analysis, for each participant and condition, we averaged the distance of the dissimilarity matrix and each model, and we show the average across participants in **Figure 3.6.B**. Then for each cluster and model, we computed a mixed effect with participants as random effects, and the condition as fixed effect, and we look at which intercepts are significantly above zero. Because we are computing ten independent statistics this way, we corrected p-values with FDR (Benjamini & Hochberg, 1995), and two clusters came out as significant at the p<.05 level: one with the symbolic model (cluster 2, p=.012) and one with the neural network model (cluster 5, p=012). The first one corresponds to the positive cluster located in the right parietal cortex, while the second corresponds to the negative, right-lateralized, ventral cluster. Zomming in on those cluster, we performed one-sample, one-sided t-tests to see which condition and population contributed most to the effect: for cluster 2, significant with the symbolic model, the effect was entirely driven by the adults in the easy condition (visible on **Figure 3.6.B**; p<.001, both hard condition and children yield p>.05). For cluster 5, significant with the neural network model, the effect was

present both in adults and children in the easy condition (resp. p=.006 and p=.02), but absent in the hard condition (p>.05).

Finally, we performed whole-brain search of significant clusters significantly associated with either model, for each condition and each population. Significance of clusters is assessed with non-parametric bootstrap similarly to the strategy used for the category localizer. There were many significant clusters associated with both models in the easy condition in adults, they are shown in **Figure 3.6.C**. In the hard condition, no cluster was significantly associated with either model, and in children a single cluster was identified, for the neural network model. To perform more targeted analyses, we then averaged the data of each participant over each significant cluster identified in the easy condition in adults, and performed one-sample, one sided t-test against chance to see whether these clusters were also relevant in other age groups and conditions. In doing so, we performed many independent tests, and therefore the *p*-values were corrected using FDR: after this, no cluster was identified as significant in the other groups at the p<.05 level, but two clusters showed a trend, both in children: one with the neural network model, at MNI coordinates (55.5, 15.5, 31.5), around BA44, with p=.092, and the other one with the symbolic model, at MNI coordinates (37.5, -6.5, 51.5) with p=.092.

Table 3.4: Location, size and significance of clusters identified for the symbolic and neural network model using searchlight RSA. *p*-values are derived with non-parametric bootstrap. No significant cluster was identified for the hard condition in adults, and a single cluster was found for children, for the neural network model.

Theory	MNI (x, y, z)	Peak t	Size (mm^3)	Weight	p-value
Symbolic	(1.5, 13.5, 49.5)	9.41	7048	4520	.0001
	(-12.5, -66.5, 5.5)	8.7	3496	2020	.002
	(41.5, -34.5, 57.5)	8.23	8648	4740	.0001
	(-10.5, -74.5, -22.5)	7.77	3160	1760	.0031
	(-32.5, -54.5, 51.5)	7.34	3928	2160	.0014
	(37.5, -6.5, 51.5)	7.3	13224	7710	<.0001
	(-36.5, -72.5, 31.5)	7.06	3136	1760	.0031
	(31.5, 31.5, -8.5)	6.89	1616	950	.0113
	(23.5, -64.5, 3.5)	6.64	1560	872	.0142
	(35.5, -74.5, -38.5)	6.34	1440	824	.0168
	(37.5, -70.5, 23.5)	6.31	2168	1210	.006
	(-30.5, -4.5, 55.5)	6.2	2344	1330	.0047
	(-30.5, -72.5, -50.5)	6.08	1272	698	.0239
	(-6.5, 65.5, -8.5)	5.88	1264	660	.0267
	(-16.5, 33.5, 45.5)	5.74	1712	943	.0114
	(-46.5, -0.5, 31.5)	5.65	1360	733	.0209
IT	(55.5, 15.5, 31.5)	6.64	3248	1790	.0007
	(-44.5, -80.5, -2.5)	6.1	1984	1120	.0048
	(35.5, 29.5, -6.5)	5.6	1544	816	.0145
	(-30.5, -82.5, 25.5)	5.3	1008	541	.0401

Theory	MNI (x, y, z)	Peak t	Size (mm³)	Weight	p-value
	(1.5, 21.5, 45.5)	5.26	1816	960	.0078
Children, IT	(-10.5, -78.5, 1.5)	6.29	1184	658	.033

Discussion

The results obtained with the intruder task offer a somewhat confusing picture. On the one hand, the behavior inside the fMRI scanner we exactly as predicted, indicating that the task was well understood and performed by subjects, in both age group and conditions. But on the other hand, results from the functional data do not warrant a strong conclusion.

In the easy condition in adults, many areas have an increased activity when the shape regularity increases, including areas in the dorsal pathways, and bilateral Broca areas. This is promising but confounded by the fact that the activity could simply increase with the difficulty of the decision part of the task, rather than being a property of the shapes themselves, and once again fall in multiple demand or problem solving (Duncan, 2010; Fedorenko et al., 2013), areas, rather than being geometric-shape specific. However, similar areas significantly correlate with the symbolic model in RSA, suggesting that the shapes themselves and not only the difficulty of the task contribute to this increased activity.

In both analyses, the clusters identified do not overlap well with the clusters identified in the category localizer. Conversely, when we test the clusters from the category localizer with RSA, only two of them are associated with either model in the RSA, once again only in the easy condition: a ventral one with the neural network model, and an intraparietal one with the symbolic model.

Another puzzling element comes from the location of the areas are well modeled by the neural network model in the RSA: while there is a strong cluster in the left inferior occipital cortex, exactly where we expected it, the strongest cluster is in the right hemisphere in the inferior frontal gyrus, around BA44, and strongly overlaps with the symbolic model: this was completely absent from the analysis in **chapter 2** where we find an early occipital cluster associated with the neural network model, and a late dorso-frontal cluster associated with the symbolic model. One possibility for this difference is that the RSA performed in **chapter 2** was restricted to two consecutive specific

time windows, one for each model: perhaps after an early correlation with the neural network in the occipital region, the information is used again later in the frontal areas, maybe to compute the symbolic properties, which would give the result we are seeing. In fact, when performing RSA not directly on the signal, but with decoders, which are more sensitive, we did find that the correlation with the neural network model was sustained, and overlapped with the symbolic model. Because these difference amounts to about 200-300ms, they would be invisible in the present fMRI study, which would just identify the two main clusters separately without being able to separate them in time.

More puzzling, neither the data from children or the data in the hard condition afford similar conclusions: neither has positive clusters in the linear contrast, and neither has strong clusters with the symbolic model in the RSA, and even testing the clusters identified in the easy condition in adults specifically does not yield significant association with either model.

What could cause these differences? Since the behavior results are as expected, it seems unlikely that participants did not form a mental representation of the shapes. What is more, the clusters identified in the category localizer naturally generalized from the adults to the children, suggesting that processing of simple geometric shapes used shared mechanisms. One possibility is that the areas identified relate to conscious access to the geometric properties: in the hard condition, the shapes were flashed very briefly, and the children may not explicitly realize that they are relying on the symbolic properties despite doing so – in conjunction with the visual properties, as evidenced by the findings of **chapter 1**.

Several additional analyses could be carried out – in particular, the current RSA analysis of the data uses theory-derived confusion matrix, but we could make used of the confusion matrices derived from behavior instead. Using a visual search task, we have already derived a confusion matrix in adults, we could similarly compute one in children, and use these two models to better understand the brain networks involved in the task.

Finally, it is important to stress out that the confusion matrices derived with RSA might be picking on the difficulty of the task, and represent a dissimilarity between difficulty levels, rather than intrinsic proper-

ties of the geometric shapes. In a sense, this is unproblematic: areas in which the difficulty pattern resembles the symbolic regularities are exactly those we are looking for. Still, it is more indirect than a passive task where there would not be any confounding factor coming from decision processes.

3.4   Conclusion

The results from the category localizer are unambiguous: in both adults and children, the perception of geometric shapes, when compared with other visual categories, elicits a strong dissociation. On the one hand, in the ventral pathway, geometric shapes generate less activity than either face, house or tools. On the other hand, geometric shapes activate the IPS bilaterally, and the ITS in adults: a result reminiscent of the network for numerosity and high-level mathematics in the brain, a result corroborated in our data by the increased activation of these areas to numbers as well.

Results from the intruder task are more ambiguous: in the condition with the highest success rate, the easy condition in adults, both linear contrasts and RSA analyses strongly yield vast networks whose activation is related to the symbolic geometric complexity, and a few additional areas that specifically appear to encode information similarly to neural networks of object recognition. This is a welcome result, but it does not replicate straightforwardly in the other conditions we tested, which prompts for additional analyses, and possibly additional data collection, both of which are ongoing work being carried out.

Chapter 4

Categorical perception of right angles in humans and baboons

Abstract

Right angles are one of the most iconic properties of geometric shapes. In this work, we systematically test humans in variations of a match-to-sample task to better understand the conditions under which their behavior with right angles is categorically different from their behavior from non-remarkable angles. We also provide a first comparison with baboons undergoing the same task in one of the conditions. In humans, we find that there are several constituents needed for right angles to elicit categorically different behavior: the task needs to give participants enough time and clues to be able to recognize right angles, and the task needs to be such that angles are the only possible strategy to answer correctly, providing insight about the nature of the mechanisms at play in the cognition of explicit geometric properties.

> When a straight line standing on a straight line makes the adjacent angles equal to one another, each of the equal angles is right, and the straight line standing on the other is called a perpendicular to that on which it stands.
>
> Euclid's Elements, Book I, Definition 10

In this chapter I will focus my attention on a single geometric object, the right angle, and compare how human and baboons perceived various angles around the right angle. Given a task involving angles, two things can occur. If the task lends itself to symbolic processing, we expect the perception of right angles to be more acute than that of other angles and to exhibit a behavior that sets it apart from other angles. If, on the other hand, behavior is a continuous function of angles as predicted by several theories of perception, then symbolic processes are not at play in that task. In this chapter, I find minimal pairs of experimental conditions that trigger one behavior or the other in humans, and provide a first insight as to the corresponding behavior in baboons.

4.1 Introduction

Right angles are central building blocks of fields as varied as engineering, painting, architecture or design. Visually, are the midpoint angle between no angle and a 180° angle, and as such they can be used to divide space equally as observed as early as Euclid' Elements (Byrne & Euclid, 1847). In turn this property grants them mechanical properties, stylistic properties, and so on, which makes right angles essential in many fields and applications. But how important and fundamental are they in human cognition?

Detection of right angles within different enough angle distractors is a task that educated adults, Mundurukus (people from an isolated Amazonian indigene group), and children can tackle with success (Dehaene et al., 2006; Izard, 2022). Angles form a continuum: if perception of angles were only bottom-up visual processing, we might expect behavior with angles to change with the angle monotonously. Indeed "association field" models of path integration suggest that disconnected segments are as easy to connect for the visual system as the angle between them is small (Field et al., 1993; Ledgeway et al., 2005).

At the same time, right angles have been argued to be perceived categorically and to behave singularly (Dillon et al., 2019): the typical amount of alteration required to detect a deviation from an angle is much lower for right angles and parallel lines than for other angles. This was empirically supported in an intruder detection task, and was shown is robust to size and orientation changes. Perhaps more striking, this effect is robust to the viewing angle: the same effect was replicated on tilted screens, but not on front-facing screens that displayed percepts retinotopically identical as the tiled screen, indicating that the sensitivity to right angles is really about the category and not about a property of that angle on the retina.

Given this elementary geometric property, we set out to compare the behavior of humans and baboons on a task designed to elicit the categorical perception of right angles in humans. We designed an experiment where we expected right angles to elicit more accurate behavior than neighboring angles in humans, and test it in baboons. In doing so, we realized that the simplest design we devised was not enough to elicit a categorical perception right angles in humans, and therefore parametrically modified two components of the design independently: presentation time, and whether other visual cues could be successfully used for answering correctly.

More specifically, we used a delayed match-to-sample experiment with angles as stimuli. An example of a trial is shown in **Figure 4.1**. Within groups, we manipulate the duration participants can look at a shape, a proxy for the encoding time (conditions "fast" and "slow"); additionally, we manipulate whether the angles are directly visible as two segments intersect or they have to be inferred from two segments slightly separated by an empty space (conditions "connected" and "disconnected". Across groups we manipulated whether the rotation of the stimuli was preserved between the target and the match in the selection screen (conditions "rotation" and "no-rotation"): when the orientation is preserved, using the angle is not required as one can rely directly on the direction of the segments to choose the match. We made this strategy hard by using distractors that shared exactly one direction with the match, but participants still seemed to favor this strategy when answering.

Because as of the writing of this work we only have data in baboons in a condition that does not trigger categorical perception of right angles in adults, the method and results of the baboon experiment is only

superficially described in the discussion.

4.2 Method

4.2.1 Participants

The experiment was performed online in two recruitment sessions, advertised on twitter, spaced 20 days apart. We started to collect data on the no-rotation condition on December 28th 2021, and stopped on January 14th 2022, with 87 participants recruited. The rotation condition spanned January 18th 2022 to January 22nd 2022, with 67 participants recruited. Following criterions established in **chapter 5** and (Sablé-Meyer, Ellis, et al., 2021) we removed only one participant, in the no-rotation condition. In total we therefore report data from 153 participants. Overall, we report data from 66 females, 84 males, and three participants who preferred not to answer that question; ages ranged from 18 to 79 with a mean of 42.9, a median of 45, and a standard deviation of 14.1; and their education ranged as follows: High School 12, Bachelors 32, Masters 73, PhD 36.

4.2.2 Stimuli

Training and generalizing stimuli were cliparts of fruits and letters, they are all displayed in **Figure 4.2**; for both categories we had 12 different stimuli, split in 6 training and 6 generalizing sets.

Angle stimuli were conveyed using white on black segments of constant length across stimuli. We generated all possible angles from 0° to 180° by increments of 10°, and used angles at 30°, 60°, 90°, 120°, and 150° as target stimuli (in the disconnected condition, we additionally used 0° and 180°, respectively the parallel and aligned conditions). We ensured that the smallest distance between the two segments was constant across angles to avoid that distance being used as a proxy for the angle when performing the task. All angle stimuli in the disconnected condition are shown in **Figure 4.1**; the connected stimuli are almost identical except the two segments are shifted left/right until they touch in an extremity.

The connected and disconnected were interleaved, but the fast and slow conditions were different enough that we decided to have par-
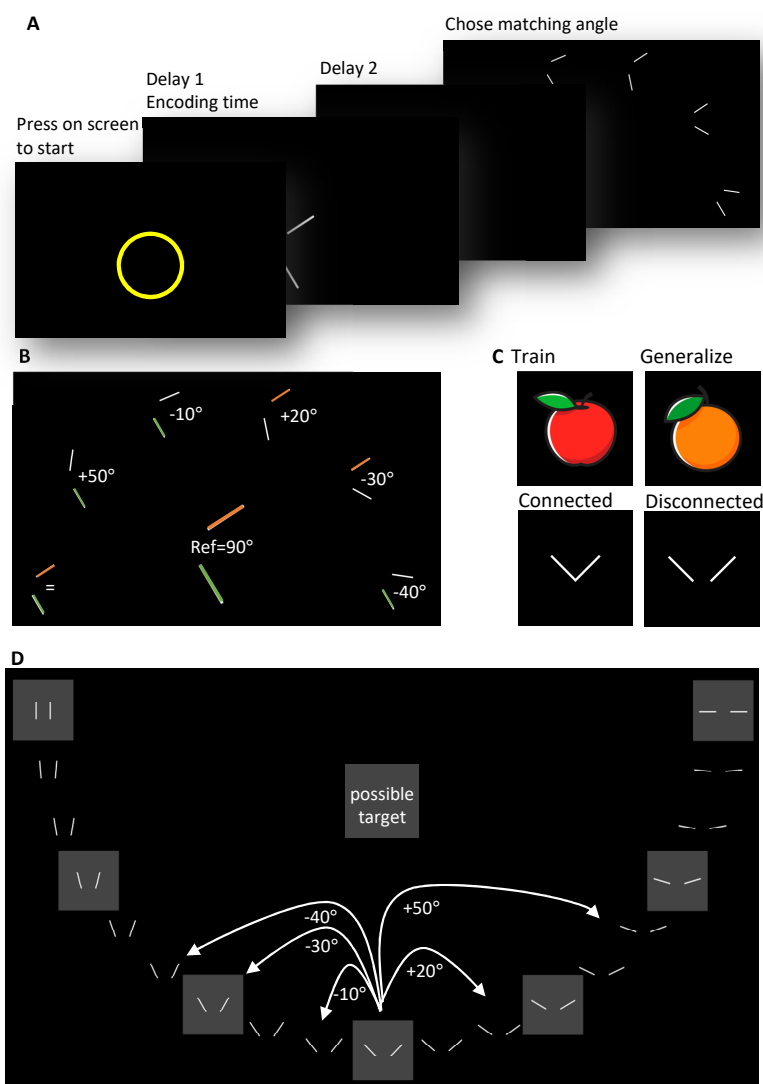
Figure 4.1: **A. General structure of the match-to-sample experiment**. The experiment follows a (possibly) delayed match-to-sample paradigm, where participants decide when to start a trial, and then see a target, possibly an empty screen, and then a set of 6 shapes in which to find the target. **B. Description of the display in A** In all cases, distractors were spaced by steps of 10° on either side of the target. In the fully rotated condition (not displayed), the orientation of each target was uniformly sampled in [0, 360°]. In the no-rotation condition, each distractor shared one side's direction with the target (see highlight in green/orange) with a 2/3 split for each direction. **C. Examples of stimuli** Exemplar stimuli for training, generalizing, and either connected or disconnected angle trials. **D. All possible disconnected stimuli**, plus all possible *targets* highlighted. Arrows indicated a possible choice of distractors for the right angle which match the example featured in A. Connected stimuli are almost identical: both parallel and "aligned" stimuli are removed, and the two segments are joined by shifting them an identical amount horizontally until they touch.
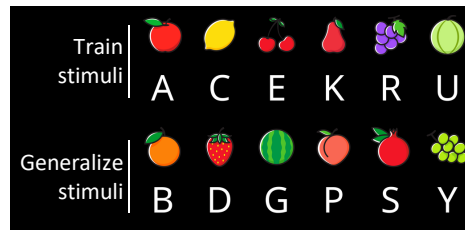
Figure 4.2: All possible training stimuli to assert understanding of the task.

ticipants take them one after the other, with repeated instructions in-between, and in randomized order.

In slow trials, the encoding time was 1000ms, and the duration between the disappearance of the target and the selection screen was again 1000ms. In fast trials, the encoding time was 100ms and was immediately (0ms) followed with the selection screen.

In total, participants took 4 trials for each angle, with either 5 or 7 possible angles (depending on the connected / disconnected condition), in two conditions, for a total of $(4*5 + 4*7)*2 = 96$ trials.

Additionally, participants received training trials before the angle trials. The speed of the training trials matched that of the first block participants would get, and they did not receive training again upon changing block. The training trials were organized in increasingly hard steps, and a success on a given step would move participants to the next step. First participant had to find the match in three distractors, then four, then five, and finally six. For all these steps, the size of the target and the match were completely identical. Then four more steps increasingly scaled up the size of the target when compared to the selection screen, with scales going from 1x to 1.25x, 1.5x, 1.75x, and finally 2x. Then entirely new stimuli were used for the generalization trials (see **Figure 4.2**), and finally participants were finished with the training.

During the entire experiment, participant would receive gamified feedback in the form of a visual cue (either V or X, respectively green or red) and auditory tones (respectively upward or downward going tones).

This training sequence was designed so that it would be possible to test non-human primates in the same experiment in possible future work, but in the present chapter we report human data only.

4.2.3 Procedure

When participants clicked the link in the tweet, they landed in a language choice page (French or English). Then they were prompted with the consent form, followed with the demographic questions (gender, age, highest degree, and touchscreen/mouse device). Then depending on which block they started with, participants received the slow or the fast instruction text, then took trials that utilized cliparts and letters at the corresponding speed, and then moved to the angles. Halfway through the experiment, participants received new instructions pertaining to the change of speed, and took the other half of the experiment without new training.

4.3 Results

Figure 4.3 shows aggregated results for all the conditions. There are several immediate results to observe. First, in all conditions, for all target angles, participants perform better than chance. This is visible on **Figure 4.3** and can be confirmed by one student test across participants on their average success rate against chance (at 1/6), for each target angle and condition: all *p*-values are <.001. Then we run an ANOVA across participants to measure the effect of whether angles were rotated or not (across participants) and the target angle, speed condition, and connected/disconnected condition (within subjects). Because all 15 interactions are computed for these four factors we only report those that reach significance, and report them in order of decreasing explained variance estimated by the η^2_G. Furthermore, some data points were removed within participants: for this analysis only, this led to some participants missing conditions, and given the total number of participants we decided to remove those 20 participants entirely from this analysis.

First comes the main effect of whether angles are connected or not (η^2_G=.049, $F_{1,131}$=162.86, p<.001), which we can see on **Figure 4.1** as the dashed lines are always below the solid lines. Then comes the main effect of delay (η^2_G=.036, $F_{1,131}$=108.21, p<.001) and the main effect of the target angle (η^2_G=.032, $F_{4,524}$=16.81, p<.001), both expected from the figure. Then we see the simple interaction terms of interest: first between the target angle and the connected condition (η^2_G=.013, $F_{4,524}$=10.75, p<.001), with the speed condition (η^2_G=.012,
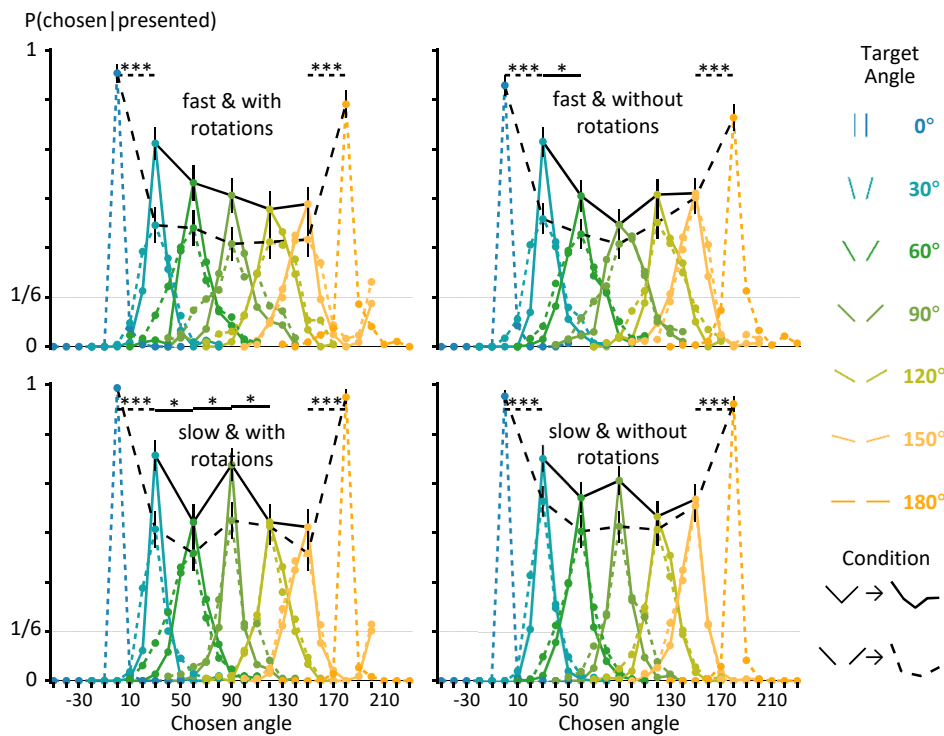
Figure 4.3: **Results of the experiment.** Each point represents the probability that a given angle is chosen if it was presented as a possible distractor, with color separating target angles, linetype separating conditions (connected or disconnected), and panels separating (no-)rotations, and fast/slow conditions. Error bars represent standard error of the mean across participants. Chance level is is at $1/6^{th}$. The value presented when the target angle and chosen angle are equal represents the success rate, while the other points of the same color show the distribution of error. Note that since not all possible distractors are presented in each trial, probabilities do not sum to one. Random behavior would yield flat behavior at chance level, while perfect behavior would show a single spike at 1 when the target angle and the chosen angle match, with values of 0 everywhere else. Stars above segments indicate that the difference between two neighboring target angles is significantly different, and again the linetype indicate the condition where there is a difference.

$F_{4,524}=10.22$, p<.001) and with the rotation condition (η_G^2=.005, $F_{4,524}=2.75$, p=.027). These three effects reflect the fact that the "right-angle" effect requires specific conditions, and we'll pinpoint it more precisely below. In addition, the interaction between the rotation condition and the connected condition is significant (η_G^2=.002, $F_{1,131}=7.60$, p=.007). No other interaction reaches significance at the p<.05 level.

The ANOVA treats angles as categorical factors. But we expect nearby angles to yield comparable behavior, and therefore we can perform Fisher tests for all pairwise neighboring angles: these tests are reported in **Figure 4.3** with segments and significance stars, at the top. We find that 0° and 180° are, as expected, very different from their neighbor angles, and therefore reach significance no matter the condition (all p<.001). In addition to this, in the connected condition a few additional pairwise comparison reach significance: the 30°-60° pair in two conditions, and the 60°-90° and 90°-120° pairs in the slow condition with rotation.

More specifically, when right angles are treated as categorical, we expect their performance to be better than both 60° and 120°, while when there are not we expect them to perform worse. This can be tested specifically by computing the contrast right angle vs. average of 60° and 120° for each participant, and see under which condition(s) the distribution is significantly above 0. This turned out to be significant at the p<.05 only for slow trials where the lines are connected (with rotation, t=5.32 and p<.001; without rotation, t=1.92 and p=.029) with a trend for slow trials with disconnected lines and lines disconnected (t=1.44 and p=.077).

4.4 Discussion

Participants in our experiments could always recognize an angle in a set of distractors better than chance; this held true no matter the angle to be recognized, the presentation speed even when flashed 100ms, whether the angles were rotated between the presentation screen and the choice screen, and whether the angles were visually immediately available or conveyed with two non-intersecting segments. Participants were excellent for parallel lines and aligned segments, and in most cases their performance were u-shaped between these two extreme points, with a minimum for 90°.

153

A. (from Field et al., 1993)

B. (from Iacaruso et al., 2017)
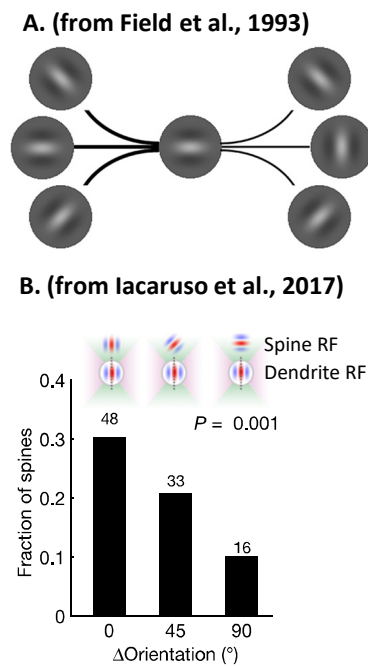
Figure 4.4: **A.** Schematic illustration of the association field model borrowed from (Field et al., 1993): when parts of a stimulus align along a smooth path with no inflection point such as on the left, they are naturally linked together in a path, but the same doesn't occur on the right. **B.** Illustration adapted from (Iacaruso et al., 2017); Frequency of spines as a function of their difference from the reference orientation either in the co-axial (left) or orthogonal (right) visual space. Number of spines indicated above bars, out of resp. 97 (left) and 62 (right), p values obtained by permutation tests out of 44 dendrites and 17 mice.

This result is in line with bottom-up perception accounts and in particular "association field" models of contour integration (Field et al., 1993; Ledgeway et al., 2005). These models propose that the strength of linking between close elements in a stimulus is proportional to their alignments along smooth paths, and show that performance in path-detection tasks degrade with the typical angle between two consecutive stimuli, a result compatible with the u-shape curved described above assuming a similar phenomenon occurs for parallel lines. Intracranial data in mice (Iacaruso et al., 2017) sheds light on the implementation level account of this phenomenon, and provides support for the idea that this phenomenon stems from simple statistics of the environment which features many roughly aligned neighboring edges.

However, under specific conditions, the behavior of right angles deviates from this and spikes well above neighboring angles: this has been referred to as the "categorical perception" of right angles (Dillon et al., 2019). More specifically, this phenomenon seems to build on three properties: (i) enough time to look at the angle, (ii) stimuli where the angle is directly visible rather than inferred from segments, and most interestingly (iii) knowing that the angle is absolutely required for the task. (i) and (ii) suggest that unlike the association field model, the mechanisms at play are not just bottom up but in fact requires efforts; on the other hand (iii) suggests that attention is required for the categorical perception to be formed, following findings in **chapter 1** and (Sablé-Meyer, Fagot, et al., 2021).
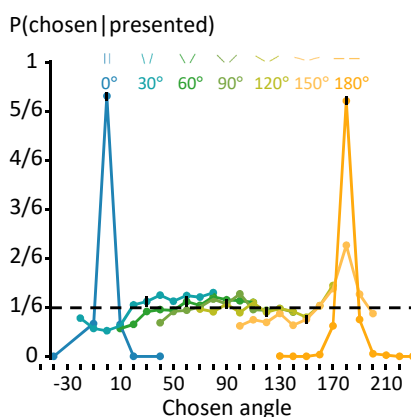
Figure 4.5: A replication of the fast version, disconnected angles, no rotation conditions of the match-to-sample task in baboons, with the same conventions.

Because this task is procedurally very similar to the task in **chapter 5**, for which baboons were trained, I briefly report here and in **Fig-**

ure 4.5 results we obtained with 8 baboons who undertook at least 67 blocks of 84 trials after averaging the remaining blocks, with a maximum of 100 blocks. Note that because of the training steps, they were never trained to ignore rotations, and therefore we tested them without changes in rotation. Furthermore we only used disconnected stimuli, and the experiment was self-paced: encoding time was stopped when the screen was touched a second time. However, they did not use this feature, and instead clicked as fast as they could with an average of 298ms of encoding time, comparable to the "fast" group in humans if a bit slower. Their performance were very good with both 0° and 180° angles (both p<.0001), but they were not significantly better than chance at the p<.05 level for any other angle (with a trend for 30°, p=.084). Like humans, there were no "right-angle" effect in this particular condition.

Unfortunately, the absence of effect in this condition is not conclusive regarding the question of whether baboons share a categorical perception of right angles with humans, because this condition does not elicit the effect in humans either. However, the data suggests that even with extensive training, the task is overall too hard for the baboons, who despite negative feedback and extensive training against the strategy still appear to choose the "aligned" stimulus when the target is 150°, indicating that the task was not properly understood. Future work should investigate whether with additional training and various conditions, the effect emerges, and if it does so in the same group of conditions as it does in the humans.

Chapter 5

A language of thought for the mental representation of geometric shapes

Abstract

In various cultures and at all spatial scales, humans produce a rich complexity of geometric shapes such as lines, circles or spirals. Here, we formalize and test the hypothesis that all humans possess a compositional language of thought that can produce line drawings as recursive combinations of a minimal set of geometric primitives. We present a programming language, similar to Logo, that combines discrete numbers and continuous integration to form higher-level structures based on repetition, concatenation and embedding, and we show that the simplest programs in this language generate the fundamental geometric shapes observed in human cultures. On the perceptual side, we propose that shape perception in humans involves searching for the shortest program that correctly draws the image (program induction). A consequence of this framework is that the mental difficulty of remembering a shape should depend on its minimum description length (MDL) in the proposed language. In two experiments, we show that encoding and processing of geometric shapes is well predicted by MDL. Furthermore, our hypotheses predict additive laws for the psychological complexity of repeated, concatenated or embedded shapes, which we confirm experimentally.

> We could never know the geometric triangle through the one we see traced on paper, if our mind had not had the idea of it elsewhere.

<div align="right">René Descartes</div>

This chapter corresponds to an article under consideration at Cognitive Psychology (current status: awaiting minor revisions). It is available as a preprint, referenced as follows: Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2021). A language of thought for the mental representation of geometric shapes. PsyArXiv. https://doi.org/10.31234/osf.io/28mg4

5.1 Introduction

The cognitive origins of geometric knowledge remain heavily debated. While several animal species possess sophisticated neural circuits for spatial navigation (including head direction, place, grid and border cells), or produce rich but systematic patterns (e.g. spiderwebs, honeycombs, or the spiral-like patterns of puffer fishes), only humans seem capable of mentally conceiving formal, symbolic geometric structures in a combinatorial and productive manner.

The formalization of geometry is traditionally dated to Euclid's *Elements*, itself rooted in Egyptian and Babylonian precursors. Yet various lines of evidence suggest that an intuitive sense of geometry is much more ancient, and that many, possibly all human cultures share a drive towards creating geometric designs (Van der Waerden, 2012). Throughout the world, at geographically distant and presumably unrelated sites, humans have produced parallel lines, circles, squares, zigzags or spirals, in activities as diverse as drawing, pottery, body paintings, rock art, land art (e.g., Nazca lines), stone-cutting (e.g., bifaces), or large-scale constructions (e.g., Stonehenge). Many Neolithic sites contain square, circular or rectangular buildings as well as large circles of stones (cromlechs) whose axes are often systematically oriented relative to geographical or astronomical landmarks (Pimenta & Tirapicos, 2015). Before the advent of artificial flight, the shape of these large structures could not be directly apprehended: from the ground, they would be perceived as a distorted quadrilateral or ellipse, at best. The fact that squares and circles appeared at many different scales suggests that their human designers possessed an abstract mental con-

cept of geometric shape that guided their architectural, artistic or practical creations.

5.1.1 Human and Animal Sensitivity to Geometric Patterns: A Brief Review

Evidence for abstract concepts of geometry, including rectilinearity, parallelism, perpendicularity and symmetries, is widespread throughout prehistory. About 70,000 years ago, *homo Sapiens* at Blombos cave carved a piece of ocher with three interlocking sets of parallel lines forming equilateral triangles, diamonds and hexagons (Henshilwood et al., 2002). Much earlier, approximately 540,000 years ago, *homo Erectus* in Java carved a zig-zag pattern on a shell (J. C. A. Joordens et al., 2015). Such a zig-zag may look simple, but it approximately respects geometric constraints of equal lengths, equal angles and parallelism, and is undoubtedly attributed to the *homo* genus. Even earlier, since ~1.8 million years, ancient humans have been carving spheroids (sphere-like stones) and bifaces — stones possessing two orthogonal planes of symmetry (Le Tensorer, 2006). The vast number of bifaces, their near-perfect symmetry (which is not required for them to operate as efficient tools (Le Tensorer, 2006)), and the archeological evidence that many were never used as tools, suggest that an aesthetic drive for symmetry was already present in ancient humans.

Contemporary cognitive anthropology corroborates those findings. Cognitive tests performed in relatively isolated human groups such as the Mundurucu from the Amazon, the Himba from Namibia, or aborigine groups from Northern Australia, show that in the absence of formal western education in mathematics, adults and even children already possess strong intuitions of numerical and geometric concepts (Amalric et al., 2017; Butterworth et al., 2008; Dehaene et al., 2006; Izard et al., 2011; Pica et al., 2004; Sablé-Meyer, Fagot, et al., 2021). Indeed, adults without formal western education share with Western preschoolers a large repertoire of abstract geometric concepts (Dehaene et al., 2006) and use them to capture the regularities in spatial sequences (Amalric et al., 2017) and quadrilateral shapes such as squares or parallelograms (Sablé-Meyer, Fagot, et al., 2021). They even possess sophisticated intuitions of how parallel lines behave under planar and spherical geometry, such as the unicity of a parallel line passing through a given point on the plane (Izard et al., 2011).

159

Another piece of evidence arises from developmental data. Preschoolers and even infants have been shown to possess sophisticated intuitions of space (Hermer & Spelke, 1994; Landau et al., 1981; Newcombe et al., 2005), spatial sequences (Amalric et al., 2017), and mirror symmetry (Bornstein et al., 1978). Indeed, preschoolers' drawings already show a tendency to represent abstract properties of objects rather than the object itself. Although they look primitive, drawings of a house as a triangle on top of a square, or a person as a stick figure with a round head, suggest a remarkable capacity for abstracting away from the actual shape and attending to its principal axes, at the expense of realism. Numerous tests leverage this geometric competence to assess a child's cognitive development by counting the number of correct or incorrect abstract properties, for instance when asked to draw a person (Goodenough, 1926; Harris, 1963; Long et al., 2019; Prewett et al., 1988; Reynolds & Hickman, 2004). There is some evidence, however limited, that this ability may be specifically human: when given pencils or a tablet computer, other non-human primates do not draw any abstract shapes or recognizable figures, but mostly generate shapeless scribbles (Saito et al., 2014; Tanaka et al., 2003).

We recently compared the perception of quadrilateral geometric shapes in humans and in baboons, using the very same task (Sablé-Meyer, Fagot, et al., 2021). We used the intruder test (Dehaene et al., 2006), which involves viewing an array of pictures and clicking on the one that looks distinctly different from the others, and is well within the grasp of human adults, children and baboons. All humans, regardless of age, culture and education, exhibited a striking effect of shape regularity: intruders amongst squares and rectangles were detected faster and more accurately than amongst other, more irregular quadrilaterals, and there was a systematic gradient of response time and error rates across shapes, from squares and rectangles to parallelograms, trapezoids, and fully irregular shapes. Strikingly, this geometric regularity effect was absent in baboons. Baboon behavior was quite consistent across individuals and could be captured by neural network models of the ventral visual pathway for object recognition. Modeling the human perception of quadrilaterals, however, required an additional assumption, namely the existence of discrete symbolic concepts of parallelism, right angle, equal length, or equal angle. We therefore argued that two strategies can be used to perform the outlier task: a visual one, available to all primates, and an

abstract, symbolic one that may be unique to humans (Sablé-Meyer, Fagot, et al., 2021).

From these observations, one is left wondering what could be the origins and evolutionary advantage of the human competence for abstract geometry. We propose that it is a specific case, in the visual domain, of a general human ability to decompose complex percepts and ideas into composable, reusable parts – an ability which led to a massive enhancement of human productions, from architecture to tool building, and of the capacity to understand the abstract features of the environment.

5.1.2   Summary of our Approach and Hypotheses

In the present paper, we formalize and put to an empirical test the hypothesis that geometry is one of the manifestations of the specifically human ability to represent and manipulate recursively embedded languages (Dehaene et al., 2015; Fitch, 2014; Fodor, 1975; Frankland & Greene, 2020; Hauser et al., 2002; Piantadosi, 2011).

Fodor (1975) famously introduced the language of thought hypothesis, according to which an inner combinatorial language underlies high-level cognition in humans and allows the creation of a vast space of mental representations by recursive recombination of preexisting ones. Hauser, Chomsky and Fitch (Hauser et al., 2002) hypothesized that recursion might be the single uniquely human ingredient that explains the emergence of the human language faculty. Fitch (2014) and Dehaene et al. (2015) later argued that recursion is not limited to linguistic communication, and that various "languages of thought", all based on a basic capacity for recursive syntax and compositional semantics, could underlie many other uniquely human abilities such as music, mathematics or theory of mind. Here, we apply this idea to the domain of geometric shape perception, a possibility which was anticipated by (Hochberg & McAlister, 1953) who summarized their proposal as "the probability of a given perceptual response to a stimulus is an inverse function of the amount of information required to define that pattern."

Our proposal builds upon the seminal work of Leeuwenberg and colleagues (Boselie & Leeuwenberg, 1986; E. L. J. Leeuwenberg, 1971), who proposed a formal coding language for 2- and 3-dimensional shapes, and showed that it could account for data on human shape

perception. Furthermore, Leeuwenberg's language outputs sequences of numbers which can then be mapped to tones to form "tunes", and the length of the smallest program for a given tune correlated with the average listening time of participants asked to repeatedly listen to the tune until they anticipated they could predict it. Later, Leyton (Leyton, 1984, 2003) argued that the shapes that humans generate arise from a set of primitives (points, lines, planes) together with the repeated mental application of a series of group transformations that duplicate, stretch, rotate, or skew them. While these elegant proposals have had a considerable influence in the design of graphics software, it is fair to say that the core aspect of productive compositionality from basic operations remained partially disconnected from the experimental psychophysical or neurophysiological literature on shape perception, while the individual transformations remained (for exceptions, see Brincat & Connor, 2004, 2006; Hung et al., 2012).

In our previous work, we introduced a much more restricted, yet more precise, language of thought for geometric sequences. Our work focused on capturing the psychological complexity of all the sequences of 8 locations that can be generated by drawing without repetition from the vertices of an octagon using either explicit prediction of the next location or eye tracking of the anticipation when looking at a sequence (Amalric et al., 2017). The basic building blocks of our proposed language were the arithmetic primitive of discrete number, the geometric primitives of rotation and of symmetry around a given axis, and a single recursive operation of repetition (possibly with variations). These operations could be embedded, thus allowing for repetitions of repetitions in a nested manner. For instance, the repeated application of a left-right symmetry operation, each time with an increment in the starting point, could generate a zig-zag pattern. A square could be generated by a 4-fold repetition of moving by 2 vertices around the octagon. As a more complex example, a sequence of two squares could be generated by two nested "for loops", i.e. a 2-fold repetition (while changing the starting point) of the 4-fold repetition that draws a square.

Amalric et al. (2017) measured empirically the difficulty that preschoolers and adults (including Mundurucu adults) had in predicting or memorizing spatial sequences of locations on an octagon. Across 11 geometric sequences, psychological complexity was determined by the complexity of their internal representation in

the proposed language. Working memory was not determined by sequence length (which, indeed, was fixed at 8 items), but by the capacity to compress the sequence into a compact internal representation using the proposed language. The central concept here, as already proposed by many others (Chater & Vitányi, 2003; Feldman, 2000, 2003; Li & Vitányi, 1997; Mathy & Feldman, 2012; Romano et al., 2013), is that psychological complexity in humans depends on **minimum description length (MDL)**, i.e. the length of the shortest mental representation which can encode the sequence, rather the literal length of the sequence. In the case of our language for geometric sequences, MDL was also shown to tightly correlates with brain activity in both functional magnetic resonance imaging (fMRI; Wang et al., 2019) and magneto-encephalography (Al Roumi et al., 2021). The very same language was also successfully extended to account for the perception of simple auditory sequences made of two discrete sounds (Planton et al., 2021).

In the present work, we move beyond discrete sequences made of points and straight lines, and tackle the mental representation of static geometric shapes such as a square, a circle or a spiral. As noted above, the square is easily captured by a language with discrete integers and repetition ("for loops"). However, continuously varying shapes such as circles and spirals raise interesting issues that arguably require more than integers. In computer languages such as Logo, such drawings are implemented using a discrete repetition instruction with a very small increment, thus drawing a quasi-continuous curve which is in fact made of straight lines. However, we find implausible the idea that humans intuitively think of such an infinitesimal and inherently discrete representation when thinking of a circle. Furthermore, computationally, the unbounded nature of such infinitesimal loops would allow short programs to generate visually complex shapes. Instead, we argue that the crucial notion of "repetition with variation" introduced in our previous work can be helpful again, but now in a continuous version. We propose that, whenever a mental primitive is available, for instance for drawing a straight line, mental control structures in humans are available to either keep its parameters constant, or to continuously vary them over time. Thus, the new version of our proposed languages includes not only discrete repetition ("for loops"), but also continuous repetition (i.e. integration). As a result, the language can conceive of a curve with a fixed amount of turning at any moment – a circle –, or a curve where the amount

of turning increases continuously – a spiral –, etc. Our proposal implies that both discrete repetition and continuous path integration are primitive concepts in the human language of geometry. Indeed, a key hypothesis of the present work is that the human mind can encode discrete as well as continuous changes and integrate them within a single language of thought. This part of our proposal is deeply related to the near-universal existence of a system of aspect in human natural languages, betraying the existence of continuous versus discrete concepts of time and repetition (compare for instance the imperfective, e.g. "the curve was turning", with the perfective, e.g. "the curve turned") (Comrie, 1976).

In summary, we propose that the human mental representation of geometric shape involves a language of thought that can produce virtually all the geometric line drawings observed in human cultures as combinations of a minimal set of geometric primitives. Our core hypothesis is that perceiving a shape, in humans, consists in finding the shortest program that suffices to reproduce it. Our proposal thus connects shape perception to the problem of program induction, i.e. the identification of a program that produces a certain output. In line with much previous work (e.g. Chater & Vitányi, 2003; Feldman, 2000, 2003; Feldman & Singh, 2006; Li & Vitányi, 1997; Mathy & Feldman, 2012; Romano et al., 2013), we hypothesize that the perceived complexity of a shape is determined by its minimum description length (MDL) in the proposed language.

As a consequence of the relation between MDL and the perceived complexity of a shape, we predict that several behavioral measures should be directly impacted by the MDL of a shape. For instance, the time it takes to learn a shape (i.e., to induce its program), as well as its subjective complexity, should scale with the MDL. Other measures, such as the time to select a known shape amongst distractors, or the accuracy of that choice, may also scale with MDL to the extent that the influence of other competing strategies based on low-level visual properties (e.g. average gray level, spatial frequency) can be mitigated. Indeed, choice time is a function of the multivariate relation between target and the distractors (Vigo & Doan, 2015), so in the following experiments, we make sure that at least one distractor is close enough to the target in low-level visual properties, thus inciting participants to adopt a higher-level strategy. Note that our language targets geometric shapes specifically and makes no claims about other kind shapes, such as the con-

tours of natural objects, whose description requires other constructs (see Wilder et al., 2016).

Two comments are in order. First, the power of the proposed language rests entirely on its capacity to encode **repetition with variation** (e.g. a square and then another square). This concept is equivalent to invariance up to a transformation (e.g. invariance of the square up to a transformation of its starting point), which in mathematics, is the definition of symmetry: an object is said to be symmetric if it (or part of it) remains unchanged after some transformation. The language we proposed recursively compresses any detectable repetition with variation, and therefore any symmetry in the object or sequence. Second, our proposal is related to, but distinct from, the psychological concept of chunking. While our language decomposes objects into coherent subgroups, this proposal goes beyond mere chunking in that (1) it applies recursively and (2) its final representation is not just a set of nested groupings (chunks of chunks) but a mental program which can generate the initial shape or sequence, possibly with variations.

Below, we describe the proposed language in detail, list its predictions, and test them in two experiments. First, we show that our language predicts which shapes are judged simple. Second, we show that any such language has to satisfy a set of additive relationships for repeated, concatenated or embedded shapes, and that those universal laws can be experimentally validated.

5.2 A Generative Language for Geometric Shapes

In this section, we make our proposal concrete by introducing a specific language, somewhat similar to Logo's turtle language, for generating a variety of geometric line drawings. The language we propose is based on two postulates. First, we assume that all humans possess a set of primitive operators or "mental routines" (Ullman, 1984) that serve as building blocks for more complex programs. We included elementary primitives that have been proven to be present in human children or adults in the absence of formal education; several of them are likely to be inherited from primate evolution. Our primitives comprise the concepts of

- Small exact integers (Feigenson et al., 2004) which can be mini-

mally generated by the successor function $s(n) = n + 1$ (Izard et al., 2008),

- Fractions, i.e. ratios of those integers (Jacob & Nieder, 2009; Siegler et al., 2011)

- Straight line (see e.g. Izard et al., 2011),

- Heading direction (Muller et al., 1996) and how it changes when we turn,

- Path integration (Dehaene et al., 2006; Gallistel, 1990; E. L. J. Leeuwenberg, 1971; McNaughton et al., 2006; O'Keefe & Nadel, 1978),

- Right angle turn (Dehaene et al., 2006; Dillon et al., 2019; Izard et al., 2011)

Extensions of this list, for instance to large approximate numbers, would be straightforward and are considered in the discussion, but as we shall see, those primitives appear to suffice to account for a broad variety of geometric shapes that humans universally consider simple.

Our second postulate is that, in humans only, a compositional language of thought allows these primitive operators to be combined into larger programs. We suppose that three composition instructions are available: concatenation; repetition; and call to a subprogram in isolation.

5.2.1 Program Instructions

The full language, described in **Figure 5.1**, contains the following instructions. First, as in the "logo" language (Abelson et al., 1974), drawing instructions dictate the movements of a pen that can move and trace curves on a plane. Those instructions are *Turn*, which changes the current heading of the pen; *Move*, which changes the position of the pen by a certain amount in the current direction without tracing; and *Trace*, which traces a curve by integrating over a set of parameters (duration, speed, acceleration, and turning speed).

Second, the three control structures are *Concatenate* (also denoted by ";") which executes one program and then another; *Repeat*, which repeats a program a certain number of times (twice by default); and *Subprogram* which saves the current state, executes a given program, and

```
Program :=
  | Program ; Program              Concatenate : run one program and then another      control
  | 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],                                                          drawing
         [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                                    arithmetic
Num :=
  | +Int     | -Int                Return a signed number
  | +Int/Int | -Int/Int            Return the signed fraction of two integers
```

Figure 5.1: Proposed language of thought for the mental representation of geometric shapes. The figure lists all primitive operators and their parameters. As indicated in the right column, **control** primitives act on programs, **drawing** primitives move the pen on the plane in various ways, and **arithmetic** instructions generate integers and fractions that are passed as parameters. Green, instructions; pink, types; blue, named parameters; gray, default values for optional arguments (denoted by brackets).

resumes the previous state for the rest of the execution, thereby isolation the execution of the subprogram.

Third, since these instructions require either discrete or continuous arguments, the language contains a number system, with integers (*Int*) and numerical (*Num*) types. For computational simplicity, in order to avoid a huge combinatorial explosion that would prevent the enumeration of all minimal programs, we did not include a full algebra, in spite of recent evidence that humans may possess one (Grace et al., 2020). Instead, the *Int*'s are built using Peano arithmetic starting from 1 (the language has a *one* primitive and a *successor* primitive, and the *Num* are either signed integers (positive or negative), or signed fractions of two integers. This is enough to generate rational numbers, but prevents nesting of fractions.

The numbers generated by our language are unitless, and they are interpreted differently depending on the functions in which they are evaluated. For the *Turn* function, an argument of 1 is interpreted as "one right angle" (i.e. the unit for angle is "right angle"). Similarly, *Move* and *Trace* instructions use implicit units of length and speed, determined such that the default values (1) on duration and turning speed imply turning by a full circle. These hypotheses, while plausible, are not crucial, since changing them would only minimally change the predicted shape complexities (e.g. if the default turn was by 180°, a right-angle turn would still be available at a minimal constant cost, as half of it).

5.2.2 Calculation of Minimum Description Length

Our language does not guarantee that each shape can be generated by a single program. Quite the opposite: whenever a shape can be generated at all, an infinity of programs are available to generate it. Our third key postulate is therefore that humans search for the shortest program that draws a given shape. We refer to the complexity of the shortest program for a given shape as its Minimum Description Length (MDL), and the corresponding program(s) as the Minimal Program(s). Notice that this is compatible with a Bayesian framework, or probabilistic program induction (Lake et al., 2015): since the number of programs of length n increases exponentially with n, the log likelihood of a given program under the hypothesis of a probabilistic grammar will be proportional to its MDL.

The complexity of a program is defined as the number of nodes in its syntax tree, or equivalently the number of primitives in the program, with two exceptions. First, whenever a signed number is required, for instance when turning by a certain angle, an additional node is needed (indicating + or -). This node was not counted, thus preventing the cost of signed values from being systematically higher than that of unsigned values (e.g. one, vs. +one). Second, concatenations did not increase MDL. This feature arose as a by-product of our implementation, which used continuations to express concatenation (all programs takes a last argument which must be a program, possibly empty, and executes it when it is done, which effectively implements concatenations). We checked that the results did not change dramatically when adding counting concatenations, as the results should not hinge on this implementation detail.

5.2.3 Examples

The minimal program to draw a square is

```
Repeat { Repeat { Trace ; Turn(angle=+one) } }
```

(where the infix operator ";" denotes concatenation). This program works because, in the absence of any argument, Repeat defaults to 2 repetitions. Since a turn of one means a right-angle turn, this program concatenates four segments, each ending with a right-angle turn. The MDL of this program is 5 (repeat + repeat + trace + turn + one).

As a slightly more complex example, the following program draws a

triskelion (⛬), a classic Celtic figure:

```
Repeat(next(next(one))) {
  Subprogram { Trace(acceleration=-one/next(next(one)),
               turningSpeed=one)
             } ;
  Turn(angle=next(next(next(one)))/next(next(one)))
}
```

This program of complexity 21 draws three identical inward spirals using the *Trace* instruction. Note that the number 3 is coded as next(next(one)) (again, this assumption is adopted for simplicity; adding primitives for numbers 2 and 3 would only minimally change the predicted MDL). The *Subprogram* instruction ensures that, after drawing a single spiral, the position is reset to the origin. The *Turn* instruction, which takes 4/3 as its argument, ensures that the three spirals are oriented at $120° = \frac{4}{3}90°$ from each other.

5.2.4 Simulation Results

We first examined the shortest programs in the proposed language, and whether they always generate shapes that are simple and frequently attested in human cultural history. To this end, we wrote a program that systematically enumerates all possible programs in order of increasing MDL, and draws the corresponding shapes (we eliminated, automatically as well as manually, the shapes that could be generated by a simpler program). **Figure 5.2** shows a random subsample of the resulting shapes after cleaning of duplicates, sorted by MDL. The simplest, lowest complexity shapes are extremely simple: they consist of a line segment (MDL=1), then a circle (MDL=2) and a spiral (MDL=3). The low-complexity shapes with MDL = 4 or 5 are also excellent candidates for cultural universals: repeating circles, dashed lines, spirals with various numbers of loops, and other simple mathematical shapes such as the square, the half-circle, or two tangent circles. At this stage, the concatenation instruction also generates less intuitive, but still culturally attested shapes such as a "sigma" (segment + circle) or a "crosier" (segment + spiral). As MDL increases, the huge combinatorial explosion of programs results in an enormous variety of shapes, only some of which are culturally observed. Nevertheless, the shapes with low MDL remain introspectively simple (this intuition is tested formally in experiment 1 further below). This

Shapes sorted by Minimum Description Length

Figure 5.2: Sample shapes generated by the enumeration of all programs in the proposed language. Programs were listed by increasing MDL. Identical or perceptually indistinguishable shapes that could be generated by a simpler program were eliminated. Starting at MDL=4, only a limited sample of 7 shapes is shown, as the number of shapes increases exponentially with MDL.

observation is in stark contrast with most other such languages, such as Logo, where the combinatorial explosion creates short programs with complex, unintuitive graphic outputs.

5.2.5   Program Induction Using DreamCoder

A crucial aspect of our proposal is that humans encode a shape mentally by inferring a simple program that could generate it. Thus, the perception of a simple shape is an act of "program induction". Yet it is implausible that humans scan through thousands of programs before recognizing a square. Otherwise, the time required to recognize a shape would grow exponentially with the length of its shortest program. Thus, it is important to show that such an inference is, at least approximately, computationally feasible in our specific case. While program induction remains a difficult challenge for computer science, we leveraged a state-of-the-art program induction technique, the DreamCoder algorithm (Ellis et al., 2021). This algorithm is given programming problems via examples of the desired behavior, and searches for the simplest program that performs the task. Here, a task reduces to a shape, and DreamCoder has to find the shortest program that generates it. DreamCoder internally represents the language as a probabilistic grammar and enumerates programs according to their likelihood using the probabilistic weights of the grammar.

Two features of DreamCoder speed up the search. First, the weights are task-dependent and are suggested by a neural network for a given task. For instance, DreamCoder may learn that shapes with straight lines call the Trace instruction without any TurningSpeed argument. The neural network can be trained without any environmental input or supervision, simply by sampling a random program, generating the corresponding shape (called "dreaming") in a top-down manner, then using this internally generated shape-program data pair in supervised learning in order to adjust the bottom-up weights from the shape to its program representation (see also Lake et al., 2017). Second, DreamCoder builds new *abstractions* for pieces of programs that are often used for a given set of tasks: for example, if the shapes contain many right angles, it may create a new abstraction Turn(angle=1), thereby increasing the likelihood of programs that use it. In a Bayesian sense, this corresponds to updating the priors over the space of programs. This abstraction mechanism is useful to capture regularities in a corpus of shapes: subprograms used to draw the simplest ones can be reused to

171

draw more complex ones. Interestingly, these two mechanisms interact. As the grammar becomes biased towards using certain program schemas, the neural network also becomes biased towards recognizing them; for example, the neural network might increase the probability of the "turn" primitive when it sees angles, or that of "repeat" when it notices repeating patterns.
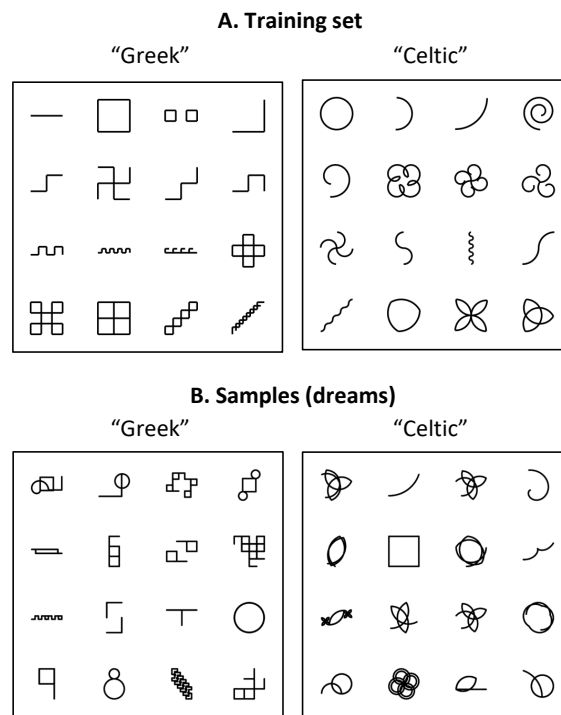
Figure 5.3: Testing the DreamCoder algorithm for program induction. A, shapes used in a training phase. We verified that, in response to all 32 shapes shown, the algorithm was able to identify a short, presumably minimal program that could generate it. The algorithm was trained either with the square shapes at left ("Greek" style) or with the circular shapes at right ("Celtic" style). B, examples of additional shapes spontaneously generated by sampling from the grammar learned during the training set, and therefore biased towards a certain geometric style.

We found that, together, those two mechanisms made program induction feasible for our language, at least for relatively simple shapes. In **Figure 5.3**, we present two separate corpora, one with mostly rectilinear shapes (referred to as "Greek"), and one with mostly curvilinear shapes (referred to as "Celtic"). After learning, DreamCoder finds fitting programs for each of these tasks: for more exhaustive evaluation of DreamCoder including train/test splits and ablation of its various components, including with drawing primitives, refer to (Ellis et al., 2021). Interestingly, the abstractions it created depending on the domain were different. This could be visualized by sampling "dreams"

from the resulting grammars. Simple shapes (circles, lines, squares) appear in both cases as they are still simple in a grammar with additional abstractions, but the grammars exhibit a bias towards shapes that resemble the training set (bottom row in **Figure 5.3**) because the new abstractions are tuned for observed shapes, and therefore yield visually similar shapes.

The DreamCoder approach opens up a number of perspectives on how human cognition could efficiently address the problem of program induction. First, it naturally accounts for cultural drifts: while shapes such as circles and squares are universally shared, cultures are also characterized by the frequent use of specific patterns (e.g. linear Greek friezes versus curvilinear Celtic spirals). This arises even though the geometric primitives are universal, because each culture adopts, initially by chance, some preferred combination of primitives, which are then internalized as frequent subprograms or program fragments and progressively cement a specific style of geometric patterns – a proposition that generalizes the "child as a hacker" hypothesis (Rule et al., 2020). Second, DreamCoder may explain how simple geometric shapes may be efficiently recognized and used by young children in the absence of much or any training (poverty of the stimulus argument). This is because the top-down system (from programs to shapes) can be used to train the bottom-up system (from shapes to programs) via the use of "dreams", i.e. internally generated training data. In an improved version of DreamCoder, the bottom-up neural network could be repurposed to directly retrieve the most plausible program for a given shape, thus providing a possible mechanism by which humans can quickly identify simple shapes. Finally, another promising aspect of this proposal, which remains to be fully explored, is the possibility of creating reusable abstractions or program templates. While the square, for instance, is not a primitive of the original language, a square-drawing program schema may become abstracted over time, thus allowing the participant to easily understand concepts such as "a square of circles" or "a square twice larger than the previous one", etc. At present, however, for simplicity, such named subprograms are not part of the current language, but solely of the DreamCoder program-induction software.

5.3 Experiment 1: Predicting Geometric Complexity

Does the proposed language have any psychological reality? In experiment 1, we test the simplest prediction of our proposal: the perceived complexity of a shape should be determined by its minimal program length. If humans represent shapes as mental programs, then for tasks involving shape perception and manipulation, the MDL of the shape should predict the difficulty of the task. Additionally, as MDL increases, the time it takes to perform program induction on the shape increases as well, and therefore it should take longer to encode the shape in working memory and to compare it with other shapes.

This prediction should hold only provided that other simple perceptual strategies do not suffice to perform those tasks. In previous work, we found that the perception of quadrilaterals could be based on two systems: a list of symbolic rules akin to those arising from the current language (e.g. right angle, equal sides), and only available to humans; and a classical invariant shape recognition system, well captured by a convolutional neural network model of the ventral visual pathway, and available to both human and non-human primates (Sablé-Meyer, Fagot, et al., 2021). Thus, to properly test the existence of the first system, it is important to cancel out the potential contributions of the second.

Here, we asked participants to memorize a sample geometric shape and perform a delayed match-to-sample task where, after a 2-second delay, that shape had to be selected from an array of 6 possible choices, some of which were perceptually quite similar (**Figure 5.4**A). We measured the choice time, but also the encoding time by letting participants view the sample shape for as long as they wished, holding down the space bar until they were ready to decide; as soon as they released the space bar, the sample shape disappeared, then after a fixed delay, the choices appeared. We hypothesized that both encoding time and choice time would be predicted by MDL.

5.3.1 Methods

Participants

Participants were 125 adults tested online (53 females, 69 males, 3 preferred not to answer; age range 20 to 78, mean and median 44 years
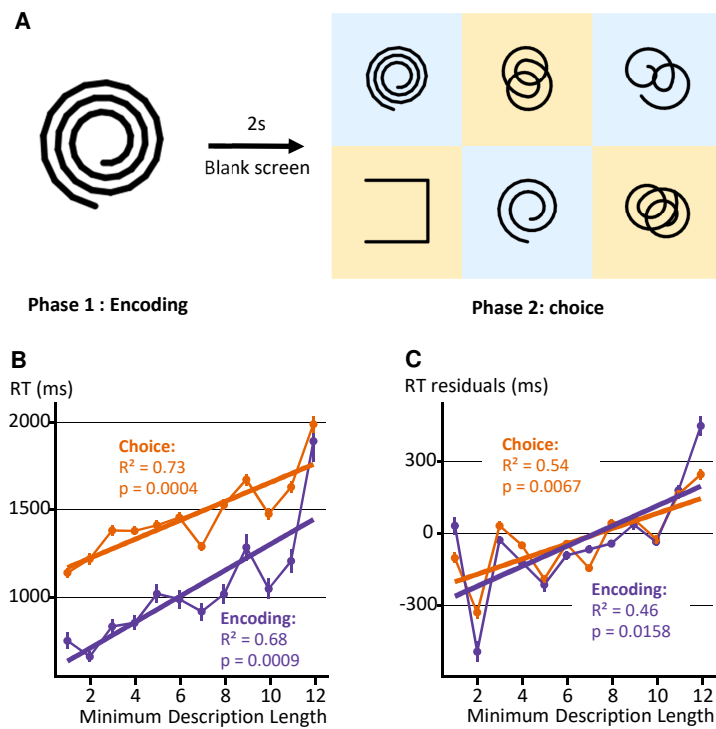
Figure 5.4: **Procedure and results for experiment 1. A, task structure.** On each trial, participants pressed a key for as long as they needed to memorize a shape (encoding phase), then, after a 2-second delay, had to select the corresponding shape among a 2x3 grid (choice phase). Note that there was a size change between phase 1 and phase 2. **B, correlation between MDL and behavior.** Average encoding time (purple) and choice time (orange) are plotted as a function of MDL. Error bars represent one standard error across participants. **C,** same results after regressing out the effect of total luminance.

175

old), recruited and tested online via messages on social media. The task was approved by our university' committee for ethical research (CER-Paris-Saclay-2019-063) and participants gave informed consent.

Procedure and Stimuli

On each trial, we showed participants a sample shape for a variable duration, then an empty screen for 2 seconds, and finally a 2x3 choice screen with 6 shapes. Participants were asked to click on the shape that was shown originally. Participants controlled how long they looked at the reference shape: when they were ready to start a trial, they clicked at the center of the screen (thus centering their mouse), then pressed the space bar on their keyboard. The sample shape was shown for as long as the space bar was depressed. Upon releasing the space bar, the shape disappeared and they were left with a blank screen. Instructions insisted on the self-paced nature of the task: "Keep the spacebar depressed for as long as you deem necessary to remember the shape well". We call the duration of the press the encoding time, as it is an indirect measure of the time that participants needed to encode the target shape. More specifically, we predict that this duration spans over several cognitive steps, some independent from the shape (e.g. motor actions) and at least one that should scale with the MDL, as participants are performing program induction to build a mental representation of the shape they remember.

The choice screen comprised six different shapes that were displayed on an isoluminant blue/yellow checkerboard (**Figure 5.4**A). At that point, participants could click on a shape, which ended the trial. We measured both accuracy and response time, which we refer to as choice time. We predicted that those variables should also impacted by MDL because, after ruling out visually implausible distractors (e.g. based an exceedingly different gray level), participants would have to either generate the target shape from their remembered mental representation and compare the output to the remaining candidates, or to encode the remaining candidates and compare with their memory representation of the target, and both strategies should scale with MDL.

The experiment comprised 6 initial training trials, then 68 trials with 68 unique testing shapes, each appearing only once as a sample. The 68 testing shapes are shown in **Figure 5.2**, while the training shapes were 6 additional shapes sample from the list of shapes with MDL=5

in our language. During training, the distractors were always the same six shapes and participants had each of them as a target once. Piloting indicated that the choice of distractors was crucial for the performance to vary from one shape to the other: if the distractors were too dissimilar from the sample shape, participants learned to press and release the spacebar as fast as they could, knowing that a purely perceptual strategy sufficed for virtually perfect accuracy and therefore bypassing the need to have an accurate mental representation of the shape. To mitigate this strategy, we selected distractors closely matched to each shape. We computed, for each shape, the four closest ones in **Figure 5.2** as defined by two metrics: (i) the average gray level (average pixel value of an image), and (ii) the difference in the vector codes of the shapes within the last layer of a convolutional neural network of object recognition, CORnet (Kubilius et al., 2019). CORnet is a convolutional neural network that figures amongst the top-scoring models of the ventral stream according to BrainScore, "a composite of multiple neural and behavioral benchmarks that score any ANN on how similar it is to the brain's mechanisms for core object recognition" (Schrimpf et al., 2018, 2020). Then, on each trial, we presented on the choice screen, at random locations: (1) the correct target shape; (2) two of the four closest shapes according to CORnet; (3) two of the four closest shapes according to average gray level, different from those selected in (2); and (4) a last shape uniformly sampled among the remaining test shapes. The selection algorithm ensured that all 6 choice shapes differed. The choice of shapes and their placement were fully randomized for each participant, independently within training and within testing.

5.3.2 Results

Overall error rate was very low (1.82%), so we concentrated our analysis on response times. We removed all participants who failed on 5 or more trials, as well as participants whose overall average encoding time or choice time was higher than the group mean plus three standard deviations (9 participants removed in total; 116 remaining). We also removed, for each shape, trials where the encoding time exceeded the average encoding time of that shape plus three standard deviations (and similarly for choice time). This procedure removed 3.8% of the total number of trials.

To test for the predicted effect of MDL on behavior, we performed simple linear regressions on the encoding times and choice times as a func-

tion of the MDL. Both measures were significantly correlated with the MDL of the target shape (**Figure 5.4**B; encoding time: R^2=.68, p<.001; choice time: R^2=.73, p<.001). We also performed between-subjects ANOVAs on the linear effect of MDL (numerical factor) across participants. The main effect of MDL was again highly significant (encoding time: F(1,115)=148.5, p<.001; choice time: F(1,115)=480.7, p<.001).

Since MDL showed a small but significant partial correlation with gray level (R^2=.07, p=.027), we replicated this analysis by first removing the main effect of gray level on response times, then performing a simple linear regression on the residuals. The effect of MDL was again significant (**Figure 5.4**C; encoding time: R^2=.46, p=.016; choice time: R^2=.54, p=.006; between-subjects ANOVAs, encoding time: F(1,115)=118.2, p<.001; choice time: F(1,115)=248.1, p<.001).

Both of these effects were replicated on error rates, despite accuracy being very high. Plots of MDL against either error rates or the residuals of error rates on gray level are shown in **Figure 5.6**. In both cases, MDL significantly correlated with the dependent variable, indicating that participants made more mistakes as MDL increased (raw R^2=.51, p=.009; after taking the residuals on gray level, R^2=.48, p=.012).

We also estimated the quality of our model for shape complexity by contrasting it with many random competing models where the shapes' MDLs were shuffled. The results are summarized in **Figure 5.5**, and allow to derive *p*-values that our model is as predictive as it is by chance, by counting the number of "better" random models generated when shuffling: for both encoding time and choice time, no shuffled model was better, ensuring that p<.0001, and for error rates 7 out of 10000 models were better, therefore p=.0007.

To compare the effects of MDL and gray level, we also performed a multiple-regression analysis with both variables (normalized) as predictors, across the 68 test shapes. For encoding time, both predictors were significant (both p<.0001; R^2=.59; betas = 157.7 (Standard Error [SE]=37.4) for MDL and 272.0 (SE=37.4) for gray level). For choice times, both predictors were also significant (MDL: p=.0003; gray p<.0001; R^2=.53; betas = 108.7 (SE=28.8) for MDL and 182.75 (SE=28.8) for gray level). For error rates, only MDL was significant (MDL: p=.003; gray p=.84, R^2=.13; betas: =.014 (SE=.004) for MDL and -.001 (SE=.004) for gray).

We also controlled for additional visual features (see **Figure 5.7**). For

**Histogram of the # of simulations ending
at given slope out of 10000 samples**
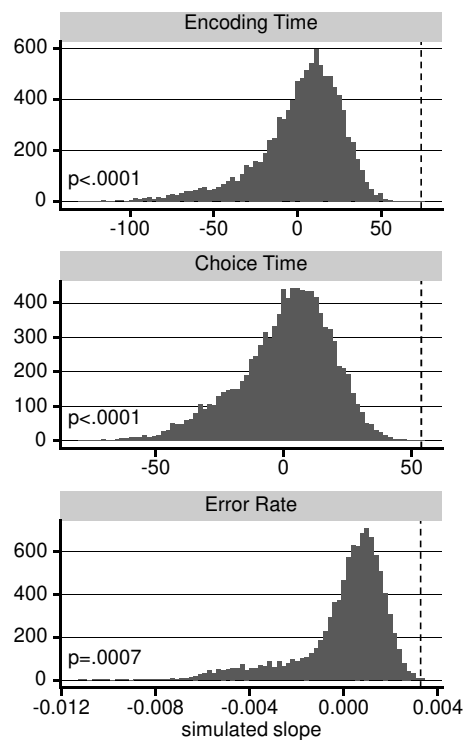
Figure 5.5: Bootstrap over the shapes' MDL. For each dependent variable in separate panels, we show the distribution of the slope of 10000 simple linear regression computed after shuffling the shapes' MDLs. The dashed line, on the right, indicates the slope of our (unshuffled) theory. From this we can derive *p*-values by computing the fraction of random theories that are better than our theory: this yields p=.0007 for error rates, and ensures p<.0001 for both encoding time and choice time as no random permutation fared better.
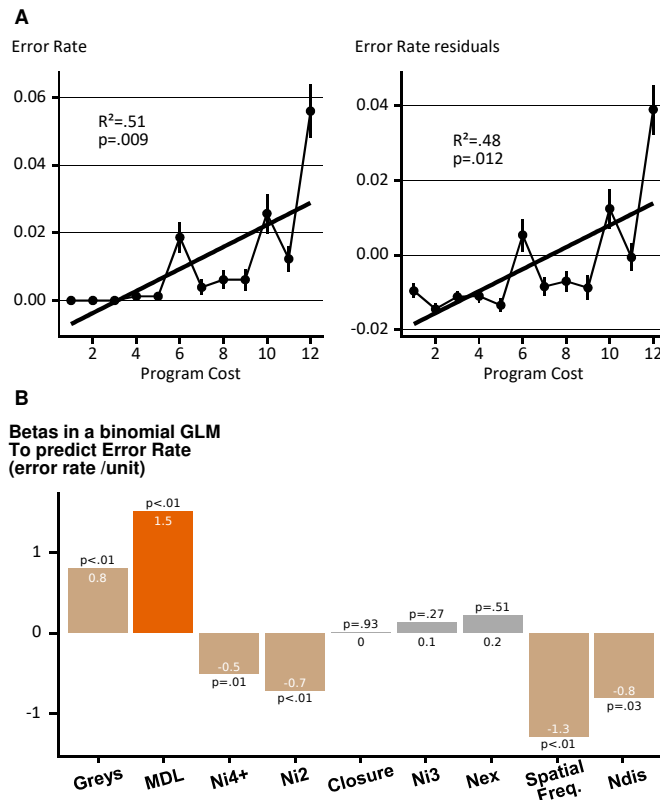
Figure 5.6: A, correlation between MDL and behavior. Left: Average Error Rates are plotted as a function of MDL. Error bars represent one standard error across participants. Right: same results after regressing out the effect of total luminance. **B, Predictive power of the visual features.** We ran a binomial GLM with all the features, normalized for comparison, as well as the MDL, and display the fitted coefficients and their significance, for the error rate. Grayed out predictors were not significant in the regression, and bright orange indicates our regressor of interest, the MDL. Predictors are ranked according to their magnitude when predicting the choice time.
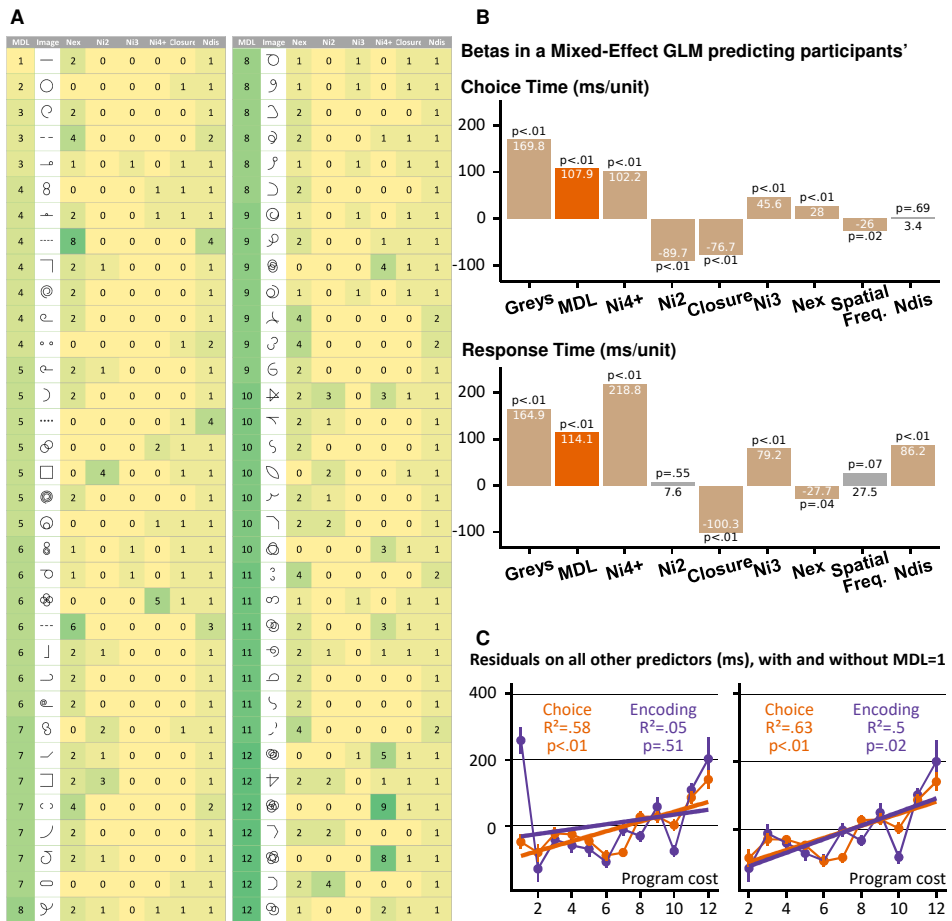
Figure 5.7: **Detailed analysis of Exp. 1 A, breakdown of the properties**. For each shape, describes the manually annotated or computed visual features. **B, Predictive power of the visual features.** We ran a mixed-effect GLM with all visual features, normalized for comparison, as well as the MDL, and display the fitted coefficients and their significance, for both choice time and response time. Grayed out predictors were not significant in the regression, and bright orange indicates our regressor of interest, the MDL. Predictors are ranked according to their magnitude when predicting the choice time. **C.** The simple linear model MDL and either choice or response time, on the residuals of a mixed-effect model with all visual features, after averaging per MDL value. On the right, we display the same regression after removing the only trial where MDL=1.

each shape, we manually counted the number of extremities (variable Nex), intersections where two lines meet (singularities, variable Ni2), intersections where three lines meet (Ni3), intersections where four or more lines meet (Ni4+), disconnected parts (Ndis), and finally the presence of a closed shape or not (closure). To confirm that the effect of MDL did not result from a spurious correlation with these features, we first ran a model comparison between a mixed-effect model with all features, gray level, and the average spatial frequency, with participants as random intercepts, and the same model plus the MDL. The second model was significantly better than the first one for both dependent measures (Likelihood Ratio Test, encoding time $\chi^2 = 106.4$, $p < .001$; choice time $\chi^2 = 169.8$, $p < .001$). The same result held for error rates using binomial linear models (error rate: $\chi^2 = 107.43$, $p < .001$; for this model no random effects were included, as the estimates could not be meaningfully fitted at the level of individual participants). **Figure 5.7** shows the breakdown of each shape's property, as well as a comparison of the predictors associated with each normalized term in the full model. For both encoding time and choice time, the predictor associated with the MDL significantly predicted the dependent variable (and also for error rate, see **Figure 5.6**). As a final, more conservative control, we computed the residuals from the first model for each participant, then averaged those residuals over participants and MDL levels, and examined the correlation of those residuals with MDL. This linear model was not significant for the encoding time (r^2=.04, p=.51), but it was for choice time (r^2=.58, p=.003). Inspection of the residuals show that the "segment" shape, which is the only shape with MDL=1, had a very high residual for the encoding time, while the other seemed to follow the MDL pattern. After removing this outlier item, both encoding time and choice time residuals were now significantly predicted (encoding time: r^2=.5, p=.015; choice time: r^2=.63, p=.003). For completeness, a similar strategy was pursued with error rates, first taking the residuals on a binomial regression on all other predictors, and then fitting a simple linear regression with MDL on those residuals. The resulting trend was similar but not significant (r^2=.23, p=.12).

5.3.3 Discussion of Experiment 1

As predicted, in a delayed match-to-sample task with geometric shapes, participants were influenced by the shape's Minimum Description Length (MDL) in our proposed language. This result held

at individual levels and for all measures of encoding time, choice time and error rate, with comparable effect sizes. The effect of MDL did not trivially arise from spurious correlation with other image properties such as number of parts, intersections, closure, etc. While behavior was also influenced by the perceptual property of average gray level, the effect of MDL remained even when controlling for this low-level effect and all other variables. Yet MDL alone did not fully predict behavior: in a regression with many predictors, several exhibited some explanatory power in our task. This finding supports our prior suggestion that there are two type of strategies available to perform such perceptual tasks with geometrical figures (Sablé-Meyer, Fagot, et al., 2021): one that encodes shapes at a visual level, based on perceptual properties such as gray level, intersections, etc; and one that represents them at an abstract, symbolic level, which is well captured by our proposed language.

While this result is promising, there are many reasons to believe that our specific language proposal is not complete, and that it would be possible to find shapes for which the fit would be poor. For instance, the complex outlines of natural objects (e.g. the contour of a recognizable animal) are completely outside of reach of our proposal, and other theories based on medial axis or shape skeletons would fare much better (Feldman & Singh, 2006). Even within geometric shapes, and as we further examine in the general discussion, there are shapes that the language cannot easily describe (e.g. ovals), or properties it cannot easily encode (e.g. while an equilateral triangle has a low MDL, there is no way to describe "any triangle"). Thus, while experiment 1 tested the specifics of our language, experiment 2 was designed to test regularities that any such reasonable language should verify.

5.4 Experiment 2: Fundamental Laws of Repetition, Concatenation and Embedding

To sidestep the constraints that come with choosing a specific language, we designed an experiment that tests, independently, the three most fundamental aspects of our proposition, namely the existence of operations of *repetition*, *concatenation* and *(nested) embedding* (that is, recursive call to a subprogram). Those operations, available in any modern programming language, have the highest impact on the compressibility of shapes and help decorrelate the length of a

183

program and the amount of "ink" of a shape. They allow a program to be extremely short and yet the shape to be complex, as long as it is highly regular. For instance, from the programs for squares and circles, a single additional instruction suffices to generate a circle of squares, or a square of circles – and thus, this addition should just have an additive effect on MDL.

These observations lead to the following quantitative prediction: in any language for geometric shapes which includes primitives of repetition, concatenation and embedding, the cost of complex shapes should be the sum of the lengths of (1) the program(s) that are being repeated, concatenated or embedded, plus (2) a fixed cost for the instruction itself and, if necessary, its operands. Consider for instance a figure formed by two shapes placed side by side: we predict its complexity to be the sum of the complexity of each shape plus some constant for the concatenation instruction. Likewise, the cost for a repetition of a shape should only depend on the cost of that shape, plus a constant to express the repetition, and a cost for the parameter "number of repetitions". Finally, the complexity of a figure consisting of one shape embedded in another (e.g. a square of circles) should be the cost of each shape plus a constant for the "embed" instruction.

In summary, the following relations should hold, where $Cplx$ stands for "complexity" (at least for $x \neq y$, see below):

1. $Cplx\left(repeat(x, n)\right) = \beta_0 + \beta_1 * Cplx(x) + \beta_2 * Cplx(n)$

2. $Cplx\left(concat(x, y)\right) = \beta_0 + \beta_1 * Cplx(x) + \beta_2 * Cplx(y)$

3. $Cplx\left(embed(x, y)\right) = \beta_0 + \beta_1 * Cplx(x) + \beta_2 * Cplx(y)$

Note that the multipliers β_1 and β_2 should be close to 1 if the length of the program is the only factor that comes into play, but might exceed 1 if additional factors intervene (e.g. interference between the two shapes in working memory). β_0, on the other hand, represent the constant cost associated with the operation at hand and should be strictly positive.

In experiment 2, we therefore replicated the delayed match-to-sample with new stimuli. We selected five base shapes spanning a broad range of predicted complexities, and used them to build new stimuli through repetition, concatenation and embedding, with the goal of testing whether their complexity could be predicted from the complexity of their base shapes. Thus, we designed a total of 60 images that served

as samples in the task : (1) five base shapes, shown in **Figure 5.8**; (2) five repetitions of those base shapes, generated by showing side by side four copies of the base shape; (3) twenty-five concatenations corresponding to all 5x5 pairs of base shapes placed side by side; and (4) twenty-five embeddings of a base shape inside another, generated by drawing the outline of one shape using 8 or 9 copies of the other. Example stimuli are shown in **Figure 5.8**.

In a separate group of participants, we also measured the subjective complexity of the same shapes. We presented participants with those 60 shapes in random order and asked them to evaluate each shape's complexity on a scale from 0-100.

5.4.1   Method

Participants

Participants were recruited via Twitter. One hundred and seventy adults participated in the main delayed match-to-sample task (71 male and 99 females), with a breakdown of 16 participants in the 18-25 age group, 77 in the 25-40, 68 in the 50-60 and 9 in the 60+. Participants were not compensated for participating in this study. An additional 27 adults participated in the subjective ratings **(**15 females and 12 males). Both tasks were approved by CER-Paris-Saclay-2019-063, and participants gave informed consent.

Procedure and Stimuli for the delayed match-to-sample task

The procedure for this experiment was identical to that of our first experiment, and only the stimuli changed. The stimuli were generated from five base shapes that were piloted to vary in encoding and choice time (square, circle, S, sigma, and square root, of respective MDL 5, 2, 10, 15, 13; see **Figure 5.8**). All shapes had similar amounts of ink, or gray level: the square, sigma and square root all had four segments of identical length, arranged differently; the circle matched the square in length; and the S shape comprised two semi-circles and was therefore matched with the circle. Those base shapes were then used to generate the 60 target stimuli for the four experimental conditions: *single shape* (5 stimuli), *repetition* (a string of 4 identical shapes; 5 stimuli), *concatenation* (2 shapes side by side; 5x5=25 stimuli), and *embedding* (an inner shape was presented at the usual size, but in 8 or 9 copies
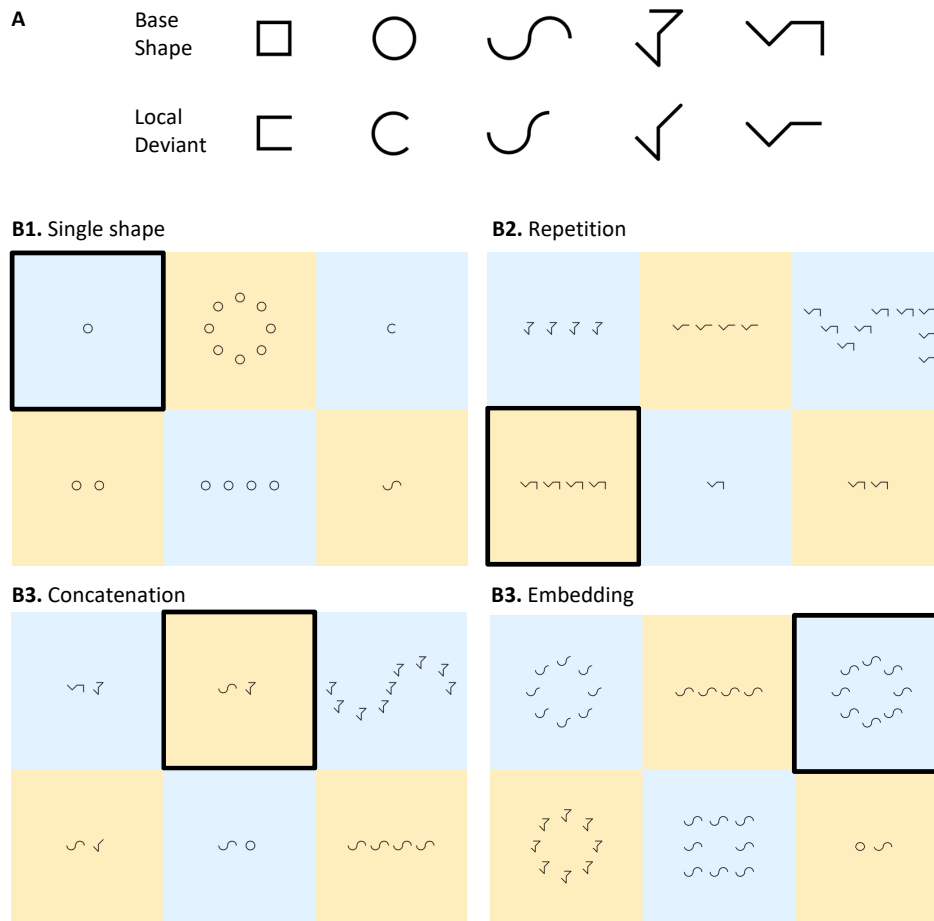
Figure 5.8: **Stimuli used in experiment 2. A**, the 5 base shapes which were presented alone and in combinations, and the corresponding deviant shapes. **B,** Examples of screens presented during the choice phase in the four conditions of the experiment (randomly intermixed): single shape, repetition of a base shape, concatenation of two base shapes, and embedding of a base shape into another. In each case, the sample image which was presented during the encoding phase is highlighted, and the other cells of the 2x3 grid illustrate the diversity of distractors. A distinct group of subjects also gave subjective ratings of complexity for each stimulus.

that formed the outlined of a second, outer shape; 5x5=25 stimuli).

Each of those 60 stimuli served as samples for the delayed match-to-sample task. During the choice period, the sample stimulus was intermixed with 5 distractors, which were generated in order to prevent short-cut strategies and maximize the need for a full identification of the target shape. For each base shape, we designed a local deviant by removing one fourth of the shape (see **Figure 5.8**). For the *single shape* and the *repetition* conditions, the distractors were (1) a distractor from the same condition, but using the local deviant instead of the target, (2) a distractor from the same condition, but using a different shape, (3) three distractors drawn from each of the other three conditions, and sharing at least one shape with the target. For the *concatenation* condition, the distractors were (1) a *concatenation* distractor where one of the two shapes was replaced by its local deviant; (2) two *concatenation* distractors where either the left or right shapes were replaced by a different base shape, (3) one distractor from the *embedding* condition, using the same two shapes as the left and right shapes, assigned randomly to embedded and embedding, (4) one distractor from the *repetition* condition, using randomly either the left or the right shape. Similarly, for the *embedding* condition, the distractors were (1) a distractor from the *embedding* condition, with the same outer shape but the inner shape replaced by its local deviant; (2) two *embedding* distractors where either the outer or the inner shape was replaced by a different base shape, (3) one distractor from the *concatenation* condition, but using the same two shapes as the inner and outer shapes, (4) one distractor from the *repetition* condition, using either the inner or the outer shape. Our logic was that this set of distractors covered a broad range of programs in the proposed language, each with a small local change or "bug" – thus forcing subjects to search for the shape whose description exactly matched the sample. **Figure 5.8**B shows, for each condition, an example of a target and five possible distractors.

In addition to those 60 trials, 10 initial training trials allowed participants to get used to the task flow and to the difficulty level of the memory task. Training trials were generated similarly, but using three different base shapes and the same exact procedure. The experiment proceeded seamlessly from training to testing trials, without any notice.

Procedure and Stimuli for Subjective Complexity Ratings

Upon clicking on a link from the twitter message, participants landed on an experiment designed with jsPsych (de Leeuw, 2015). The experiment started with a consent form as well as a small demographic questionnaire for age group and sex. Then they were presented with instructions for the task: using sliders from 0-100, they had to give a complexity rating to each of the 60 sample shapes. To familiarize them with the type of shapes, the instructions included 24 shapes that did not appear afterward. We highlighted in the instructions that they should focus on trying to be consistent across shapes. Participants were then presented with all 60 shapes in a shuffled order. Participants could freely look at the shapes in any order and change their rating until they were satisfied. The task took a median time of 6:34 minutes to answer (1st quartile, 4:48; 3rd quartile, 9:36.).

5.4.2 Results

None of the subjective rating data were rejected. For the delayed match-to-sample task, the error rate was low (3.39%), and we removed data following the same strategies as in experiment 1 (7 participants and 3.16% trials rejected). Because the accuracy was very high across most conditions, we restricted our analyses to encoding time, choice time and subjective complexity. **Figure 5.9** shows these three dependent variables for each of the 60 stimuli, as a function of the base shape(s) used to generate them.

First, we verified that, in the single-shape condition, all three dependent measures varied across the 5 base shapes we selected (ANOVAs where participants with missing data were removed, all p<.01). For encoding and choice times, the increase was roughly as follows: square and circle were roughly on par, and then the response times increased for the S, sigma and square-root shapes, in this order. Note that this order is close to, but not strictly identical to the MDL ordering, which was circle, square, S shape, square root and sigma. The subjective ratings followed a noisier profile, but still ranking the last two shapes as more complex than the first 2, with an overall ordering close to the one predicted by MDL. A one-tailed student test on the distribution of slopes across participants in a simple linear regression with MDL confirmed that MDL significantly correlated with all three measures (encoding time p<.001, average r² across subjects r^2_{avg}=.30; choice time p<.001,
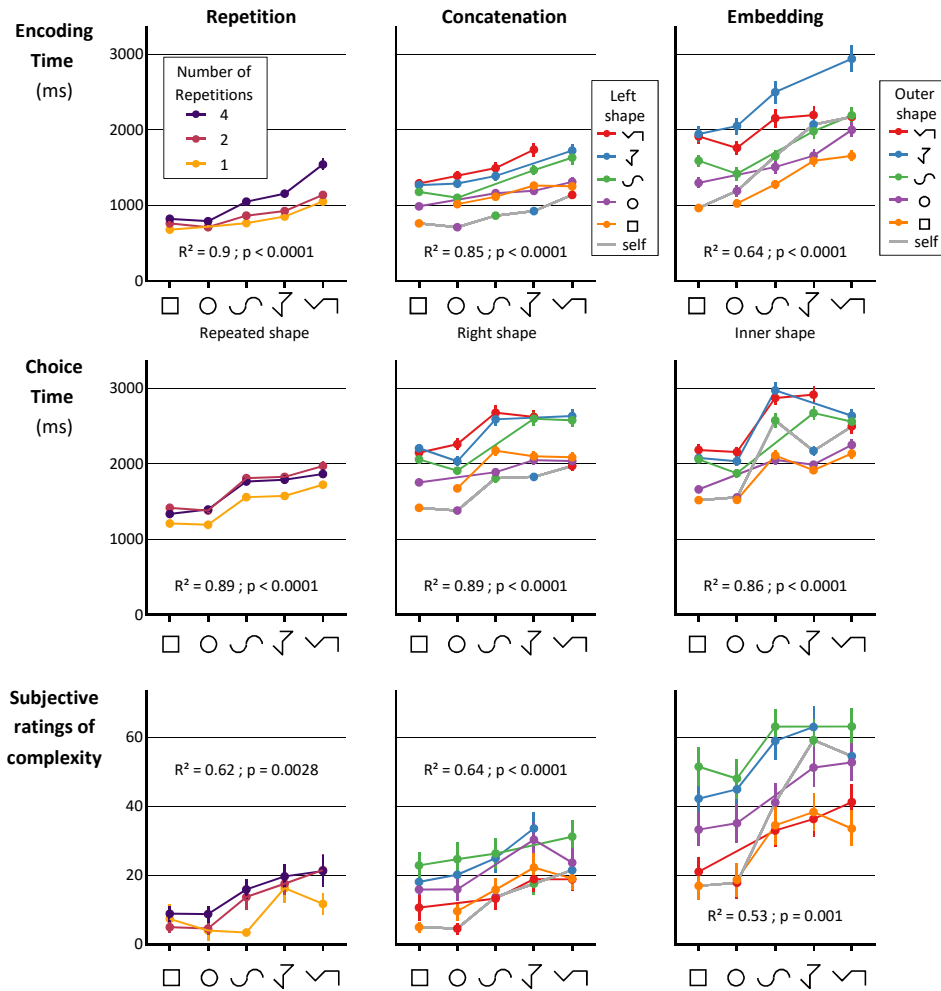
Figure 5.9: **Results of experiment 2.** The three behavioral measures (encoding time, choice time, and subjective complexity) are shown for the 60 different stimuli, grouped into repetition (left), concatenation (middle), and embedding (right). The conditions are sorted as a function of base shape complexity. Note that the single-shape condition is shown in the repetition panel (number of repetitions = 1), and that the data from the concatenation of two identical shapes appears twice (as number of repetitions = 2 in the left panels, and as dots connected by a gray line in the middle panels). Similarly, the data for "self-embedded" shapes (e.g. a square of squares) appear with a gray line in the right panels). In each panel, we show the coefficient of determination and statistical significance of the associated model (equations [1-3] in the main text). Error bars indicate standard errors across subjects.

r^2_{avg}=.40; subjective p=.004, r^2_{avg}=.55).

We then examined whether, when those shapes entered into more complex stimuli, the complexity still varied in a similar way, predictable from equations [1-3]. In the *repetition* condition with 4 identical shapes, a significant, monotonic profile of response was observed for all three dependent measures (ANOVAs where participants with relevant missing data were removed, encoding time F(4,576)=53.88; response time F(4,576)=36.87; subjective rating F(4,104)= 6.63; all p<.001). Similarly, we analyzed just the stimuli of the *concatenation* condition where two identical shapes were placed side by side, thus corresponding to a repetition of 2. Again, for all dependent measures, a significant, nearly monotonic increase was seen across the 5 base shapes (participants with missing data removed; encoding time, F(4,576)=32.20; choice time F(4,576)=41.01, ; subjective F(4,104)=15.29; all p<.001). Importantly, the curves for single shape, 2 repetitions and 4 repetitions were nearly parallel to each other, as predicted by our equations for MDL. To test this idea, we entered all 3 conditions into a mixed-effect linear model with two main factors and their interaction, i.e. the repeated shape (5 levels), and the number of repetitions (numerical factor spanning 1, 2 or 4 repetitions). As shown in **Figure 5.10**, we found main effects for each measure, in agreement with equation [1]. There was no significant interaction with the number of repetitions for both subjective rating and choice time, in agreement with equation [1], but a significant interaction was found for encoding time. To measure the relative importance of the interaction term, we estimated the amount of the final model's variance due to the interaction term (last columns of **Figure 5.10**) by comparing the marginal r^2 (Nakagawa et al., 2017) of both the full model and the model without the interaction. For the repetition condition, this indicated that even when the interaction term is significant, it accounted for less than 10% of the explained variance in all three dependent measures, and was therefore dominated by the main effect of shape and number.

For the concatenation and embedding conditions, as shown in **Figure 5.9**, encoding time, choice time, and subjective complexity also increased with each of the two shapes involved (left/right or inner/outer). Equations [2] and [3] predicted that these effects should be similar to the single shape condition and should not interact. **Figure 5.10** shows the results of mixed effect linear models, but now

Repetition		Shape		Number		Interaction		r² inter.
	Subjective	$F_{4,369}=2.95$	p=.020	$F_{1,369}=11.56$	p<.001	$F_{4,369}=0.88$	n.s.	0.05
	Encoding	$F_{4,2215.25}=4.66$	p<.001	$F_{1,2215.10}=185.21$	p<.001	$F_{4,2215.17}=14.86$	p<.001	0.08
	Choice	$F_{4,2215.41}=28.31$	p<.001	$F_{1,2215.23}=35.75$	p<.001	$F_{4,2215.31}=0.76$	n.s.	<.01

Concatenation		Left shape		Right shape		Interaction		r² inter.
Including	Subjective	$F_{4,624}=22.16$	p<.001	$F_{4,624}=16.71$	p<.001	$F_{16,624}=3.83$	p<.001	0.28
self-concatenated	Encoding	$F_{4,3796.63}=52.88$	p<.001	$F_{4,3796.54}=45.02$	p<.001	$F_{16,3796.56}=27.67$	p<.001	0.53
shapes	Choice	$F_{4,3797.13}=76.72$	p<.001	$F_{4,3796.97}=58.94$	p<.001	$F_{16,3797.01}=25.77$	p<.001	0.43
Excluding	Subjective	$F_{4,494}=25.85$	p<.001	$F_{4,494}=20.80$	p<.001	$F_{11,494}=0.45$	n.s.	0.03
self-concatenated	Encoding	$F_{4,3004.67}=57.78$	p<.001	$F_{4,3004.59}=50.76$	p<.001	$F_{11,3004.66}=1.85$	p=.041	0.05
shapes	Choice	$F_{4,3005.32}=78.42$	p<.001	$F_{4,3005.18}=63.97$	p<.001	$F_{11,3005.31}=2.16$	p=.014	0.05

Embedding		Outer shape		Inner shape		Interaction		r² inter.
Including	Subjective	$F_{4,624}=66.58$	p<.001	$F_{4,624}=28.54$	p<.001	$F_{16,624}=0.90$	n.s.	0.03
self-embedded	Encoding	$F_{4,3786.68}=119.23$	p<.001	$F_{4,3786.67}=61.62$	p<.001	$F_{16,3786.70}=3.64$	p<.001	0.07
shapes	Choice	$F_{4,3787.10}=115.12$	p<.001	$F_{4,3787.08}=118.95$	p<.001	$F_{16,3787.12}=6.74$	p<.001	0.1
Excluding	Subjective	$F_{4,494}=55.57$	p<.001	$F_{4,494}=24.90$	p<.001	$F_{11,494}=0.42$	n.s.	<.01
self-embedded	Encoding	$F_{4,2996.89}=103.49$	p<.001	$F_{4,2996.79}=55.88$	p<.001	$F_{11,2996.82}=1.20$	n.s.	0.02
shapes	Choice	$F_{4,2997.45}=111.20$	p<.001	$F_{4,2997.29}=110.49$	p<.001	$F_{11,2997.33}=3.74$	p<.001	0.05

Figure 5.10: **Mixed effect modeling of data from Experiment 2** The table shows the statistics of mixed effect models applied to our three main conditions (Repetition, Concatenation and Embedding) and to our three dependent variables (Subjective Rating, Encoding time and Choice time). In each case our model had two main effects and their interaction, plus a single random effect of the participant. P-values were computed using the Kenward-Roger approximation for degrees of freedom. For Repetition, the main effects were the repeated shape (5 levels), and the number of repetitions (numerical factor in 1, 2 or 4, respectively for the shape alone, for the concatenation of two identical shapes, and for the repetition condition). For Concatenation, the main factors were the shape on the left and the shape on the right. For embedding, the main effects were the outer/embedding shape, and the inner/embedded shape. For Concatenation and Embedding, we ran the model a second time after removing the stimuli in which the same base shape was used twice (e.g. two squares, or a square of squares). The rightmost columns measure the proportion of the full model's r² (fixed-effects only, approximated using the marginal r² from (Lüdecke et al., 2021; Nakagawa et al., 2017)) that is lost when removing the interaction term, to approximate the explained variance due to the interaction terms.

with the shapes as two 5-level factors. Starting with the Concatenation condition, for all 3 dependent variables, we found significant effects of both the shape on the left and the one on the right, as predicted. Importantly, the interaction term was also significant. However, there was a very simple explanation: the 5x5 design matrix for concatenation included a diagonal of 5 stimuli in which the left and right shapes were identical. In this case, our theory predicts that these stimuli should be compressible using the repetition instruction, and therefore easier to perceive than other concatenation stimuli. This is exactly what we found: as seen in the middle panels of **Figure 5.9** (where, for simplicity, those data points are connected by a gray line), the repeated stimuli stood out as faster and subjectively less complex than the corresponding concatenation stimuli. Correspondingly, the fraction of explained variance due to the interaction term in the full model was high (ranging from 28% to 53%) across all three measures. However, when we removed the conditions where both shapes were identical from the mixed-effect model (**Figure 5.10**), the proportion of explained variance collapsed to 5% or less for all dependent measures, even when the term remained significant, in accordance with equation [2].

Finally, we ran the same analysis on the *embedding* condition, using both embedded and embedding shape as the 2 factors, plus their interactions. Again, both main effects were significant on all 3 dependent measures, as predicted by equation [3]. However, as in the concatenation condition, the interaction term had a significant effect on encoding time and choice time, which our additive equations did not predict – albeit these interaction terms' proportion of explained variance were low, ranging from 3% to 10%. Inspired by our analysis of concatenation, we plot separately the "self-embedding" trials, in which the same shape was used at the inner and outer levels (e.g. a square of squares, a circle of circles, etc). In **Figure 5.9** (right panels), we can see that those data points again yielded lower values than the others (i.e. lower subjective complexity, faster encoding and choice times). When we excluded those self-embedding trials from the mixed-effect model, the proportion of explained variance associated with the interaction term were reduced, indicating that the model's explained variance was dominated by the main effects, especially in the absence of self-embedded shapes. Although those observations were not predicted, they are minor compared to the main effects, and can easily be accommodated: it appears that the

mental representation of a "square of squares" involves a saving, because both the embedded and the embedding shapes are identical, and thus presumably parts of the mental program are reused twice.

We further tested the predictions of equations [1-3] using General Linear Models (GLMs). Those equations imply that we should be able to accurately reconstruct the complexity of composite shapes from that of the five base shapes. To test this, we fitted generalized linear models on each of our dependent variables after averaging data across participants for each item (**Figure 5.9**; all fitted values are provided in **Figure 5.1**). First, we modeled the repetition conditions (base shape, two shapes, and four shapes) by predicting, for each dependent measure, its value for a given trial as a linear function of its value in the single shape condition and the number of repetitions. This model was significant for each of our three dependent variables, and all predictors were significantly different from 0 (all p<.05; subjective rating, R^2=.62; encoding time, R^2=.90; choice time, R^2=.89). The fitted coefficients for the shape term were all close to one (non-significantly different from 1 for both subjective complexity and choice time, and only slightly larger than 1 for encoding time), suggesting that base shape complexity was directly reflected in the complexity of the repeated shape.

Similarly, we modeled the concatenation condition by predicting the complexity of a trial with a linear combination of the complexity of the left shape, the complexity of the right shape, and a dummy variable (termed IsSelf) for whether the left and right shapes were identical (as these trials are instances of repeat and should have a lower perceptual complexity). This model was significant for each of our three dependent variables, and all predictors were significantly different from 0 (see Table 2; all p<.05; subjective, R^2=.64; encoding, R^2=.85; choice, R^2=.89). The coefficients for both left and right shapes were not significantly different from 1 at the p<.05 level, indicating that both contributed equally and directly, as predicted from equation 2. The IsSelf predictor was always significantly negative, indicating a saving when both shapes are identical.

Finally, we applied the same model to our embedding condition. Again, a good fit was found, with significant effects of both shapes for all dependent measures (see Table 2; all p<0.05; subjective, R^2=.53; encoding R^2=.64; choice R^2=.86). In this between-items analysis the impact of self-embedding did not reach significance but the trend was present (subjective rating p=.33; encoding time p=.14; response time p=.06).

This is compatible with the low proportion of explained variance associated with the interaction term described in the mixed-effect models: in order to reach significance, a less sensitive analysis may not indicate that a given small effect significant despite identifying the correct trend. This time, the slopes tended to be higher than 1, although this reached significance only in the case of encoding time for the outer shape. We also observed a trend towards a larger influence of the outer shape compared to the inner shape, consistent with the outer shape's greater visual impact, but again this effect was only significant in a single dependent measure (subjective ratings). Those minor trends notwithstanding, the main finding is that the data supported equation [3]: even when we made the visual pattern much more complex by embedding one shape inside another, thus creating for instance a circle of squares, the final complexity was still only a linear function of the complexity of the individual shapes.

Table 5.1: Coefficients resulting from the General Linear Modeling of Experiment 2. The table shows, for each shape type (repeat, concatenate, and embedding), each dependent variable (subjective rating, encoding time or choice time), and for each variable in the model (in columns), the value of the regression slope, associated standard error, and whether the effect is significantly different from 0 (*, $p<.05$; **, $p<.01$; ***, $p<.001$).

Repeat		Shape	Number of repetitions	
	Subjective	0.88 ± 0.23**	1.98 ± 0.89*	
	Encoding	1.37 ± 0.15***	87.69 ± 16.45***	
	Choice Time	1.05 ± 0.11***	47.31 ± 19.1*	
Concatenate		**Left shape**	**Right shape**	**IsSelf**
	Subjective	0.72 ± 0.2**	0.67 ± 0.2**	-8.33 ± 2.37**
	Encoding	0.92 ± 0.17***	0.89 ± 0.17***	-433.15 ± 55.29***
	Choice Time	0.96 ± 0.12***	0.81 ± 0.12***	-521.96 ± 63.28***
Embedding		**Outer shape**	**Inner shape**	**IsSelf**
	Subjective	1.84 ± 0.43***	0.95 ± 0.43*	-5.26 ± 5.27
	Encoding	1.94 ± 0.45***	1.83 ± 0.45***	-222.82 ± 148.07
	Choice Time	1.29 ± 0.16***	1.23 ± 0.16***	-169.22 ± 85.43

5.4.3 Discussion of Experiment 2.

Participants' behavior matched the additive properties predicted by equations 1 through 3: the complexity of complex patterns could be decomposed into the sum of the complexity of constituents. This property was observed in three different metrics that we collected as proxies for the complexity of the mental representation of shapes: subjective complexity ratings, encoding times and choice times in a delayed match-to-sample task.

More specifically, and following the equations, the following three results were verified: [1] the complexity of the mental representation of *n* identical shapes is the complexity of the mental representation of the shape, together with a cost that increases with *n*; [2] the complexity of the mental representation of two different shapes side by side is the sum of their respective complexities; and [3] the complexity of the mental representation of a shape drawn in outline using several small copies of a different shape (which we called embedding) is predicted by the sum of their respective complexities. The latter finding is the most interesting, as intuition alone might have predicted a product operation – after all, the overall pattern comprises as many copies of the inner shape as needed to draw the outer shape. However, the prediction from the language of thought perspective is clear enough: drawing a square of circles is not much more complex than drawing a square itself – it merely requires stopping the square program at regular intervals to call a subprogram for drawing a circle, and in first approximation, such embedding only has a linear effect on total complexity. Of course, this is only true in first approximation – our square of circles, for instance, comprised additional circles not only at the vertices of the square, but also in the middle of its sides, thus requiring a slightly more complex square-drawing program. Such subtleties, which require further investigation, may explain why the slope measuring the impact of the outer shape on the complexity of the overall pattern tended to be larger than 1 for embedded shapes (see Table 2).

Two other salient effects emerged, one which could be predicted from equations [1-3] and another which could not. First, when concatenating two identical shapes, the resulting shape can be either described as a repetition or as a concatenation – but our language of thought predicts that programs involving repetition are shorter, and therefore that the complexity of a pair of shapes should be lower than predicted by concatenation alone. Our data supports this prediction: concatenations of two identical shapes have a lower complexity than predicted by the sum of the complexities of the two shapes, each with their respective coefficients, indicating that identical shapes induce a saving. Second, unexpectedly, the same phenomenon occurred in the embedding condition when a shape was outlined using a smaller version of itself ("self-embed" trials, e.g. a square of squares). Once again, the data points to the lower complexity of those trials, compared to those using two different shapes. Such a saving is not captured by our equations. It suggests that in the mental representation of e.g. "a square of squares",

195

the two squares may be represented by a single mental program or at least by some degree of sharing of working memory resources. This interesting finding supports the idea that human mental representations allow for named subprograms, recursive calls, or higher-order functions over functions, which are not fully captured by the present language.

Overall, Experiment 2 goes beyond experiment 1 in showing that, over and above the specific complexity predicted by the particular geometrical language we proposed, there are several properties of additivity that must be satisfied by any such language, and that these properties are true of the human working memory for geometric shapes.

5.5   General Discussion

Previous research has emphasized that all humans inherit, from evolution, core knowledge of space and number that they share with many other animal species (Dehaene et al., 2006; Feigenson et al., 2004). Here, we propose that, in humans, those core systems can also be recombined using a language of thought in order to form complex mental programs. As a result, humans are able to form complex, compositional thoughts such as "three parallel lines", "repeat a pattern four times", or "arrange some circles in the shape of a square".

In the present work, we argued that such combinatorial mental representations underlie human perception and working memory for geometric shapes, and we put this hypothesis to several tests. First, we proposed a concrete language inspired by observations on prehistoric and ethnographic geometric patterns (featuring abstract patterns, right angles, parallel lines, circles and spirals) as well as elements from the core-knowledge literature (number sense). We tested this language in Experiment 1 and showed that it could predict the behavior of participants in a shape memory task, above and beyond lower-lever perception mechanisms. In Experiment 2, we further showed that theoretically motivated additive equations for the complexity of composite shapes characterize humans' subjective ratings and objective behavior in a delayed match-to-sample task, thereby constraining the properties that must be satisfied by any proposition for a language of thought for geometric shape.

The proposed theory of shape perception assumes that humans can

efficiently infer a mental program from a visual percept, a problem known in computer science as "program induction". The idea that humans infer mental programs was previously shown to successfully account for human concept learning (Lake et al., 2015; Rule et al., 2020), including hand-drawn sketches (Ellis et al., 2018). However, program induction is a computationally challenging problem. Enumerating all possible programs until a match is discovered is not a plausible strategy, as the search time would scale exponentially with the size of the program. We show how recent work from the program-induction literature helps tackle this problem by using DreamCoder (Ellis et al., 2021) to find the best representation for several shapes. DreamCoder uses a bottom-up neural network to speed up the search for the relevant program. In DreamCoder, the network is trained to map each visual shape onto biases that accelerate the search for the relevant program. Future versions could incorporate direct mappings from shapes to programs, thus leading to the immediate recognition of shapes close enough to the training set. What is remarkable about this idea is that the system does not need any external training data (although successfully solved problems will be used when training, much like replays, in addition to "dreams"): it can generate its own supervised learning dataset by sampling programs, executing them to produce the corresponding shapes, and then training a neural network to perform the backward inference from shape to program. The notion of "inner training" is an interesting metaphor for how humans may explore, in a purely mental manner, the domain of geometry and discover interesting properties on their own, without external inputs. This is possible if we assume, as René Descartes did in the introductory citation, that our mind already has the ideas, at least in the form of a large space of potential mental programs.

To further accelerate its search, the DreamCoder algorithm looks for subprograms that are reused across several shapes. This mechanism is useful to go from simple shapes to more and more complex ones, as each new success offers the possibility of discovering new abstractions. This mechanism changes the topography of the search space by bringing certain shapes (those that leverage the discovered abstractions) closer to the effectively searchable threshold for program induction: while all shapes remain accessible, some of them become much simpler as they can be expressed more succinctly using the discovered abstractions. This mechanism could explain cultural drifts in the style of geometrical patterns, whereby a given human culture focuses on cer-

tain shapes and their variants, thus producing, over time, a variety of similar-looking patterns.

The core of our proposal builds upon Leyton's (Leyton, 1984, 2003) seminal proposal of a generative theory of shapes. Leyton's theory stipulates that all shapes are constructed in a bottom-up fashion by a sequence of operations, called "control groups", starting from a single point. For example, the mental representation of a cube would be the extrusion across the z-axis of a square, which itself would come from turning a segment 4 times around a central point. The segment itself would be built by translating the starting point along an axis. The unpublished experiments briefly mentioned by Leyton (Leyton, 1984) seem to have probed those hypotheses only indirectly, for example by asking participants to perform intuitive judgements about the stability of certain shapes. While less general and confined to two dimensions, our proposal is supported by direct empirical tests of the mental complexity of shapes.

Earlier work by Leeuwenberg (E. L. Leeuwenberg, 1971) also introduced a language for shapes, particularly focused on the idea of nested repetitions with variations and the concept of continuous integration as a complement of discrete repetition capable of tracing curves. However, Leeuwenberg's language only satisfied one direction of the intended correspondence between mental and linguistic complexity: the language was such that low-complexity mental representation corresponded to a short program, but the converse was not true as some short programs generated shapes that were not easily parsed by humans. Nevertheless, the behavioral results he reported are in line with ours: he found the length of the shortest program is a good predictor of subjective ratings of complexity, as well as objective performance in shape copying and other similar tasks (Boselie & Leeuwenberg, 1986; E. L. Leeuwenberg, 1969, 1971; see also Restle, 1970, 1973).

More recently, several articles highlighted the importance of the notion of repetition with variation in the human perception of geometric and auditory sequences (Amalric et al., 2017; Piantadosi et al., 2016; Planton et al., 2021; Roumi et al., 2021; Simon, 1972). The language first proposed by Amalric et al. (2017) included two distinct notions of repetitions with variations, either changing the starting point with each repetition, or changing a parameter with each repetition. Our geometric language generalizes this idea to the case of continuous curve tracing. Our tracing primitive can be considered a particular case of

infinitesimal repetition with variation: at each time step, parameters are updated for the computation of the next time step. Importantly, we did not allow for arbitrary infinitesimal repetition like what is present in the Logo language (Abelson et al., 1974), where at each time step arbitrary computations can occur. While such computations were required in the original logo in order to draw curves by infinitesimally changing the heading at each time step, it also opened the possibility of short programs having extremely complex outputs. By limiting ourselves to simple, linear variations of speed or turning angle over time, our language fits with the universal presence of a limited set of shapes (mostly lines, circles and spirals) in human geometric patterns.

Another, more speculative aspect of our proposal is that a recursively compositional language of thought for geometry is unique to humans. As also proposed by others, only humans would possess a recursive compositional capacity (Dehaene et al., 2015; Fitch, 2014; Hauser et al., 2002; Penn et al., 2008). Thus, an important goal for the future is to explicitly test the perception of the present shapes in non-human primates. Our recent work (Sablé-Meyer, Fagot, et al., 2021) shows that, even for shapes as simple as quadrilaterals, human behavior differs strikingly from baboons and is characterized by a symbolic geometrical regularity effect: in humans only, regular quadrilaterals such as squares, rectangles or parallelograms (which can be compressed in the present language) are easier to perceive than less regular ones. Baboon behavior was not random either, it but could be captured by existing neural network models of the ventral visual pathway for object recognition. Accounting for behavioral data from preschoolers and human adults without formal education, however, always required both the object recognition model and the symbolic geometry model. Thus, two strategies seem to be available for geometric shape perception: a purely visual strategy, available to both baboons and humans, and a symbolic geometry strategy, putatively available only to humans. It is important to take this dual-route model into consideration when testing the present idea: if the stimuli can be too easily discriminated by object recognition alone, then humans may not make the effort to encode them symbolically, in the proposed language of thought, and MDL may cease to determine performance (as we observed in pilot data).

5.5.1 Limits of our language

The language we proposed is severely limited in scope: it only applies to the perception and drawing of geometric shapes. While it could plausibly draw a human stick-figure that might meaningfully compare with, say, children's drawings, tracing the contours of a human body or any other realistic shape is outside the reach of our proposal. This aspect separates the present work from related research on contour perception (Feldman, 2001; Feldman & Singh, 2005; Wilder et al., 2016), which offers theories of smooth continuous contours for natural objects or artificial blobs. It also sets our work apart from a longstanding literature on the determinants of shape perception, such as skeletal representation or medial axis extraction (Ayzenberg et al., 2019; Ayzenberg & Lourenco, 2019; Blum, 1973; Firestone & Scholl, 2014; Lowet et al., 2018). In the future, it would be interesting to examine whether the two approaches can be unified, as they share some converging ideas, such as the use of probabilistically generative models for finding a candidate skeleton of a shape (Feldman & Singh, 2006) or finding subgroups within a complex shape (Froyen et al., 2015), but also some challenging experimental results, such as the automatic extraction of medial axes even for triangles or rectangles (Firestone & Scholl, 2014). In a sense, the two approaches are complementary: there are many visual domains for which our language is not well suited, e.g. predicting the complexity of the contour outline of an animal, where a skeleton-based approach is superior; but our proposal is intended to explain another domain, abstract geometric shape, and identify the core tools required to account for their perception and production in humans.

Simple arguments show that the proposed language can generate most of the geometric shapes that humans find simple and that are frequently attested in human cultures as well as in the history of geometry. Any regular polygon, for instance, such as an equilateral triangle, can be generated by a program similar to the square, but using fractions of a right-angle turn; the pseudocode would read: `Repeat p times { Trace ; Turn(angle=4/p ) }`. Stars with various numbers of branches can be similarly generated. Less regular polygons, such as a rectangle approximating the proportions of the famous golden section, can be generated using fractions (e.g. 5/3 or 8/5). Symmetrical patterns and friezes such as 卐 or 回回 arise naturally from recursive combinations of repetitions and alternations.

Finally, using concatenation or embedding, these patterns can be combined to generate, for instance, a pentagram (star inside a circle), a circle of circles, etc.

Nevertheless, some simple figures remain difficult to generate with a short program. A trapezoid, for instance, comprises two parallel sides interrupted by two arbitrary segments. Drawing a trapezoid in our language requires a turn by an arbitrary angle α, followed by a second turn by 2- α to restore parallelism. However, our language does not have variables that could store the value α, and hence does not find this shape simpler than an arbitrary quadrilateral with turns α and β. In general, our approach is unable to encode a partial regularity inside an otherwise arbitrary figure. The addition of local variables could address this limitation, but exploration of this idea suggested that straightforwardly adding variables has a high cost: while they allow one to express otherwise hard to describe simple shapes, they also make very complex shapes easy to describe, an undesirable feature that is hard to keep in check.

There are also shapes for which our language proposes implausible programs. For instance, the minimal program that draws a "+" shape repeats four segments starting from the center, instead of drawing the shape using two intersecting segments. As for continuous shapes, our program cannot account for some simple shapes such as the ellipse or parabola. These shapes might be better represented as visual transformations of the outputs of another program (e.g. compressing a circle to get an ellipse). Such a dual-mode system, combining the generative and transformative capacities of mental imagery, has been proposed by others (e.g. Kosslyn, 1980; Leyton, 2003; Shepard & Cooper, 1982). The addition of a buffer in which mental operations such as rotations or shearing could be applied would be an important addition to the present work.

Although our model was not designed with Gestalt configurations in mind, it might capture some, if not all, of the perceptual and memory savings associated with Gestalt stimuli. For instance, shorter programs would be allocated to displays in which shapes are aligned, repeated, or otherwise obey "good continuation" rules. Some configurations based on closure may be captured when the phenomenon requires alignments of parts of the stimuli, as those may be generated with shorter programs than the counterparts that don't induce closure (e.g. a Kanisza triangle versus a "misaligned" alternative,

201

which is costlier to express). However, as noted in (Kanizsa, 1976), "[Geometric regularity] is not a necessary condition for the formation of subjective surfaces and contours. Amorphous shapes are possible and irregular figures can generate contours." Such examples, as well as figure-ground configurations, seem out of the scope of the current proposition, but have been given Bayesian accounts, see for example (Feldman, 2001; Goldreich & Peterson, 2012).

Finally, in the present framework, each program must contain all the instructions to generate a given shape in every detail. An unfortunate consequence is that a large part of this program is repeated when a subject learns a new but related shape. For instance, the triskelion (⚛) includes a detailed subprogram for a spiral, although the same program may have been used to describe other shapes. The DreamCoder algorithm sidesteps this by inventing abstractions over several iterations, but this feature is not part of the language itself. It may be useful to add to our language a capacity for named subprograms, such that the spiral program would be encoded just once and evoked by a single call to its procedure name ("spiral"), thus further increasing the human memory compression rate. The idea that humans construct complex geometrical concepts by progressively developing a vocabulary of nested, increasingly complex ones, is an appealing view of mathematical development that should be the focus of future work.

5.5.2 Future Directions

We see several promising research directions to go beyond the present work. A first one, drawing inspiration from Kosslyn, Shepard and Leyton's work, would allow for further mental manipulation of the outputs of the present language, thus performing operations such as deformations, rotations, or even extrusions as a post-processing step. This would open up to the modeling of 3-D shape, for instance using rotation around a fixed axis. A second would be develop a library of increasingly complex and reusable named subprograms, such as "square", "spiral", etc. Finally, a third possibility would be to integrate ideas from a different programming paradigm. The language we used is imperative in nature: its programs describe, in every detail, the sequence of operations needed to draw a shape. However, not all programming languages work that way. Logic programming languages such as Prolog describe the logic of the computation and its constraints rather than the details of its execution flow. Such constraint-based programming

may be closer to how humans think, particularly in the mathematical domain. The canonical definition of a circle, for instance, involves a constraint (equidistance from a center point) rather than a generative program (turn by a fixed curvature, as in the present language). Integrating both declarative and imperative elements may provide a better account of specific shapes, such as trapezoids or generic triangles, which only possess some properties (e.g. two parallel sides, or three sides) while leaving the other details unspecified. Some geometrical shapes could thus be defined in terms of properties that they satisfy, others in more detailed imperative instructions, and both could be reused in control structures such as embed or repeat. Meanwhile, we surmise that the present proposal merely brings us one step closer to understanding how abstract mathematical and geometrical objects are mentally encoded.

5.6 Addendum Post-Publication

A few additional results came out since the submission of the article that constitutes **chapter 5**, and while they are not final yet, they are worth mentioning. In pursuing the idea that access to the symbolic representation of geometric shapes, the first experiment of this chapter was adapted to allow baboons to take it. Minor modifications were required, and subsequently we retested humans with the new variation, providing us with a replication.

First, we tried to replicate the exact same task with a slightly different spatial organization of the target and the distractors in baboons. However, baboons did not utilize at all the possibilities offered by the self-paced looking time, and instead all subjects looked as briefly as they could (average 100ms) to move on to the next step, making the experiment substantially different: with very little time to study the shape, the selection amidst distractors becomes much harder and may be driven by different factors. Therefore, we also replicated the experiment with a forced, 1000ms looking time; we therefore distinguish the "fast" condition and the "slow" condition. Separately, replicated both the slow and the fast condition in new groups of adults tested online.

5.6.1 Experiment 3
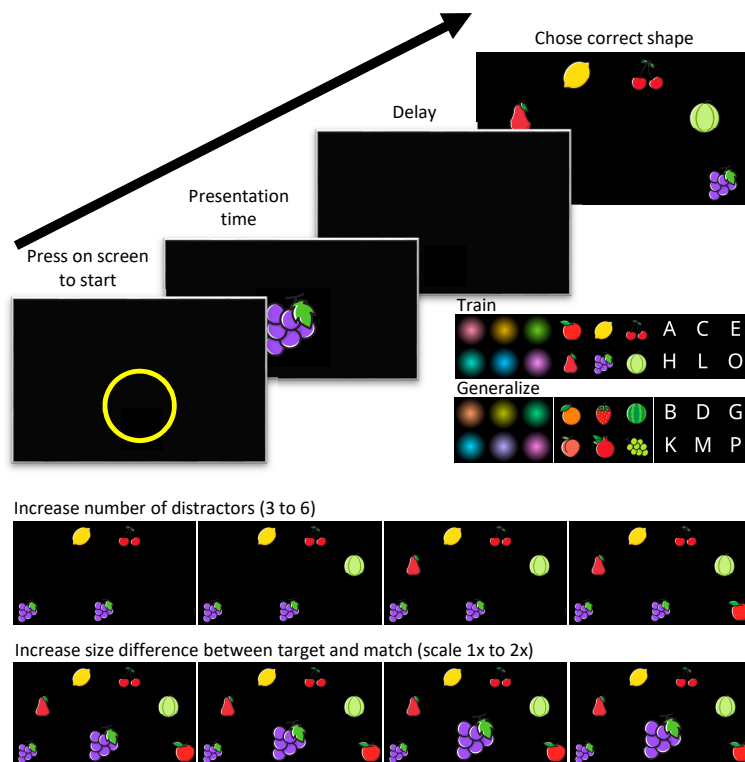
Figure 5.11: Training procedure used with the baboons. The task was a delayed match-to-sample, and several training steps were designed. Top-left: structure of a trial. Center-right: all training stimuli used. Bottom: example trials from 8 out of the 20 successive training steps.

Population

The baboon population was identical to the one tested in **chapter 1**. In total, 12 baboons reached the end of the training and performed enough test to be reported in the "fast" condition, and 9 in the "slow" condition. Additional methodological and ethical details are provided in the method section of **chapter 1**.

One additional group was recruited for the human adult experiments, in the fast condition; data for the "slow" condition is the one reported in the experiment 1 above, with self-paced encoding time. In the fast conditions, 560 participants were recruited. Demographics were self-reported and a few aberrant values were given but we did not remove participants based on that: the distribution averages at 37.7 years old, 1st quartile 28, 3rd quartile 45, SD=19.1; with 130 Females, 408 Males, and 22 who did not identify as either. The same criteria were applied as the first experiment; four participants were removed from the analysis as well as 1.85% of the data.

Method

The display of the stimuli was modified so that baboons could use the touchscreen to click on the initial target without hiding any part of the screen used for distractors afterward, whence the semicircle seen in **Figure 5.11**. In order for them to understand the task, a long training procedure was devised: first, they had to lean the match-to-sample task, using perceptually simple stimuli. First, they had to learn the task with choosing display feature the match and two distractors, then three, then four, and finally five for a total of 6 items on display. Then, we trained them to ignore scaling differences by making the target bigger and bigger, until it reached a scaling factor of 2 which matches the experimental condition in humans. Each baboon would move on to the next step whenever the average performance on the last block of 108 trials at a given step was better than 80% success. This was done in a first version of the task that tried to replicate the self-pace strategy: a trial started with an empty yellow circle, which was replaced with the target upon being touched. The target disappeared when touched again, and immediately after the distractor appeared.

Another minor difference is that the choice of possible distractors was slightly modified for the baboons' experiments, and correspondingly for the fast version of the human experiment: instead of two distrac-

205

tors close for the IT measure, two distractors close for the gray-level measure, and one chosen randomly uniformly in the remaining shapes, there were only one IT and one gray-level distractors, and the three others were chosen randomly. We decided to do this to make the task slightly simpler, as harder tasks may discourage baboons from participating at all.

This data was analyzed, and it became clear that the baboons did not use the ability offered by the self-pacing, making the experiment very different: there was no significant effect of the target shape on the encoding time (tested with an ANOVA, $F_{67,737}=1.07$, p=.33) and its median value was 108ms, much lower than even the minimum in adults. This confirmed the observed behavior that the baboons seemed to make very quick double-tap on the screen to start the screen and get to the distractors as fast as possible, a strategy which flashed the target long enough for them to perform correctly on the training stimuli and maximize their expected reward through the sheer number of trials per minute they could perform. But if indeed some shapes take longer to encode than others, such a behavior destroys the complexity of the MDL on the behavior.

This first data was collected from September 8th, 2021 to September 27th, 2021, and subsequently analyzed. When this discrepancy with the human adults was discovered, we decided to run an additional experiment where the looking time was constrained, to give enough time for the shapes to be encoded. To further replicate the human experiment, the delay between the target and the choice screen was also reintroduced. This led to an additional 12 training steps that take place before the first training step of the previous design: starting with 1-in-3 choice screen and 100ms presentation time / no delay, we slowly built up the presentation time duration (200ms, 300ms, 500ms, 700ms, 1000ms) and the delay (0ms, 100ms, 200ms, 300ms, 500ms, 700ms, 1000ms). Once both had reached 1000ms, we resumed the previous training procedure with these delays: adding back more distractors, then changing the scale, generalizing, and finally testing on the geometric shapes. This second experiment spanned from March 22nd, 2022 to April 09th, 2022.

Between these two experiments, we replicated the fast version of the experiment in adults, with a fixed presentation time of 100ms and no delay between presentation and choice. Additionally, we reused the semicircle display used in the baboons instead of the 3x2 grid, and we

had participants took the training steps in order: any success would move them to the next step, to avoid participants staying too long on the training as the experiment was conducted online.

Results
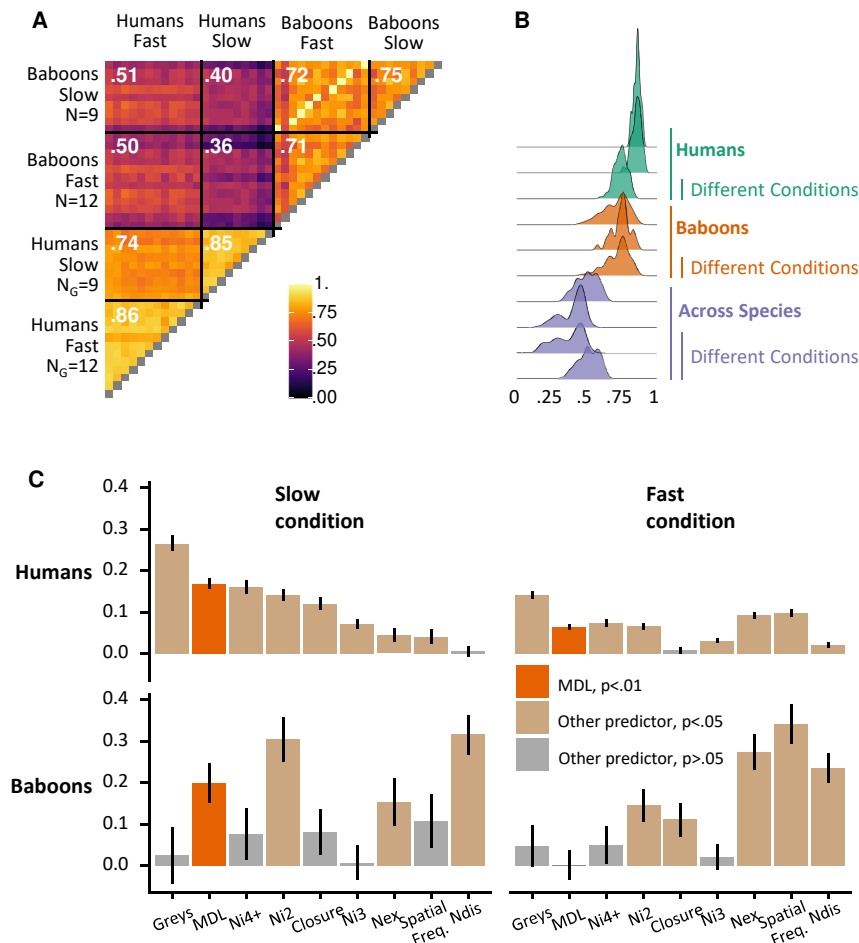
Figure 5.12: **A.** r coefficient associated with the correlation of the response time across shapes for (i) each baboon in the slow condition, (ii) each baboon in the fast condition, (iii) nine groups of humans (groups of equal size, randomly attributed) in the slow condition and (iv) twelve groups of humans (idem) in the fast condition. Values indicate average r within the "big cell", after removing correlation between the same individual baboon across conditions as they were very close to 1. **B.** distribution of the r coefficient in each "big cell" of the matrix in A; to visualize how distribution compare across species and conditions. **C.** Plot similar to **Figure 5.7.B** with (i) the estimates and standard error instead of the t-values, (ii) absolute value to visualize explained variance more straightforwardly, and (iii) a subplot for each population and condition.

After training, nine baboons in the slow condition and twelve baboons in the fast condition performed trials with the geometric shapes described in experiment 1.

First, we report the correlation of the response times across shapes for each baboon in each condition, and for random groups of humans. Individual participants having had a single trial per shape, and being more numerous than the baboons, we decided to randomly form as many groups of participants as there are individual baboons in a condition, to approximate stability in the population by averaging within group, and then measure the correlation across groups, treating human groups as we would individual baboons. The full correlation matrix is displayed in **Figure 5.12.A**, and distribution of r-coefficients within "big cells" are reported in **Figure 5.12.B**, where "big" cells refer to a population-condition pair.

First off, we notice that correlations are higher for within-species pairs than across species, irrelevant of the task, as made clear by the two triangles on the bottom left and top right in **Figure 5.12.A**, and the first two triplets of density estimates in **Figure 5.12.B.** Interestingly, within humans the correlation is higher for within condition (fast or slow) pairs than for the across condition pair, and that difference is significant (two t-test, both p<.001), however, in baboons neither is significant (p=.24 for one pair and p=.08 for the other, removing from the distributions the "same individual" data points).

While the correlations between humans and baboons are much lower than the within population correlations, another effect is visible: the fast, under-pressure humans are more correlated with the baboons (both conditions) than the self-paced humans. This is true whether we correlate fast humans with the slow baboons or the fast baboons.

Finally, we used the same model as for experiment 1 and ran a mixed-effect glm predicting the response time with many regressors to see whether the effect of the MDL stands out in both human conditions and in both baboons conditions: the coefficients after scaling both the dependent variable and all predictors are presented in **Figure 5.12.C**, sorted according to the absolute value of the predictor for the data already provided with experiment 1. We can see the predictors are all reduced and that the distribution changes a bit in the humans between the two conditions, but the gray level is still the dominant predictor and the MDL is still a significant predictor.

The coefficients of the baboons' data look very different from the humans, for both conditions. In the slow condition, the MDL is also significant in baboons, after the number of angles and the number of dis-

connected parts, and above the number of extremities. In the fast condition, the effect of MDL collapses, while the effect of closure becomes significant, as does the spatial frequency. Note that the standard error bars are much bigger for baboons than for humans because we use mixed models and have many more human participants than baboons.

5.6.2 Discussion and Future Work

These analyses are still preliminary, but they already point toward a number of interesting directions. First, the effect of population dominates the effect of conditions: humans behave like humans no matter the condition, and likewise for baboons. And yet, humans in the fast conditions are significantly closer to baboons than humans in the slow condition: this is predicted by the theory under the assumption that the flashed shape cannot be encoded – or at least, that some shapes cannot be encoded in such a short period, forcing participants to rely on more superficial visual properties.

Analysis of the coefficients in the mixed effect model indicate that the data the fast condition, when compared to the slow condition, is overall noisier in humans as almost all predictors take a toll, but overall very comparable and yielding comparable kind of behavior. Indeed, while the general ordering changes a bit to give more importance to the spatial frequency and the number of parts, the MDL still significantly impacts the behavior, and the overall distribution of predictors resembles that of the slow condition.

In baboons, the data in the fast condition has a very different look from the human pattern, with almost dissociated set of important predictors: the gray level, MDL and number of 4+ way intersections, which dominate in the human data, are not significant in baboons, and conversely the number of parts, spatial frequency, and number of extremities dominates in baboons and are mostly weakest in humans. However, the general pattern of predictors changes quite a lot in the slow condition, indicating that changing condition does indeed have an effect on the behavior beyond making participants more or less prone to mistakes.

Ultimately, these results show the limit of the current analyses. Unfortunately, it is meaningless to analyze error rates in humans as the performance are almost always perfect, and at the same time the response time in baboons is often disfavored and replaced with error rates be-

cause subjects tend to respond as fast as possible, quenching the variability in response time. This suggests moving to a speed-accuracy trade-off metric such as IES, RCS, LISAS or BIS (Liesefeld & Janczyk, 2019), however, many of these metrics require intra-individual estimation of the variance of the accuracy and response time, and we can't estimate this information in our human participants.

In addition to this, current models of the behavior only consider the target shape's properties: given the importance of the choice of distractors highlighted previously, it is crucial for a more accurate model of the behavior to take into account the distractors as well. One promising avenue for this is to use two models as done in **chapter 1** and **chapter 2**, a symbolic one and a perceptual one. The symbolic model could be a distance metric over the space of programs, perhaps the Hamming distance (which measures the number of additions, deletions and mutations required to go from one program to another). For the perceptual model, we can compute the confusion matrix across all shapes using a neural network of object recognition, to get a notion of perceptual similarity. These models, then, predict how difficult a given trial should be as a function of the list of distractors present, and we can separately account for response time variability in humans, and error rate variability in baboons.

Summary

Natural language is not the only hallmark of humans' singular cognitive abilities. In the work presented here, and in line with the language of thought literature, I argue that there could exist several internal "languages of thought", and that geometric shapes could be a pivotal way to study non-linguistic, high-level structured cognitive processes. To do so, I show that cognition involving geometric shapes requires a set of discrete, symbolic mental representations that act as an internal mental language with compositional properties, and provide evidence that non-human primates cannot access these representations. Under that view, perceiving a shape is comparable to the process of program induction: finding the shape's shortest representation in the internal mental language.

First, leveraging an intruder task with quadrilateral shapes, I show that humans of diverse education, age and culture share a sense of geometric complexity wherein some shapes are consistently simpler than others. I replicate the finding using several different paradigms and the same stimuli, including introspective reports of geometric complexity. However, baboons lack this sense even after adequate extensive training to the task.

Artificial neural networks of object recognition fit all baboons' data well, but explaining humans' behavior requires enriching the neural network with additional symbolic properties such as the presence of right angles – and in fact, in educated adults the symbolic model is enough. I show that the symbolic model also generalizes to other tasks,

and that data-driven methods for characterizing the behavior in humans yield dominant accounts that align well with the symbolic model.

This sharp dissociation suggests that two strategies are available to encode geometric shapes: both humans and non-human primates share a perceptual strategy, well captured by models of the ventral visual pathway, but only human have access to an exact, symbolic strategy. Using magnetoencephalography (MEG) during a passive shape perception task, I identify strong neural correlates of both strategies, well separated in location and temporal unfolding: I shed light on an early occipital response that resembles the neural network models, followed by a slower, more dorso-frontal response similar to the symbolic model. At the same time, using electroencephalography, I provide preliminary evidence for the existence of the symbolic strategy already in three-month-old infants. In fMRI, I find that passive geometric shape perception, when compared to the passive visual perception of other shapes such as faces, objects or letters, over-activates areas previously argued to belong to a vast, non-linguistic network for mathematical reasoning. I also localize more precisely areas specifically linked to the two models, and find interesting overlap between the two which may indicate where the visual information is integrated into the higher-level, symbolic mental representation.

Finally, I show that if a symbolic strategy exists redundantly with a lower-level strategy which is rich enough to perform a given task, humans may not always use the symbolic strategy at all. This is compatible with the argument made from previous experiments that the symbolic strategy is not pre-attentive, and importantly this suggests ways of preventing participants from deploying symbolic strategies, thereby pushing them toward a non-human private behavior. Conversely, this could provide new ways to test non-human primates, allowing for a finer-grained characterization of the differences between human and non-human primates.

Going beyond a small set of highly controlled quadrilaterals, I set to try to account for *all geometric* shapes produced by humans: I make a concrete proposition for a generative mental language of geometric shapes inspired by attested human geometric productions. I argue that in humans, perceiving a shape means finding the shortest program in this language that generates the shape, connecting shape perception in humans to the literature on program induction.

Equipped with this language, I show that program induction is in principle a tractable problem for this domain using the DreamCoder algorithm I helped implement. Then, I use this language to enumerate shapes of increasingly high complexity, and show that humans' performance in a match-to-sample task for a shape correlate with the length of its shortest program. To decouple this result from the exact language proposition, I also derive more general additive rules that any alternative languages must obey; then with new experiments, I provide strong empirical evidence for the existence of these laws as models of human behavior.

Taken all together, these results support the existence of a discrete, symbolic set of mental representations for geometric shapes, which coexists with bottom-up visual representations. This work paves the way for integrated neuro-symbolic models of shape perception while challenging currently dominant object-recognition based models of visual perception.

Future Work

The work presented here is, unsurprisingly, far from settling all the questions it sets out to consider. In particular, there are several research directions that I want to emphasize as being good candidates for future work, and list them below. They correspond to the main properties of the phenomenon I describe: its universality (in humans, and in particular in babies), its uniqueness (as compared to other species), its cognitive plausibility (symbolic approaches must somehow be implemented in real networks of real neurons) and its emergence (both at the evolutionary timescale, and at the developmental timescale).

Extensive Comparison with Non-Human Primates

In order to make a strong argument of the human singularity of a language of thought for geometric shapes, the comparison with a single species is not enough: after all, the baboons could be singular in not sharing this ability with other primates.

While I don't think this alternative theory will eventually prove correct, gathering data from more non-human primates, and in particular great apes, would be very informative. During my PhD we started

a collaboration with the "Ape Cognition and Conservation Initiative" (ACCI, "dedicated to the science, conservation and protection of Great Apes worldwide"). There, in collaboration with Jared Taglialatela and Amanda Epping, we started a training procedure similar to the one performed with baboons on the quadrilateral intruder test with seven bonobos. Unfortunately, for a variety of methodological and administrative reasons, the training procedure was slowed down immensely, forcing us to adapt the training steps on the fly and focus on a subset of the available subjects, namely Nyota, Teco, and the infamous Kanzi.

While these three subjects eventually reached acceptable performance on the training trials, it took a very long time, and went over the time and financial budget we could allocate for this work. The trials we collected for the main geometric task are inconclusive about the research questions, as subjects were still at chance across all possible geometric shapes, but it allows us to at least conclude about a failure to generalize straightforwardly from the training procedure in three individuals, in stark contrast with what even five-years-old children appear to do effortlessly.

Without additional data, further scientific claims about bonobos are unwarranted. But the duration of the training and the lack of generalization are interesting observations alone – in particular in the case of Kanzi who is said to have learned to use over 200 lexigrams (not mentioning the combinatorial explosion), most of which are colorful geometric shapes, to communicate (Greenfield & Savage-Rumbaugh, 1990).

Going beyond behavioral data, intracranial data from primates could make comparing the neural dynamic of geometric shape perception in humans even more straightforward, thanks to the MEG and fMRI data acquired already. Using passive presentation paradigms and high-resolution recording arrays over several brain areas, as is already done e.g. in occipital areas (Ponce et al., 2019) or in the prefrontal cortex (Bellet et al., 2021; Xie et al., 2022), would be particularly impactful.

Indeed, such neuronal data would unlock several possibilities to test hypotheses. First, it could be used to test whether the early, neural network-like response observed in humans in **chapter 2** already exists in non-human primates – this is likely, given that the exceptional predictiveness of the neural network model on the behavioral data. But it could also help adjudicate between different hypotheses of why ba-

boons' behavior doesn't correlate at all with the symbolic model: if we find that there is a late, frontal response that resembles the symbolic model in non-human primates, then this will strongly suggest that the absence of "geometric complexity effect" in **chapter 1** is a byproduct of the design of the task itself rather than a limitation in the cognitive abilities of baboons. If, however, no brain area at all seems to correspond to the symbolic model a hypothesis supported by our data but also by recent experiments using geometrical sequences (H. Zhang et al., 2022), this will strongly strengthen the argument I present in the present work.

Neural Networks, Neurosymbolic Models, and Program Induction

In computer science and AI, attempts at solving the problem of Program Induction has historically been dominated by symbolic strategies, recently improved with neurosymbolic models. One reason for this is the inherent roughness of the space of programs (for any expressive enough programming language, one can find two programs that are arbitrarily close in "source space", or in the space of the language, with arbitrarily different outputs[1]).

In an interesting turn of events, it appears that neural networks trained for natural language processing on gigantic web-scrapped corpora end up learning a fair amount of programming in different languages. For example, GPT-3 (Brown et al., 2020), when prompted with text formatted using LaTeX's syntax, well produce continuations that are often syntactically valid LaTeX markup. For common enough problems, prompting the model with the beginning of a python solution can make it produce a valid continuation.

Neither of these uses were explicit goal of the training procedure, but this realization prompted the creation of Codex (Chen et al., 2021), which powers the current version of GitHub's "copilot" AI for code suggestion. Its use range over all sorts of tasks, such as (i) writing a function given a short comment describing it, (ii) explaining a regular expression in a comment just above, or (iii) writing a syntactically valid LaTeX equation for a mathematical computation in code; and it does so it does so across seven programming so far.

---

[1]I want to thanks Steven Piantadosi for pointing this apparently unpublished result to me

And while natural language is bound by existing productions (Hoffmann et al., 2022) as there are no oracles for what constitutes valid utterances in a given language, this does not hold true for computer languages. In DreamCoder (Ellis et al., 2021), a neurosymbolic model, we leverage this property to generate training data from scratch. Recent work has taken a purely-non-symbolic approach to this problem (Haluptzok et al., 2022) in a more ambitious way: they generate "programming puzzles" randomly, and train a large language model to solve them, thereby breaking free of the existing training data in tackling the program induction in a purely "large language model" way.

This strategy might soon provide us with remarkably good tools for solving program induction problems. But unlike natural languages, there is no debate that program induction is a symbolic problem at its core. Therefore, we might have very interesting models to study how a purely symbolic problem can be solved in a purely neural-network model – whether this will be informative about the cognition remains to be seen.

Cognitive Plausibility and Implementation of Program Induction

The other side of the coin is that "language(s) of thought" must be implemented in the brain: a real network of real neurons. What cortical and subcortical mechanisms could implement symbolic manipulation and program induction?

One interesting avenue, which I intended to explore during my postdoc, is that hippocampal replay, a mechanism by which the brain rapidly reactivates in ways that are reminiscent of a stimulus (Diba & Buzsáki, 2007; Foster & Wilson, 2006), might be repurposed to implement program induction. It is now possible to detect replay of a given stimulus in humans with MEG data by using decoders trained on the reference activations for the stimulus, and tested during resting or deliberating states. Recently, it has been hypothesized that generative replay plays a role in hypothesis testing or model building (Schwartenbeck, 2021). As such, generative replay is an explicit neurobiological proposition of how a mechanism such as program induction might be implemented in the brain, and it could prove a pivotal bridge between the theoretical cognitive proposals and the

neurobiological mechanisms underpinning mental representations.

At the same time, the mechanism of replay is far from being unique to humans. Studying the link between replay and the act of building abstract representations is therefore useful in two ways: it offers a concrete challenge to the program induction theory, and suggests line of research to explore the limits of what might make humans appear singular in the animal kingdom from a cognitive standpoint.

Evolution, at Different Timescales, of the Language of Thought Hypothesis

There are two timescales at which the emergence and progression of the Language(s) of Thought are particularly critical for a better understanding of its roots: the individual developmental, and the evolutionary timescale.

Throughout my work I insist on the universality of the existence of symbols in human cognition, at least in the case of geometric shapes. But evidence provided as early as in **chapter 1** indicates that, unsurprisingly, there **are** differences between adults and children. I have not tried to model those differences explicitly yet, but already this observation prompts several questions.

In particular, it would be interesting to try to understand whether the effect of education is to develop the symbolic strategy itself, or to train explicitly to overemphasize one strategy against the other. Previous work in approximate number system and math abilities (Elliott et al., 2018; Nys et al., 2013; Piazza et al., 2013), showing that explicit training in mathematics appears to positively impact the approximate number system, suggests that explicit teaching in geometry could similarly impact the non-symbolic strategy. Additionally, given the fMRI evidence proposed in **chapter 3** and the apparent proximity between geometric shape perception and mathematics, but not language, another immediate question is whether reading scores, language score and math scores would correlate with performance in our intruder task.

And what about the evolutionary timescale? Under the current framework it seems that when compared to other primates, evolution would affect multiple, broad and bilateral areas, encompassing at least the inferior and middle prefrontal cortex, as well as the associative pari-

etal and temporal areas (Eichert et al., 2020; Hill et al., 2010; Mars et al., 2018). What could underscore specifically the language(s) of thought? One exciting possibility is that a mutation made composition-ality available independently to many brain areas, perhaps by chang-ing low-level properties of the neural substrate, either the cells, or the laminal cortical organization, or other. Under that hypothesis, a single set of mutation could have led to the emergence of new abilities in var-ious areas, which would then end up specializing to different domains according to their preexisting functional connectivity to the cortex.

Closing Words

> If a mathematician of the past, an Archimedes or even a Descartes, could view the field of geometry in its present condition, the first feature to impress him would be its lack of concreteness. There are whole classes of geometric theories which proceed [...] without the slightest (apparent) use of the spatial intuition.

Edward Kasner in "The present problems of geometry" (Kasner, 1905)

The quest for understanding the natural of abstract representations through an examination of the geometric abilities of humans, which started more than two thousand years ago with Plato, is unsurpris-ingly far from over. That quest has featured many great names from both cognitive psychology, cognitive neuroscience, computer science and mathematics.

Much like mathematicians of the past would be surprised by what ge-ometry looked like in 1905 (years before Grothendieck was even born – what would Kasner think about *that* revolution), I hope cognitive sci-ence about geometry and abstraction will keep on surprising and en-lightening us. It seems that now things are moving faster and faster: artificial neural networks achieve superhuman performance in some vision-related tasks and in some natural language processing tasks, neuroimaging techniques make it possible to find an algebraic encod-ing of the unfolding of a geometrical sequence over time in primates' brains, and testing thousands of humans on a single task has never been easier – what an exciting time it is to work in this environment!

Acknowledgements

Alexandre, Léo, Kevin, Kévin, Ava, Micha, Élin, Victor, Hsiao, Nicolas, Martin, Antoine, et André.

Open Data Statement

It is hard to overstate how many practical steps in the process of pursuing this work were made possible by the open-source community, its members, and its mindset of easily hackable, openly available tools. This includes programming languages and their packages (R, python, ggplot, tidyverse, afex, nilearn, mne, numpy, pandas, javascript, jsPsych, and many, many others), browsers (firefox), OS kernels, distributions, and desktop environments (GNU/Linux, archlinux, wayland, sway), text manipulation and formatting tools (vim, LaTeX, pandoc, openoffice), image manipulation tools (inkscape, gimp), and a plethora of other quality-of-life day to day tools. To the people that develop these tools, that respond to questions, that add features or fix bugs, and lay the very foundation of so many modern scientific computing setups: thank you.

I have submitted minor contributions to some of these tools across many years, but there's only so much I can provide with my expertise. However, I can follow the guiding principles and set free as much of my own code, experiments, analyses and data as I can. Each chapter roughly maps onto an independently published (or to be published, hopefully) article, which will have its accompanying dataset and scripts. In the event that, upon reading this, you find yourself unable to find an associated chunk of data or script, do not hesitate to reach out, I will do my best to both make it available to you and make sure that others find it if they need in the future.

A Word on Public Services and Education

I want to extensively acknowledge the French public institutions that brought me to where I am. My entire education has taken place inside the public sector. It has cost my family or me virtually nothing, while providing for all of my intellectual needs, and then some. Indeed, starting with what would be the equivalent of the third year of university, and every year since, my education has been publicly funded, providing me with a modest but most welcome salary – all this without being possessive: I have spent half a year in Oxford, UK, half a year at MIT, USA, and a full (unpaid) year sailing. During all these years I have been

granted a lot of freedom, together with adequate support, to organize my curriculum as I deemed most fitting.

The system is far from perfect in many ways, notably in terms of inclusiveness, and I have been known to be grouchy when it comes to bureaucracy and administrative burdens – but this does not prevent me from stating that in my view, efforts toward improving and expanding the existing system are far better spent than efforts trying to replace it or abandon it altogether. There is no individual to thank for this, but I want to give appropriate credit to the ideals that gave us this, and the institution that support it on a daily basis. Or, in the words of a ride during a hitchhiking trip abroad (*marked surprise, heaving Eastern-Europe accent*): "You travel with money you get from your government as a student? It's a good government!"

References

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

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, *113*(18), 4909–4917. https://doi.org/10.1073/pnas.1603205113

Amalric, M., & Dehaene, S. (2017). Cortical circuits for mathematical knowledge: Evidence for a major subdivision within the brain's semantic networks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1740). https://doi.org/10.1098/rstb.2016.0515

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.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193. https://doi.org/10.1037/h0054663

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Ayzenberg, V., Chen, Y., Yousif, S. R., & Lourenco, S. F. (2019). Skeletal representations of shape in human vision: Evidence for a pruned medial axis model. *Journal of Vision*, *19*(6), 6–6.

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-45268-y

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, *14*(12), e1006613. https://doi.org/10.1371/journal.pcbi.1006613

Balog, M., Gaunt, A. L., Brockschmidt, M., Nowozin, S., & Tarlow, D. (2017). Deep-Coder: Learning to Write Programs. *ArXiv:1611.01989 [Cs]*. http://arxiv.org/abs/1611.01989

Beckers, G. J. L., Berwick, R. C., Okanoya, K., & Bolhuis, J. J. (2016). What do animals learn in artificial grammar studies? *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2016.12.021

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bellet, M. E., Gay, M., Bellet, J., Jarraya, B., Dehaene, S., van Kerkoerle, T., & Panagiotaropoulos, T. I. (2021). Prefrontal neural ensembles encode an internal model of visual sequences and their violations. *BioRxiv*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Berwick, R. C., & Chomsky, N. (2016). *Why Only Us: Language and Evolution*. The MIT Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115.

Blum, H. (1967). Models for the perception of speech and visual form. *Ch. A Transformation for Extracting New Descriptors of Shape*, 362–380.

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

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

Borg, I., Groenen, P. J., & Mair, P. (2018). *Applied multidimensional scaling and unfolding*. Springer.

Bornstein, M. H., Gross, C. G., & Wolf, J. Z. (1978). Perceptual similarity of mirror images in infancy. *Cognition*, *6*(2), 89–116.

Boselie, F., & Leeuwenberg, E. (1986). A Test of the Minimum Principle Requires a Perceptual Coding System. *Perception*, *15*(3), 331–354. https://doi.org/10.1068/p150331

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci*, *7*(8), 880–886.

Brincat, S. L., & Connor, C. E. (2006). Dynamic Shape Synthesis in Posterior Inferotemporal Cortex. *Neuron*, *49*(1), 17–24. https://doi.org/10.1016/j.neuron.2005.11.026

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & others. (2020). Language models are few-shot

learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Brüderlin, B. (1985). Using prolog for constructing geometric objects defined by constraints. In B. F. Caviness (Ed.), *EUROCAL '85* (Vol. 204, pp. 448–459). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-15984-3_309

Butterworth, B., Reeve, R., Reynolds, F., & Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proc Natl Acad Sci U S A*, *105*(35), 13179–13184.

Byrne, O., & Euclid. (1847). *The first six books of the Elements of Euclid: In which coloured diagrams and symbols are used instead of letters for the greater ease of learners*. William Pickering.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS One*, *5*(6), e10729.

Calero, C. I., Shalom, D. E., Spelke, E. S., & Sigman, M. (2019). Language, gesture, and judgment: Children's paths to abstract geometry. *Journal of Experimental Child Psychology*, *177*, 70–85. https://doi.org/10.1016/j.jecp.2018.07.015

Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biol*, *5*(12), e328.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/978019

Chaplin, T. A., Yu, H.-H., Soares, J. G. M., Gattass, R., & Rosa, M. G. P. (2013). A Conserved Pattern of Differential Expansion of Cortical Areas in Simian Primates. *Journal of Neuroscience*, *33*(38), 15120–15125. https://doi.org/10.1523/JNEUROSCI.2909-13.2013

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291. https://doi.org/10.1016/j.tics.2006.05.007

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

Chaudhuri, S., Ellis, K., Polozov, O., Singh, R., Solar-Lezama, A., & Yue, Y. (2021). Neurosymbolic Programming. *Foundations and Trends® in Programming Languages*, *7*(3), 158–243. https://doi.org/10.1561/2500000049

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., & others. (2021). Evaluating large language models trained on code. *ArXiv Preprint ArXiv:2107.03374*.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*(2), 149–178.

Cheng, K., Huttenlocher, J., & Newcombe, N. S. (2013). 25 years of research on the use of geometry in spatial reorientation: A current theoretical perspective. *Psychonomic Bulletin & Review*, *20*(6), 1033–1054.

Cheng, K., & Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin &amp$\mathsemicolon$*

Review, *12*(1), 1–23. https://doi.org/10.3758/bf03196346

Church, A. (1932). A set of postulates for the foundation of logic. *Annals of Mathematics*, 346–366.

Ciccione, L., & Dehaene, S. (2021). Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cognitive Psychology*, *128*, 101406. https://doi.org/10.1016/j.cogpsych.2021.101406

Ciccione, L., Sablé-Meyer, M., & Dehaene, S. (2022). Analyzing the misperception of exponential growth in graphs. *Cognition*, *225*, 105112. https://doi.org/10.1016/j.cognition.2022.105112

Comrie, B. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends Cogn Sci*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Cusumano-Towner, M. F., Saad, F. A., Lew, A. K., & Mansinghka, V. K. (2019). Gen: A general-purpose probabilistic programming system with programmable inference. *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*, 221–236. https://doi.org/10.1145/3314221.3314642

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Davidoff, J., Roberson, D., & Shapiro, L. (2002). Squaring the Circle: The Cultural Relativity of "Good" Shape. *Journal of Cognition and Culture*, *2*(1), 29–51. https://doi.org/10.1163/156853702753693299

De Cruz, H. (2009). An enhanced argument for innate elementary geometric knowledge and its philosophical implications. *New Perspectives on Mathematical Practices. Essays in Philosophy and History of Mathematics*, 185–206.

De Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*, 1–30.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Company.

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*, S1364661322001413. https://doi.org/10.1016/j.tics.2022.06.010

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*, *311*, 381–384.

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

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. *Science*, *284*(5416), 970–974. https://doi.org/10.1126/science.284.5416.970

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, *16*(3), e2004103. https://doi.org/10.1371/journal.pbio.2004103

Dehaene-Lambertz, G., & Spelke, E. S. (2015). The Infancy of the Human Brain. *Neuron*, *88*(1), 93–109. https://doi.org/10.1016/j.neuron.2015.09.026

Devlin, J., Uesato, J., Bhupatiraju, S., Singh, R., Mohamed, A., & Kohli, P. (2017). RobustFill: Neural Program Learning under Noisy I/O. *ArXiv:1703.07469 [Cs]*. http://arxiv.org/abs/1703.07469

Dillon, M. R., Duyck, M., Dehaene, S., Amalric, M., & Izard, V. (2019). Geometric categories in cognition. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000663

Dillon, M. R., Izard, V., & Spelke, E. S. (2020). Infants' sensitivity to shape changes in 2D visual forms. *Infancy*, *25*(5), 618–639. https://doi.org/10.1111/infa.12343

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Eichert, N., Robinson, E. C., Bryant, K. L., Jbabdi, S., Jenkinson, M., Li, L., Krug, K., Watkins, K. E., & Mars, R. B. (2020). Cross-species cortical alignment identifies different types of anatomical reorganization in the primate temporal lobe. *ELife*, *9*, e53232. https://doi.org/10.7554/eLife.53232

Elliott, L., Feigenson, L., Halberda, J., & Libertus, M. E. (2018). Bidirectional, Longitudinal Associations Between Math Ability and Approximate Number System Precision in Childhood. *Journal of Cognition and Development*, *20*(1), 56–74. https://doi.org/10.1080/15248372.2018.1551218

Ellis, K., Ritchie, D., Solar-Lezama, A., & Tenenbaum, J. (2018). Learning to infer graphics programs from hand-drawn images. *Advances in Neural Information Processing Systems*, *31*.

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., Cary, L., Solar-Lezama, A., & Tenenbaum, J. B. (2021). DreamCoder: Bootstrapping inductive program synthesis with wake-sleep library learning. *Proceedings of the 42nd ACM SIG-*

PLAN International Conference on Programming Language Design and Implementation, 835–850. https://doi.org/10.1145/3453483.3454080

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). FMRIPrep. *Software*. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*. https://doi.org/10.1038/s41592-018-0235-4

Evans, A., Janke, A., Collins, D., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, *62*(2), 911–922. https://doi.org/10.1016/j.neuroimage.2012.01.024

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

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616–16621. https://doi.org/10.1073/pnas.1315235110

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.*, *8*(7), 307–314.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630.

Feldman, J. (2001). Bayesian contour integration. *Perception & Psychophysics*, *63*(7), 1171–1182. https://doi.org/10.3758/BF03194532

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science (Wiley-Blackwell)*, *12*(6), 227–232. https://doi.org/10.1046/j.0963-7214.2003.01267.x

Feldman, J., & Singh, M. (2005). Information Along Contours and Object Boundaries. *Psychological Review*, *112*(1), 243–252. https://doi.org/10.1037/0033-295X.112.1.243

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

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Science Advances*, *6*(26), eaaz1002. https://doi.org/10.1126/sciadv.aaz1002

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*(2), 173–193. https://doi.org/10.1016/0042-6989(93)90156-Q

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. https://doi.org/10.1177/0956797613507584

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364. https://doi.org/10.1016/j.plrev.2014.04.005

Fló, A., Gennari, G., Benjamin, L., & Dehaene-Lambertz, G. (2022). Automated Pipeline for Infants Continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. *Developmental Cognitive Neuroscience*, *54*, 101077. https://doi.org/10.1016/j.dcn.2022.101077

Fobes, J. L., & King, J. E. (1982). *Primate behavior*. Academic Pr.

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

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47, Supplement 1*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Forget, J., Buiatti, M., & Dehaene, S. (2009). Temporal integration in visual word recognition. *J Cogn Neurosci*, *22*(5), 1054–1068. https://doi.org/10.1162/jocn.2009.21300

Frankland, S. M., & Greene, J. D. (2020). Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annual Review of Psychology*, *71*(1), 273–303. https://doi.org/10.1146/annurev-psych-122216-011829

Franklin, W. R., Wu, P. Y., Samaddar, S., & Nichols, M. (1986). Prolog and geometry projects. *IEEE Computer Graphics and Applications*, *6*(11), 46–55.

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

Gallistel, C. R. (1990). *The organization of learning*. MIT Press.

Garcez, A. d'Avila, & Lamb, L. C. (2020). Neurosymbolic AI: the 3rd wave. *ArXiv Preprint ArXiv:2012.05876*.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *ArXiv:1811.12231 [Cs, q-Bio, Stat]*. http://arxiv.org/abs/1811.12231

George, D., Lehrach, W., Kansky, K., Lázaro-Gredilla, M., Laan, C., Marthi, B., Lou, X., Meng, Z., Liu, Y., Wang, H., Lavin, A., & Phoenix, D. S. (2017). A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs. *Science (New York, N.Y.)*, *358*(6368). https://doi.org/10.1126/science.aag2612

Giaquinto, M. (2005). From Symmetry Perception to Basic Geometry. In *Visualization, Explanation and Reasoning Styles in Mathematics* (pp. 31–55). Springer Netherlands. https://doi.org/10.1007/1-4020-3335-4_3

Goldreich, D., & Peterson, M. A. (2012). A Bayesian Observer Replicates Convexity Context Effects in Figure–Ground Perception. *Seeing and Perceiving*, *25*(3–4), 365–395. https://doi.org/10.1163/187847612X634445

Gombrich, E. H. (1994). *The Sense of Order: A Study in the Psychology of Decorative Art*. Phaidon Press.

Goodenough, F. L. (1926). *Measurement of intelligence by drawings.*

Goodman, N., Mansinghka, V., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2012). Church: A language for generative models. *ArXiv Preprint ArXiv:1206.3255*.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. *Software*. https://doi.org/10.5281/zenodo.596855

Gould, S. J., & Eldredge, N. (1977). Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered. *Paleobiology*, *3*(2), 115–151. JSTOR.

Grace, R. C., Carvell, G. E., Morton, N. J., Grice, M., Wilson, A. J., & Kemp, S. (2020). On the origins of computationally complex behavior. *Journal of Experimental Psychology: Animal Learning and Cognition*, *46*(1), 1–15. https://doi.org/10.1037/xan0000227

Greene, E. (2016). Information persistence evaluated with low-density dot patterns. *Acta Psychologica*, *170*, 215–225. https://doi.org/10.1016/j.actpsy.2016.08.005

Greenfield, P. M., & Savage-Rumbaugh, E. S. (1990). Grammatical combination in $\less$i$\greater$Pan paniscus$\less$/i$\greater$: Processes of learning and invention in the evolution and development of language. In *"Language" and intelligence in monkeys and apes* (pp. 540–578). Cambridge University Press. https://doi.org/10.1017/cbo9780511665486.022

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Halat, E., & Dağli, Ü. Y. (2016). Preschool students' understanding of a geometric shape, the square. *Bolema: Boletim de Educação Matemática*, *30*, 830–848.

Haluptzok, P., Bowers, M., & Kalai, A. T. (2022). Language Models Can Teach Themselves to Program Better. *ArXiv Preprint ArXiv:2207.14502*.

Harris, D. B. (1963). *Children's Drawings as Measures of Intellectual Maturity: Revision of Goodenough Draw-a-man Test*.

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), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hauser, M. D., & Watumull, J. (2017). The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2016.10.005

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

Henshilwood, C. S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G. A. T., Mercier, N., Sealy, J. C., Valladas, H., Watts, I., & Wintle, A. G. (2002). Emergence of Modern Human Behavior: Middle Stone Age Engravings from South Africa. *Science*, *295*(5558), 1278–1280. https://doi.org/10.1126/science.1067575

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370*(6484), 57–59.

Hermes, D., Rangarajan, V., Foster, B. L., King, J.-R., Kasikci, I., Miller, K. J., & Parvizi, J. (2017). Electrophysiological responses in the ventral temporal cortex during reading of numerals and calculation. *Cerebral Cortex*, *27*(1), 567–575.

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science*, *317*(5843), 1360–1366. https://doi.org/10.1126/science.1146282

Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., & Van Essen, D. (2010). Similar patterns of cortical expansion during human development and evolution. *Proceedings of the National Academy of Sciences*, *107*(29), 13135–13140. https://doi.org/10.1073/pnas.1001229107

Hochberg, J., & McAlister, E. (1953). A quantitative approach, to figural" goodness". *Journal of Experimental Psychology*, *46*(5), 361.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. de L., Hendricks, L. A., Welbl, J., Clark, A., & others. (2022). Training Compute-Optimal Large Language Models. *ArXiv Preprint ArXiv:2203.15556*.

Hofstadter, D. R. (2001). Analogy as the core of cognition. *The Analogical Mind: Perspectives from Cognitive Science*, 499–538.

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic books.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*(2), 229–289.

Hung, C.-C., Carlson, E. T., & Connor, C. E. (2012). Medial Axis Shape Coding in Macaque Inferotemporal Cortex. *Neuron*, *74*(6), 1099–1113. https://doi.org/10.1016/j.neuron.2012.04.029

Hupbach, A., & Nadel, L. (2005). Reorientation in a rhombic environment: No evidence for an encapsulated geometric module. *Cognitive Development*, *20*(2), 279–302. https://doi.org/10.1016/j.cogdev.2005.04.003

Hyde, D. C., Winkler-Rhoades, N., Lee, S.-A., Izard, V., Shapiro, K. A., & Spelke, E. S. (2011). Spatial and numerical abilities without a complete natural language. *Neuropsychologia*, *49*(5), 924–936. https://doi.org/10.1016/j.neuropsychologia.2010.12.017

Iacaruso, M. F., Gasler, I. T., & Hofer, S. B. (2017). Synaptic organization of visual space in primary visual cortex. *Nature*, *547*(7664), 449–452. https://doi.org/10.1038/nature23019

Izard, V. (2022). Visual foundations of Euclidean geometry. *Cognitive Psychology*, 24.

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2008). Exact Equality and Successor Function: Two Key Concepts on the Path towards Understanding Exact Numbers. *Philosophical Psychology*, *21*(4), 491–505. https://doi.org/10.1080/09515080802285354

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

Jacob, S. N., & Nieder, A. (2009). Notation-independent representation of fractions in the human parietal cortex. *J Neurosci*, *29*(14), 4652–4657.

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations, Quality Assessments, and Good Practices. *Frontiers in Neuroscience*, *12*, 530. https://doi.org/10.3389/fnins.2018.00530

Jatoi, M. A., Kamel, N., Malik, A. S., & Faye, I. (2014). EEG based brain source localization comparison of sLORETA and eLORETA. *Australasian Physical & Engineering Sciences in Medicine*, *37*(4), 713–721.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Joordens, J. C. A., d'Errico, F., Wesselingh, F. P., Munro, S., de Vos, J., Wallinga, J., Ankjærgaard, C., Reimann, T., Wijbrans, J. R., Kuiper, K. F., Mücher, H. J., Coqueugniot, H., Prié, V., Joosten, I., van Os, B., Schulp, A. S., Panuel, M., van der Haas, V., Lustenhouwer, W., … Roebroeks, W. (2015). Homo erectus at Trinil on Java used shells for tool production and engraving. *Nature*, *518*(7538), 228–231. https://doi.org/10.1038/nature13962

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., & others. (2015). Homo erectus at Trinil on Java used shells for tool production and engraving. *Nature*, *518*(7538), 228.

Kanizsa, G. (1976). Subjective Contours. *SCIENTIFIC AMERICAN*, 7.

232

Kasner, E. (1905). The present problems of geometry. *Bulletin of the American Mathematical Society*, *11*(6), 283–314.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271–304.

Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. *ArXiv:1312.6114 [Cs, Stat]*. http://arxiv.org/abs/1312.6114

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

Koller, D., McAllester, D., & Pfeffer, A. (1997). Effective Bayesian inference for stochastic programs. *AAAI/IAAI*, 740–747.

Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A*, 369–376.

Kosslyn, S. M. (1980). *Image and Mind*. Harvard University_Press.

Kovács, I., Fehér, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, *38*(15–16), 2323–2333.

Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N., Issa, E., Bashivan, P., Prescott-Roy, J., & Schmidt, K. (2019). Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. *Advances in Neural Information Processing Systems*, *32*, 12805–12816.

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

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. https://doi.org/10.1017/S0140525X16001837

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from* (Vol. 6). New York: Basic Books.

Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007

Landau, B., Gleitman, H., & Spelke, E. (1981). Spatial knowledge and geometric representation in a child blind from birth. *Science*, *213*(4513), 1275–1278.

Le Tensorer, J.-M. (2006). Les cultures acheuléennes et la question de l'émergence de la pensée symbolique chez Homo erectus à partir des données relatives à la forme symétrique et harmonique des bifaces. *Comptes Rendus Palevol*, *5*(1–2), 127–135.

Ledgeway, T., Hess, R. F., & Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: Empirical tests of "association field" models of contour integration. *Vision Research*, *45*(19), 2511–2522. https://doi.org/10.1016/j.visres.2005.04.002

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology-Paris*, *97*(2–3), 121–139.

Leeuwen, C. van. (1990). Perceptual-learning systems as conservative structures: Is economy an attractor? *Psychological Research*, *52*(2–3), 145–152. https://doi.org/10.1007/bf00877522

Leeuwenberg, E. L. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, *76*(2), 216.

Leeuwenberg, E. L. (1971). A perceptual coding language for visual and auditory patterns. *The American Journal of Psychology*, 307–349.

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

Leyton, M. (1984). Perceptual organization as nested control. *Biological Cybernetics*, *51*(3), 141–153.

Leyton, M. (2003). *A generative theory of shape* (Vol. 2145). Springer.

Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. Springer Heidelberg. http://cs.ioc.ee/yik/lib/7/Li1.html

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

Long, B., Fan, J., Chai, Z., & Frank, M. C. (2019). *Developmental changes in the ability to draw distinctive features of object categories* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/8rzku

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, *80*(5), 1278–1289. https://doi.org/10.3758/s13414-017-1457-8

Mach, E. (1914). *The analysis of sensations, and the relation of the physical to the psychical.* Open Court Publishing Company.

Mair, P., Groenen, P. J. F., & de Leeuw, J. (2022). More on Multidimensional Scaling and Unfolding in *R*: **Smacof** Version 2. *Journal of Statistical Software*, *102*(10). https://doi.org/10.18637/jss.v102.i10

Malassis, R., Dehaene, S., & Fagot, J. (2020). Baboons (Papio papio) Process a Context-Free but Not a Context-Sensitive Grammar. *Scientific Reports*, *10*(1), 7381.

https://doi.org/10.1038/s41598-020-64244-5

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294.

Mars, R. B., Sotiropoulos, S. N., Passingham, R. E., Sallet, J., Verhagen, L., Khrapitchev, A. A., Sibson, N., & Jbabdi, S. (2018). Whole brain comparative anatomy using connectivity blueprints. *ELife*, *7*, e35237. https://doi.org/10.7554/eLife.35237

Maruyama, M., Pallier, C., Jobert, A., Sigman, M., & Dehaene, S. (2012). The cortical representation of simple mathematical expressions. *NeuroImage*, *61*(4), 1444–1460. https://doi.org/10.1016/j.neuroimage.2012.04.020

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346–362. https://doi.org/10.1016/j.cognition.2011.11.003

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*(6014), 57–59.

McBride, T., Arnold, S. E., & Gur, R. C. (1999). A Comparative Volumetric Analysis of the Prefrontal Cortex in Human and Baboon MRI. *Brain, Behavior and Evolution*, *54*(3), 159–166. https://doi.org/10.1159/000006620

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the "cognitive map." *Nat Rev Neurosci*, *7*(8), 663–678. https://doi.org/10.1038/nrn1932

Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., Brázdil, M., & Krupa, P. (2008). Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*, *26*(4), 490–503. https://doi.org/10.1016/j.mri.2007.08.006

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought beyond language: Neural dissociation of algebra and natural language. *Psychological Science*, *23*(8), 914–922. https://doi.org/10.1177/0956797612437427

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., Cross, C. P., Evans, C., Kearney, R., de la Torre, I., Whiten, A., & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms7029

Muller, R. U., Ranck, J. B., & Taube, J. S. (1996). Head direction cells: Properties and functional significance. *Current Opinion in Neurobiology*, *6*(2), 196–206.

Mumma, J. (2009). Proofs, pictures, and Euclid. *Synthese*, *175*(2), 255–287. https://doi.org/10.1007/s11229-009-9509-9

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213.

Newcombe, N. S., Sluzenski, J., & Huttenlocher, J. (2005). Preexisting knowledge versus on-line learning: What do young infants really know about spatial location? *Psychol Sci*, *16*(3), 222–227.

Newcombe, N. S., & Uttal, D. H. (2006). Whorf versus Socrates, round 10. *Trends in Cognitive Sciences*, *10*(9), 394–396. https://doi.org/10.1016/j.tics.2006.07.008

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annu Rev Neurosci*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Niso, G., Gorgolewski, K. J., Bock, E., Brooks, T. L., Flandin, G., Gramfort, A., Henson, R. N., Jas, M., Litvak, V., T. Moreau, J., Oostenveld, R., Schoffelen, J.-M., Tadel, F., Wexler, J., & Baillet, S. (2018). MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific Data*, *5*(1), 180110. https://doi.org/10.1038/sdata.2018.110

Noser, R., & Byrne, R. W. (2007). Travel routes and planning of visits to out-of-sight resources in wild chacma baboons, Papio ursinus. *Animal Behaviour*, *73*(2), 257–266. https://doi.org/10.1016/j.anbehav.2006.04.012

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education*, *2*(1), 13–22. https://doi.org/10.1016/j.tine.2013.01.001

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press.

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, *8*(7), 315–324. https://doi.org/10.1016/j.tics.2004.05.009

Pascual-Marqui, R. D., Faber, P., Kinoshita, T., Kochi, K., Milz, P., Nishida, K., & Yoshimura, M. (2018). Comparing EEG/MEG neuroimaging methods based on localization error, false positive activity, and false positive connectivity. *BioRxiv*, 269753.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behav Brain Sci*, *31*(2), 109–130; discussion 130-178. https://doi.org/10.1017/S0140525X08003543

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1), 103. https://doi.org/10.1038/s41597-019-0104-8

Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 1075–1089. https://doi.org/10.1037/0096-1523.17.4.1075

Philbrick, O. (1966). *A study of shape recognition using the medial axis transformation*. Air Force Cambridge Research Laboratories, Office of Aerospace Research ….

Piantadosi, S. T. (2011). *Learning and the language of thought* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/68423

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), 392–424. https://doi.org/10.1037/a0039980

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education Enhances the Acuity of the Nonverbal Approximate Number System. *Psychological Science*, *24*(6), 1037–1043. https://doi.org/10.1177/0956797612464057

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503.

Pimenta, F., & Tirapicos, L. (2015). Megalithic Cromlechs of Iberia. *Handbook of Archaeoastronomy and Ethnoastronomy*, 1153–1162.

Planton, S., Kerkoerle, T. van, Abbih, L., Maheu, M., Meyniel, F., Sigman, M., Wang, L., Figueira, S., Romano, S., & Dehaene, S. (2020). *Mental compression of binary sequences in a language of thought*. PsyArXiv. https://doi.org/10.31234/osf.io/aez4w

Planton, S., Kerkoerle, T. van, Abbih, L., Maheu, M., Meyniel, F., Sigman, M., Wang, L., Figueira, S., Romano, S., & Dehaene, S. (2021). A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLOS Computational Biology*, *17*(1), e1008598. https://doi.org/10.1371/journal.pcbi.1008598

Ponce, C. R., Xiao, W., Schade, P. F., Hartmann, T. S., Kreiman, G., & Livingstone, M. S. (2019). Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences. *Cell*, *177*(4), 999-1009.e10. https://doi.org/10.1016/j.cell.2019.04.005

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*(Supplement C), 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Premack, D. (1988). Minds with and without language. In W. L. (Ed.), *Thought without language* (pp. 46–65). Clarenton Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *The Behavioral and Brain Sciences*, *4*, 515–526.

Prewett, P. N., Bardos, A. N., & Naglieri, J. A. (1988). Use of the Matrix Analogies Test-Short Form and the Draw a Person: A quantitative scoring system with learning-disabled and normal students. *Journal of Psychoeducational Assessment*, *6*(4), 347–353.

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences*, *107*(27), 12116–12120. https://doi.org/10.1073/pnas.0914044107

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–365. https://doi.org/10.1017/S0140525X99002022

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*(6), 481–495. https://doi.org/10.1037/h0029964

Restle, F. (1973). Serial pattern learning: Higher order transitions. *Journal of Experimental Psychology*, *99*(1), 61–69. https://doi.org/10.1037/h0034751

Reynolds, C. R., & Hickman, J. A. (2004). *Draw-a-person Intellectual Ability Test for Children, Adolescents and Adults*. Par, Incorporated.

Rieber, R. (2013). *Dialogues on the Psychology of Language and Thought*. Springer Science & Business Media.

Romano, S., Sigman, M., & Figueira, S. (2013). LT^2C^2: A language of thought with Turing-computable Kolmogorov complexity. *Papers in Physics*, *5*, 050001.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Roumi, F. A., Marti, S., Wang, L., Amalric, M., & Dehaene, S. (2021). Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. *Neuron*, *0*(0). https://doi.org/10.1016/j.neuron.2021.06.009

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., Ellis, K., Tenenbaum, J., & Dehaene, S. (2021). *A language of thought for the mental representation of geometric shapes*. PsyArXiv. https://doi.org/10.31234/osf.io/28mg4

Sablé-Meyer, M., Fagot, J., Caparos, S., Kerkoerle, T. van, 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*, *85*(6), 2232–2246. https://doi.org/10.1111/cdev.12319

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Satlow, E., & Newcombe, N. (1998). When is a triangle not a triangle? Young children's developing concepts of geometric shape. *Cognitive Development*, *13*(4), 547–559. https://doi.org/10.1016/S0885-2014(98)90006-5

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1), 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Geiger, F., Schmidt, K., Yamins, D. L. K., & DiCarlo, J. J. (2018). *Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?* [Preprint]. Neuroscience. https://doi.org/10.1101/407007

Schrimpf, M., Kubilius, J., Lee, M. J., Murty, N. A. R., Ajemian, R., & DiCarlo, J. J. (2020). Integrative Benchmarking to Advance Neurally Mechanistic Models of Human Intelligence. *Neuron*, *108*(3), 413–423. https://doi.org/10.1016/j.neuron.2020.07.040

Schwartz, M., Day, R. H., & Cohen, L. B. (1979). Visual Shape Perception in Early Infancy. *Monographs of the Society for Research in Child Development*, *44*(7), 1. https://doi.org/10.2307/1165963

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1963). Cultural Differences in the Perception of Geometric Illusions. *Science*, *139*(3556), 769–771. https://doi.org/10.1126/science.139.3556.769

Semendeferi, K., Lu, A., Schenker, N., & Damasio, H. (2002). Humans and great apes share a large frontal cortex. *Nature Neuroscience*, *5*(3), 272–276. https://doi.org/10.1038/nn814

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. MIT Press.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Sherwood, C. C., & Smaers, J. B. (2013). What's the fuss over human frontal lobe evolution? *Trends in Cognitive Sciences*, *17*(9), 432–433. https://doi.org/10.1016/j.tics.2013.06.008

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., Miller, K. J., & Parvizi, J. (2013). A Brain Area for Visual Numerals. *Journal of Neuroscience*, *33*(16), 6709–6715. https://doi.org/10.1523/JNEUROSCI.4558-12.2013

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*(4), 273–296. https://doi.org/10.1016/j.cogpsych.2011.03.001

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*(5), 369.

239

Singh, M., Seyranian, G. D., & Hoffman, D. D. (1999). Parsing silhouettes: The shortcut rule. *Perception & Psychophysics*, *61*(4), 636–660.

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *J Exp Psychol Gen*, *133*(3), 398–414.

Solomonoff, R. (1964). A Formal Theory of Inductive Inference: Parts 1 and 2. *Information and Control*, *7*, 1–22 and 224–254.

Spaepen, E., Coppola, M., Flaherty, M., Spelke, E., & Goldin-Meadow, S. (2013). Generating a lexicon without a language model: Do words for number count? *Journal of Memory and Language*, *69*(4), 496–505. https://doi.org/10.1016/j.jml.2013.05.004

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences*, *108*(8), 3163–3168. https://doi.org/10.1073/pnas.1015975108

Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond Core Knowledge: Natural Geometry. *Cognitive Science*, *34*(5), 863–884. https://doi.org/10.1111/j.1551-6709.2010.01110.x

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In *Language in mind: Advances in the study of language and thought.* (pp. 277–311). MIT Press.

Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., & Kriegeskorte, N. (2020). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLOS Computational Biology*, *16*(10), e1008215. https://doi.org/10.1371/journal.pcbi.1008215

Sun, Z., & Firestone, C. (2021). Curious Objects: How Visual Complexity Guides Attention and Engagement. *Cognitive Science*, *45*(4). https://doi.org/10.1111/cogs.12933

Sun, Z., & Firestone, C. (2022). Seeing and speaking: How verbal "description length" encodes visual complexity. *Journal of Experimental Psychology: General*, *151*(1), 82–96. https://doi.org/10.1037/xge0001076

Tanaka, M., Tomonaga, M., & Matsuzawa, T. (2003). Finger drawing by infant chimpanzees (Pan troglodytes). *Animal Cognition*, *6*(4), 245–251. https://doi.org/10.1007/s10071-003-0198-3

Taulu, S., & Kajola, M. (2005). Presentation of electromagnetic multichannel data: The signal space separation method. *Journal of Applied Physics*, *97*(12), 124905.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tomasello, M. (2000). *The cultural origins of human cognition*. Harvard University Press.

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

Tootell, R. B. H., Dale, A. M., Sereno, M. I., & Malach, R. (1996). New images from human visual cortex. *Trends in Neurosciences*, *19*(11), 481–489. https://doi.org/10.1016/S0166-2236(96)10053-9

Treisman, A., & Gormican, S. (1988). *Feature Analysis in Early Vision: Evidence From Search Asymmetries*. 34.

Treisman, A., & Souther, J. (1985). *Search Asymmetry: A Diagnostic for Preattentive Processing of Separable Features*. 26.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. of Math*, *58*(345–363), 5.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Twyman, A. D., & Newcombe, N. S. (2010). Five Reasons to Doubt the Existence of a Geometric Module. *Cognitive Science*, *34*(7), 1315–1356. https://doi.org/10.1111/j.1551-6709.2009.01081.x

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences*, *113*(10), 2744–2749. https://doi.org/10.1073/pnas.1513198113

Uusitalo, M. A., & Ilmoniemi, R. J. (1997). Signal-space projection method for separating MEG or EEG into components. *Medical and Biological Engineering and Computing*, *35*(2), 135–140.

Van der Waerden, B. L. (2012). *Geometry and algebra in ancient civilizations*. Springer Science & Business Media.

Vigo, R., & Doan, C. A. (2015). The structure of choice. *Cognitive Systems Research*, *36–37*, 1–14. https://doi.org/10.1016/j.cogsys.2015.02.001

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, *55*(1), 143–156. https://doi.org/10.1016/j.neuron.2007.05.031

Von Petzinger, G. (2009). *Making the abstract concrete: The place of geometric signs in French Upper Paleolithic parietal art* [PhD Thesis].

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172.

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.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*,

188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., Sigman, M., & Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage*, *186*, 245–255. https://doi.org/10.1016/j.neuroimage.2018.10.061

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Current Biology: CB*, *25*(15), 1966–1974. https://doi.org/10.1016/j.cub.2015.06.035

Wertheimer, M. (1912). Experimentelle studien uber das sehen von bewegung. *Zeitschrift Fur Psychologie*, *61*.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology.* (pp. 71–88). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005

Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: Effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2007–2022. https://doi.org/10.1098/rstb.2012.0098

Wilder, J., Feldman, J., & Singh, M. (2016). The role of shape complexity in the detection of closed contours. *Vision Research*, *126*, 220–231. https://doi.org/10.1016/j.visres.2015.10.011

Wolfe, J. M. (1998). What Can 1 Million Trials Tell Us About Visual Search? *Psychological Science*, *9*(1), 33–39. https://doi.org/10.1111/1467-9280.00006

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics*, *63*(3), 381–389. https://doi.org/10.3758/BF03194406

Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X.-J., Yang, T., Dehaene, S., Tang, S., Min, B., & Wang, L. (2022). Geometry of sequence working memory in macaque prefrontal cortex. *Science*, *375*(6581), 632–639. https://doi.org/10.1126/science.abm0204

Xu, T., Nenning, K.-H., Schwartz, E., Hong, S.-J., Vogelstein, J. T., Goulas, A., Fair, D. A., Schroeder, C. E., Margulies, D. S., Smallwood, J., Milham, M. P., & Langs, G. (2020). Cross-species functional alignment reveals evolutionary hierarchy within the connectome. *NeuroImage*, *223*, 117346. https://doi.org/10.1016/j.neuroimage.2020.117346

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 201216803. https://doi.org/10.1073/pnas.1216803110

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240.

Zhan, M., Goebel, R., & de Gelder, B. (2018). Ventral and Dorsal Pathways Relate Differently to Visual Awareness of Body Postures under Continuous Flash Suppression. *Eneuro*, *5*(1), ENEURO.0285-17.2017. https://doi.org/10.1523/ENEURO.0285-17.2017

Zhang, H., Zhen, Y., Yu, S., Long, T., Zhang, B., Jiang, X., Li, J., Fang, W., Sigman, M., Dehaene, S., & Wang, L. (2022). Working Memory for Spatial Sequences: Developmental and Evolutionary Factors in Encoding Ordinal and Relational Structures. *The Journal of Neuroscience*, *42*(5), 850–864. https://doi.org/10.1523/JNEUROSCI.0603-21.2021

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Long Summary (English)

> Ainsi certes, nous ne pourrions jamais connaître le triangle géométrique par celui que nous voyons tracé sur le papier, si notre esprit d'ailleurs n'en avait eu l'idée
>
> Descartes, Cinq Rép., AT VII, 382 ; OC IV-1, 574.

What is a point? Euclid famously kickstarted geometry as we know it today with his definition n°1, "A point is that which has no part" ("Σημειον εστιν, ου μερος ουθεν" (Byrne & Euclid, 1847)). There is no physical entity to which this definition would apply; a point must therefore exist only in the mind of the beholder. What mental and neural mechanisms make it possible to entertain such concepts? Are those mechanisms only available to humans, and are they deeply tied to natural language? In my PhD work, I propose that even the simplest geometric concepts are uniquely human, and that they lie at the foundation of a rich generative system of shapes that behaves like an internal mental language. I argue that humans across ages, cultures and education levels share this sense of geometry, and I explore its neural mechanisms.

In my thesis I explore the possibility that all humans, and humans only, mentally represent shapes as much more than their visual impressions by superimposing structure onto their perception. More specifically, I argue that humans possess the ability to represent geometric shapes using an internal language specialized for that purpose. I provide evidence that they naturally deploy this mechanism when faced with geo-

metric shapes, in addition to other mechanisms of perception. I show that this is true independently from education, culture and age by testing diverse populations on the same task. As a contrasting point, I provide evidence that baboons either do not possess this competence, or do not deploy it even when it would be extremely useful. I model the difference between humans and non-human primates with two very contrasting classes of models.

I then find neural evidence in favor of cognitive processes related to these two models, using both MEG and fMRI in human adults, 6-year-old children, and to a limited extent, 3-month-old infants. Finally, building on computer science and AI together with cognitive psychology, I make a concrete proposal for a candidate mental language for geometric shapes. I test its behavioral validity, show that a neurosymbolic model can implement a theory of shape perception as program induction, and independently verify that some pivotal hypotheses underlying the language I propose must hold true of other propositions as well.

Below, I will provide a very brief overview of the production of geometric shapes across history and culture, together with a review of the experimental cognitive science literature regarding geometry in humans. Then I will introduce the Language of Thought Hypothesis (LOTH), with an emphasis on *programs* as candidates for mental representations. I will connect the LOTH to cognitive science research performed using information theory. Finally, will outline and briefly summarize the structure of the work reported in my thesis, chapter by chapter.

Evidence of Geometrical Productions in Humans

Paleontological Evidence

Evidence for abstract concepts of geometry, including rectilinearity, parallelism, perpendicularity and symmetries, is widespread throughout prehistory. About 70,000 years ago, **Homo Sapiens** at Blombos cave carved a piece of ocher with three interlocking sets of parallel lines forming equilateral triangles, diamonds and hexagons (Henshilwood et al., 2002). Much earlier, approximately 540,000 years ago, **homo Erectus** in Java carved a zig-zag pattern on a shell (J. C. A. Joordens et al., 2015). Such a zig-zag may look simple, but it approximately respects geometric constraints of equal lengths, equal

Figure 13: **Geometric shapes in human cultural history.** A, examples of small- and large-scale geometric drawings and constructions (From left to right and top to bottom: an engraved slab from Blombos caves dating about 70.000 years ago (Henshilwood et al., 2002); zigzag pattern engraved on a shell in Java approximately 540.000 years ago (J. C. A. Joordens et al., 2015); Boscawen-Ûn's Bronze Age elliptical cromlech in Cornwall; spiral stone engraving on Signal Hill in Saguaro National Park, Arizona, dated 550 to 1550 years ago; geometrical shapes below the painting of a Megaloceros in Lascaux, France, typically dated to be 17,000 years old)

angles and parallelism, and is undoubtedly attributed to the *homo* genus.

Even earlier, about ~1.8 million years ago, ancient humans have been carving spheroids (sphere-like stones) and bifaces — stones possessing two orthogonal planes of symmetry (Le Tensorer, 2006). The vast number of bifaces, their near-perfect symmetry (which is not required for them to operate as efficient tools (Le Tensorer, 2006)), and the archeological evidence that many were never used as tools, suggest that an aesthetic drive for symmetry was already present in ancient humans.

Anthropological and Cross-Cultural Evidence

Contemporary cognitive anthropology corroborates those findings. Cognitive tests performed in relatively isolated human groups such as the Mundurucu from the Amazon, the Himba from Namibia, or indigenous groups from Northern Australia, show that in the absence of formal western education in mathematics, adults and even children

already possess strong intuitions of numerical and geometric concepts (Amalric et al., 2017; Butterworth et al., 2008; Dehaene et al., 2006; Izard et al., 2011; Pica et al., 2004; Sablé-Meyer, Fagot, et al., 2021).

Indeed, adults without formal western education share with Western preschoolers a large repertoire of abstract geometric concepts (Dehaene et al., 2006) and use them to capture the regularities in spatial sequences (Amalric et al., 2017) and quadrilateral shapes such as squares or parallelograms (Sablé-Meyer, Fagot, et al., 2021). They also possess sophisticated intuitions of how parallel lines behave under planar and spherical geometry, such as the unicity of a parallel line passing through a given point on the plane (Izard et al., 2011).

Developmental Evidence

Another piece of evidence arises from developmental data. Preschoolers and even infants have been shown to possess sophisticated intuitions of space (Hermer & Spelke, 1994; Landau et al., 1981; Newcombe et al., 2005), spatial sequences (Amalric et al., 2017), and mirror symmetry (Bornstein et al., 1978). Indeed, preschoolers' drawings already show a tendency to represent abstract properties of objects rather than the object itself. Although they look primitive, drawings of a house as a triangle on top of a square, or a person as a stick figure with a round head, suggest a remarkable capacity for abstracting away from the actual shape and attending to its principal axes, at the expense of realism. Numerous tests leverage this geometric competence to assess a child's cognitive development by counting the number of correct or incorrect abstract properties, for instance when asked to draw a person (Goodenough, 1926; Harris, 1963; Long et al., 2019; Prewett et al., 1988; Reynolds & Hickman, 2004). There is some evidence, however limited, that this ability may be specifically human: when given pencils or a tablet computer, other non-human primates do not draw any abstract shapes or recognizable figures, but mostly generate shapeless scribbles (Saito et al., 2014; Tanaka et al., 2003).

Programs as Candidates for Mental Representations: A Take on the Language of Thought Hypothesis

> "Humans have a multi-domain capacity and proclivity to infer tree structures from strings, to a degree that is difficult or impossible for most non-human animal species"
>
> W. Tecumseh Fitch (2014)

Geometric Primitives of Cognition

In cognitive science, the first application of information theory to visual perception comes from work by Attneave (1954). He observed that a given a visual perception is redundant, in the sense that many portions could be hidden and yet successfully inferred or recovered by a human observer. From this, he argues that the mental representation of a visual percept might leverage this redundancy to compress the visual information. For geometric shapes specifically, the earliest theory of programs representing geometric shapes comes from Leeuwenberg and colleagues(E. L. Leeuwenberg, 1969, 1971; Boselie & Leeuwenberg, 1986), who proposed a formal coding language for 2- and 3-dimensional shapes. They argue that the mental representation of a shape is as complex as the smallest program in that language, a property I also defend in **chapter 5**. In fact, they already observe that some elements of their proposed language are quite general, and that they could be applied to the compression of auditory sequences as well, a property that was found again in recent work on geometric sequences in the visual and auditory domain (Amalric et al., 2017; Planton et al., 2021). In addition to the theoretical claims, they provide some empirical support for their proposition, but conclude in saying, "[...] for the time being [the proposed coding procedure] will hardly lend itself to computer programming", but nowadays this does not hold true anymore, and I tackle this in **chapter 5**. I compare this other approaches, including from Leyton (Leyton, 1984, 2003), in the thesis.

Kolmogorov Complexity & MDL

A long-standing cognitive hypothesis is that the brain excels at compressing information. Indeed, in the presence of structure in stimuli, either visual, auditory or other, participants' score improves in a wide

variety of tasks. The first observation of this phenomenon comes from (Miller, 1956), who states: "I have fallen into the custom of distinguishing between bits of information and chunks of information. [...] The span of immediate memory seems to be almost independent of the number of bits per chunk, at least over the range that has been examined to date." Immediate examples include remembering words, where the main factor is the number of words and not the number of letters, but similar observations are pervasive in psychology.

A strong version of this hypothesis states that the brain finds structure in a richer way than chunking, and relies on generative (programming) languages: the complexity of a given piece of information is the length of the shortest program that generates that information (Goodman et al., 2012; Tenenbaum et al., 2011). This hypothesis connects to the predictive coding hypothesis, as the ability to predict and generate can be deeply tied to the generative languages. Furthermore, for probabilistic programing languages, generative language theories have the ability to account for both success and mistakes in human behavior: a crucial feature of a good candidate theory of cognition. What's more, such theories can entertain the coexistence of several possible mental representation with different probabilities, and therefore account for ambiguous representations.

A lot hinges on the choice of the programming language, as various propositions might make wildly different predictions. Because of this, it's unclear that one can find a unifying proposition that can account for very different stimuli (auditory, visual, intuitive physics, etc.). Thankfully, an entire research domain studies a related problem: the field of information theory. The field studies theoretical questions such as "what is the most information-efficient way to represent a given set and its elements". In the introduction of my thesis I describe two related notion from that field and show how they relate: Kolmogorov Complexity, and Minimum Description Length (MDL). I discuss how influential and useful they can be when specifying cognitive theories.

Cognition and Program Induction

If mental representations have program-like properties, then it is crucial to offer a theory of how the programs are inferred: how do we go from the sensory inputs to the structured representation? What are good models of "mental program building"?

In computer science, the subfield tackles this question is that of "Program Induction": the problem of program induction is the problem of finding a program given, typically, a set of input-output example pairs. An immediate observation is that this is in principle an impossible problem: there are arbitrarily many programs that may work, but some will behave differently on new input – in the absence of which the notion of "correct program" cannot be decided. But given enough examples, the trivial program that encodes explicitly the input/output pairs becomes very costly, and the length of the program can be used as a selection strategy: I claim that humans search, and find, the shortest program when they understand a *geometric shape*, and therefore the complexity of a shape will be predicted by its MDL.

A baseline for solving program induction is program enumeration: recursively enumerate all possible programs in a programming language's grammar until you hit a program that satisfies the constraints. But this procedure is a poor candidate for a cognitive implementation of program induction under the MDL hypothesis: if we consider that the complexity of a certain program is a function of its MDL, then under that approach the complexity of *finding* that program would grow exponentially with its MDL. However, several methods have been devised to try and keep this combinatorial explosion in check, I describe them in the introduction of the thesis and I rely on DreamCoder (Ellis et al., 2021), which I helped implement, in **chapter 5**.

Structure of the chapters

In chapter 1, I show that even the detection of an intruder among quadrilaterals distinguishes humans from non-human primates. Using an intruder task with quadrilateral shapes of different regularity, from highly regular squares to irregular quadrilaterals, I show that humans of diverse education, age and culture share a sense of geometric complexity: some shapes systematically elicit better performance than others. For example, when looking for an intruder among squares, participants are fast and accurate, but when looking for an intruder among identical irregular shapes, participants are slow and inaccurate. I show that this is also true with different paradigms such as visual search or introspective rating, making the case for a strong, easy to replicate effect.

Chapter 1: Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity

Task: pick the intruder

Populations:
- N≈800 Educated Adults
- N=44 Himbas
- N=28 Kindergartners
- N=156 First graders
- N≈26 Baboons

Several tasks tested in educated adults.

Models: Several Neural Networks, custom symbolic models

Modality: touchscreens, online or in the lab, measuring errors and Response Times

Result: strong dissociation between humans and Baboons

Predictor amplitude

Symbolic Model   Neural Network Model

Humans   Baboons

Chapter 2: M/EEG evidence for symbolic and non-symbolic neural mechanisms of geometric shape perception in adults and infants

Task: passive presentation with oddballs

Miniblock with 1st shape

SOA 1s, ISI 200ms

6 reference shapes   20 reference shapes, 4 oddballs

Modality and population

N=20 Adults, MEG & anatomical MRI

N=40 3-months-olds with EEG nets Still ongoing

Results: spatial & temporal dissociation of the models

Neural Network Model   Symbolic Model

Onset   400ms   time

Chapter 3: Geometric Shape Perception Activates Brain Regions Associated with Mathematical Cognition, an fMRI study

Task: Category Localizer with geometric shapes

BON   走   2 - 1

Modality and population

N=20 Adults 1h30 of fMRI

N=20 Six-years-old 1h00 of fMRI

Results: In both age groups, comparing single items to shapes reveals a lower activation in ventral pathways and a math-like activation in the IPS

Adults   6 y.o.

Contrast: ⬡ > ( )

t   -10   0   10

Chapter 4: Categorical perception of right angles in humans and baboons

Task: delayed match-to-sample; **Method**: online with touchscreen

Population:
N≈153 educated adults
N=26 baboons

Stimuli:
Target angle in 0°, 30°, 60°, 90°, 120°, 150°, 180°
Distractors: random with ±10°, 20°, 30°, 40°, 50°

Results: categorical perception of right angle only in a subset of conditions in humans, no conclusion yet in baboons

P(chosen|presented)

0° 30° 60° 90° 120°150°180°

1   5/6   4/6   3/6   2/6   1/6   0

-30 10 50 90 130 170 210

Chapter 5: A Language of Thought for the Mental Representation of Geometric Shapes

Core proposition: shape perception is performing program induction, Thus shape complexity increases With Minimal Description Length

Tasks: delayed match to sample with shapes of ranging MDL

repeat

3   concat

draw   draw

Modeling: DreamCoder, State of the Art Program induction model

Modality & population: touchscreens, online

Exp. 1: N=57 humans, N=26 baboons
Exp. 2: N=170 humans

Results: shape complexity correlates with MDL Concatenation, repetition and embedding are pervasive in geometric shape perception

Residuals on all other visual predictors

400   200   0

Choice Time
$R^2 = 0.63$
$p < 0.01$

Encoding Time
$R^2 = 0.5$
$p = 0.02$

Program cost

2   4   6   8   10   12

Figure 14

On the other hand, baboons lack this sense even after adequate extensive training: over a sequence of increasingly hard tasks, I confirm that baboons can understand the intruder task and generalize across stimuli, but fail to generalize to quadrilateral shapes. With extensive training on the quadrilateral shapes, their overall average performance eventually reaches to the level of untrained 5-year-old children, but baboons display no evidence of the geometric regularity effect detected in humans: all shapes elicit similar behavior.

Many different artificial neural networks of object recognition, of different architecture, fit all baboons' data well, but explaining humans' behavior requires using additional symbolic properties such as the presence of right angles (I zoom in on this specific symbolic property in **chapter 4**). Symbolic models of the humans' behavior generalize to several related tasks, which validates the robustness of the choice of geometric primitives included in the model. This sharp dissociation suggests that two strategies are available to encode geometric shapes: both humans and non-human primates share a perceptual strategy, well captured by models of the ventral visual pathway, but only human have access to an exact, symbolic strategy. This indicates a putative signature of human singularity and provides a challenge for nonsymbolic models of human shape perception going forward.

This chapter corresponds to an article published in PNAS under the following reference: **Sablé-Meyer, M.**, Fagot, J., Caparos, S., Kerkoerle, T. van, 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).

In chapter 2, I use neuroimaging techniques to shed light on the neural implementation of the two strategies put forward in **chapter 1**.

First, I use a new behavioral visual search task to measure the confusion matrix across the quadrilaterals designed in **chapter 1**, and model it with the symbolic model and the neural network. I also show that a data-driven decomposition of the complexity matrix coincides with the already proposed symbolic model. This provides me with both a replication of the results from **chapter 1**, a data-driven way to see what drives the geometrical regularity effect, and a confusion matrix that I will be able to use both in **chapter 2** and in **chapter 3** to model neuroimaging data.

Then, to make a task compatible with both adults and infants, I design a purely passive oddball paradigm: participants are shown the quadrilateral shapes of **chapter 1** centered on a screen with random scaling and rotation, once per second, with the possibility of a deviant every once in a while. In twenty adults tested for 90 minutes in MEG, I can decode the brain signal associated with oddballs, and the performance of the decoder replicates the geometric regularity effect. Using Representational Similarity Analysis, I show that the brain signal first resembles the neural network model, and then the symbolic model, indicating that the two models exist in adults despite the fact that their behavior only reflects the symbolic model. Using source reconstruction, I show that the neural network model corresponds to a bilateral occipital cluster of sources, while the symbolic model is associated with a wide cluster which includes frontal sources and sources in the dorsal pathway.

This provides a replication of the behavioral data with a passive presentation paradigm, wherein intruders are detectable inasmuch as they are intruder of regular shapes. On top of this, this confirms the proposition from **chapter 1** that both strategies are present in educated adults, a result invisible from the behavior alone but necessary to understand the data from the first chapter. Finally, this provides both a spatial and temporal dissociation of the bottom-up ventral visual pathway of shape processing, early and transitory, located in the occipital area, followed by a top-down symbolic processing, spanning over a much broader brain area.

In two groups of 20 three-month-old infants, I try to replicate this experiment while measuring the EEG activation. Preliminary results are inconclusive as to the intruder detection, but it appears that the different shapes are represented differently from one another in a way that is compatible with the geometric regularity effect in infants already. The results from infants are preliminary and I am still working on this project.

In chapter 3, I use 3T fMRI in twenty adults and twenty 6-year-old children to more accurately localize the areas associated with geometric shape perception across education.

Using a category localizer, with typical visual categories (faces, objects, etc.) as well as words, numbers, and geometric shapes, I find that passive geometric perception, when contrasted to other visual

objects, over activates a network of bilateral IPS brain areas typically associated with non-linguistic mathematic and number processing. Conversely, geometric shapes are under-activating the ventral visual pathways, when compared to the other visual categories such as faces, objects and houses. This result holds for both age groups, indicating that the core mechanisms at play in geometric shape are already present in six-year-old children. Further exploration of the ventral, under activated cluster reveals that even though intuitively one might think that shape processing resembles letter reading, or visual processing of faces, or visual processing of objects, or other general visual processing, none of the areas typically associated to these categories are strongly activated by geometric shapes.

Inside the fMRI, participants additionally performed a variation of the intruder test from **chapter 1** where they had to indicate the location of an intruding shape on screen. Both age groups performed a simple version of the task, and additionally adults performed a harder version of the task. The behavior from inside the scanner replicates our previous experiments really well, including an overall ranking of performance across age groups and difficulty, indicating that even under moderate time pressure the geometric regularity effect is present.

Data from adults in the easy condition yielded broad, bilateral clusters that show increased activity as a function of the complexity of the shape. On top of this, Representational Similarity Analysis detected significant clusters associated with both the neural network model, mostly inside visual areas, and the symbolic model, mostly in around the IPS. Both models were as defined in **chapter 1** and **chapter 2**. There were also some areas associated to both models, mostly in decision processing areas, suggesting an integrating process weighting the two sources of information, possibly when participants have to answer.

In chapter 4, I focus on a single geometric property: right angles. I compare the behavior of educated adults and baboons in a delayed match-to-sample task involving various angles and relative distractors. The question is whether specific angles, namely 0°, 90° and 180° (corresponding to aligned lines, perpendicular lines, and parallel lines) elicit behavior that is categorically different from other angles, as indicated mainly by a much sharper sensitivity to deviation from the reference angle. The task tests equally spaced angles: 0°, 30°, 60°,

90°, 120°, 150° and 180°. With each trial, distractors are sampled from neighboring angles distant from 10°, 20°, 30°, 40°, 50° on either side. Then the comparison of the performance between an angle and its neighbors (e.g. 90° versus 80° and 100°, compared to 30° versus 20° and 40°) informs us on participants' sensitivity to specific angles.

In trying to replicate classical results on the categorical perception of right angles in adults, I put forward the fact that several properties are required for right angles to elicit behavior that differs from other neighboring angles: (i) the stimuli need to be displayed long enough, and (ii) no other low-level property can be used to perform the task. Both results coincide with the idea that symbolic properties require attention as put forward in **chapter 1**, but these are new results that were not evident given the literature on the categorical perception of right angles.

I also present early data collected in baboons where the performance is very good for 0° and 180°, and poor everywhere else. However, data collected so far only correspond to a subset condition where humans do not display categorical perception of geometric shape either, and therefore its absence in baboons is not theoretically very informative.

In chapter 5, I go beyond a small set of highly controlled quadrilaterals and I set to try to account for *all geometric shapes* produced by humans. To do so, I make a concrete proposition for a generative mental language of geometric shapes inspired by attested human geometric productions, which has to satisfy two constraints: short programs must generate simple shapes (for humans), and simple shapes must have small programs.

It is in that chapter that I fully develop the argument that perceiving a shape is equivalent to program induction: finding the shortest mental program in this internal language that generates that particular shape. First, for the language I propose, I show that program induction is in principle a tractable problem: to that end, I use the DreamCoder algorithm I helped implement before my PhD, and show that it can successfully find the shortest program for a given shape. Then, I use this language to generate shapes of increasingly high complexity and show that humans' performance in a match-to-sample task for a shape correlates with the length of its shortest program, above and beyond many other perceptual features that are otherwise typically attested in the literature.

On top of this, I put forward the observation that DreamCoder implements a plausible theory of how different cultures produce drawings that look very different, e.g. Celtic curvilinear productions versus Greek rectilinear productions. Although DreamCoder has a single language for geometry to start with, when it is trained it learns both new primitives and new biases as it solves new tasks. In doing so, slowly DreamCoder drifts in specific directions throughout its training procedure. The same could be true of cultural drifts, where basic shapes are shared by everyone (circles and squares for example), but complex shapes reuse intermediary building blocks that are learned from the basic primitives, and over time this constitutes a visible drift in the productions of different cultures.

Finally, to decouple my result from my exact language proposition, I also derive general additive rules that any alternative languages must obey. I design a new experiment with a new set of shapes to test this hypothesis and confirm that concatenation, repetition and embedding are essential to capture the compositional nature of geometric shape complexity. To make this more concrete, I show that the mental representation of a shape that is "a square made out of smaller squares" induces a saving, wherein a saving is induced by the fact that both the local and the global shapes are identical, indicating the necessity for recursion in humans' geometric shape perception.

This chapter corresponds to an article published in Cognitive Psychology under the following reference: **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.

Conclusion

Natural language is not the only hallmark of humans' singular cognitive abilities. In the work presented here, and in line with the language of thought literature, I argue that there could exist several internal "languages of thought", and that geometric shapes could be a pivotal way to study non-linguistic, high-level structured cognitive processes. To do so, I show that cognition involving geometric shapes requires a set of discrete, symbolic mental representations that act as an internal mental language with compositional properties. With cultural psychology and developmental psychology methodology, I argue that this ability

does not depend on education, age or culture. Using comparative cognition methods, I provide evidence that non-human primates cannot access these symbolic representations. I claim that under that view, perceiving a shape is comparable to the process of program induction: finding the shape's shortest representation in the internal mental language. I show neural evidence of this in humans, using both MEG and fMRI, and I model the data with state of the arts models inspired by several fields: neural networks from the AI literature, neurosymbolic models from AI and computer science, and discrete symbolic models commonly employed in psychology.

Résumé Détaillé (Français)

> Ainsi certes, nous ne pourrions jamais connaître le triangle géométrique par celui que nous voyons tracé sur le papier, si notre esprit d'ailleurs n'en avait eu l'idée
>
> ———————————————————————————
>
> Descartes, Cinq Rép., AT VII, 382 ; OC IV-1, 574.

Qu'est-ce qu'un point ? Euclide a donné le coup d'envoi de la géométrie telle que nous la connaissons aujourd'hui avec sa définition n°1, "Un point est ce qui n'a aucune partie" ("Σημειον εστιν, ου μερος ουθεν" (Byrne & Euclid, 1847)). Il n'existe aucune entité physique à laquelle cette définition puisse s'appliquer ; un point ne peut donc exister que dans l'esprit de celui qui y pense. Quels mécanismes mentaux et neuronaux permettent de former de tels concepts ? Ces mécanismes sont-ils uniquement présents chez l'humain, et sont-ils liés irrémédiablement à sa possession du langage naturel ? Dans mon travail de thèse, je défends l'idée que même les concepts géométriques les plus simples ne sont présents que chez les humains, et qu'ils sont à la base d'un système génératif de formes qui se comporte comme un langage mental interne. Je montre que les humains, quels que soient leur âge, leur culture et leur niveau d'éducation, partagent ce sens de la géométrie, et j'explore les mécanismes neuronaux qui sous-tendent cette capacité.

Dans ma thèse, j'expose l'idée que tous les humains, et seulement les humains, représentent mentalement les formes comme étant bien plus que leurs impressions visuelles, en superposant une structure à la perception visuelle. Plus précisément, je dis que les humains possèdent la capacité de représenter les formes géométriques en utilisant un langage interne, spécialisé pour cette représentation. Je fournis

des preuves expérimentales que les humains déploient naturellement ce mécanisme lorsqu'ils sont confrontés à des formes géométriques, en plus d'autres mécanismes de perception. Je montre que ce résultat ne dépend ni de l'éducation, ni de la culture, ni de l'âge, en testant diverses populations sur la même tâche. À titre de contraste, je fournis des preuves que les babouins ne possèdent pas cette compétence, ou ne la déploient pas même lorsqu'elle serait extrêmement utile. Je modélise le contraste entre les humains et les primates non humains avec deux classes de modèles fondamentalement différents.

Je mets ensuite en avant des preuves neuronales en faveur de processus cognitifs liés à ces deux modèles, en utilisant à la fois la MEG et l'IRMf chez des adultes humains, des enfants de 6 ans et, dans une mesure limitée, des nourrissons de 3 mois. Enfin, en m'appuyant sur l'informatique et l'IA ainsi que sur la psychologie cognitive, je fais une proposition concrète pour un langage mental pour les formes géométriques. Je teste sa validité comportementale, puis je fais la démonstration qu'un modèle neuro-symbolique peut implémenter une théorie de la perception des formes sous forme d'induction de programme, et je vérifie de que certaines hypothèses cruciales qui sous-tendent ma proposition de langage doivent être vraies pour tout autre propositions, en faisant ainsi un résultat plus général.

Ci-dessous, je commence par donner un très bref aperçu de la littérature sur la production de formes géométriques à travers l'histoire et les cultures, ainsi qu'une revue de la littérature en sciences cognitives concernant la géométrie chez l'humain. Je présenterai ensuite l'hypothèse du langage de la pensée (Language Of Thought Hypothesis, LOTH), en mettant l'accent sur les *programmes* comme candidats pour la nature des représentations mentales. Je relierai la LOTH à la recherche en sciences cognitives effectuée à l'aide de la théorie de l'information. Enfin, je présenterai et résumerai brièvement la structure du travail rapporté dans ma thèse, chapitre par chapitre.

Preuve des productions géométriques chez l'homme

Preuves paléontologiques

Les preuves de l'existence de concepts abstraits en géométrie, notamment la rectilinéarité, le parallélisme, la perpendicularité et les symétries, sont très répandues tout au long de la préhistoire. Il y a environ 70 000 ans, un **Homo Sapiens** de la grotte de Blombos a

Figure 15: Les formes géométriques dans l'histoire de l'humanité. A, exemples de dessins et de constructions géométriques à petite et grande échelle (De gauche à droite et de haut en bas : une dalle gravée des grottes de Blombos datant d'environ 70.000 ans (Henshilwood et al., 2002); motif en zigzag gravé sur un coquillage à Java il y a environ 540.000 ans (J. C. A. Joordens et al., 2015); cromlech elliptique de l'âge du bronze de Boscawen-Ûn en Cornouailles ; gravure en spirale sur pierre sur Signal Hill dans le parc national de Saguaro, Arizona, datée de 550 à 1550 ans ; formes géométriques sous la peinture d'un Mégalocéros à Lascaux, France, souvent datée de 17.000 ans)

sculpté un morceau d'ocre avec trois ensembles imbriqués de lignes parallèles formant des triangles équilatéraux, des diamants et des hexagones (Henshilwood et al., 2002). Bien plus tôt, il y a environ 540 000 ans, un **Homo Erectus** de Java a gravé un motif en zigzag sur un coquillage (J. C. A. Joordens et al., 2015). Un tel zig-zag peut sembler simple, mais il respecte approximativement les contraintes géométriques de longueurs égales, d'angles égaux et de parallélisme, et est sans aucun doute attribué au genre *homo*.

Plus tôt encore, il y a environ 1,8 million d'années, les anciens humains sculptaient des sphéroïdes (pierres en forme de sphère) et des bifaces - pierres possédant deux plans de symétrie orthogonaux (Le Tensorer, 2006). Le grand nombre de bifaces, leur symétrie quasi-parfaite (qui n'est pas nécessaire pour qu'ils fonctionnent comme des outils efficaces (Le Tensorer, 2006)), et les preuves archéologiques que beaucoup d'entre eux n'ont jamais été utilisés comme outils, suggèrent qu'un désir esthétique de symétrie était déjà présent chez ces anciens humains.

Preuves anthropologiques et interculturelles

L'anthropologie cognitive contemporaine corrobore ces résultats. Des tests cognitifs réalisés dans des groupes humains relativement isolés, comme les Mundurucu d'Amazonie, les Himba de Namibie ou les groupes indigènes du nord de l'Australie, montrent qu'en l'absence d'une éducation occidentale formelle en mathématiques, les adultes et même les enfants possèdent déjà de fortes intuitions des concepts numériques et géométriques (Amalric et al., 2017; Butterworth et al., 2008; Dehaene et al., 2006; Izard et al., 2011; Pica et al., 2004; Sablé-Meyer, Fagot, et al., 2021).

En effet, les adultes sans éducation occidentale formelle partagent avec les enfants occidentaux d'âge préscolaire un vaste répertoire de concepts géométriques abstraits (Dehaene et al., 2006) et les utilisent pour saisir les régularités des séquences spatiales (Amalric et al., 2017) et les formes quadrilatérales telles que les carrés ou les parallélogrammes (Sablé-Meyer, Fagot, et al., 2021). Ils possèdent également des intuitions sophistiquées sur le comportement des lignes parallèles en géométrie planaire et sphérique, comme l'unicité d'une ligne parallèle passant par un point donné du plan (Izard et al., 2011).

Preuves développementales

Un autre élément de preuve découle des données sur le développement. Il a été démontré que les enfants d'âge préscolaire et même les nourrissons possèdent des intuitions sophistiquées de l'espace (Hermer & Spelke, 1994; Landau et al., 1981; Newcombe et al., 2005), des séquences spatiales (Amalric et al., 2017), et de la symétrie des miroirs (Bornstein et al., 1978). En effet, les dessins des enfants d'âge préscolaire montrent déjà une tendance à représenter les propriétés abstraites des objets plutôt que l'objet lui-même. Bien qu'ils aient l'air primitifs, les dessins d'une maison sous la forme d'un triangle au-dessus d'un carré, ou d'une personne sous la forme d'un bâton avec une tête ronde, suggèrent une capacité remarquable à faire abstraction de la forme réelle et à s'intéresser à ses axes principaux, au détriment du réalisme. De nombreux tests exploitent cette compétence géométrique pour évaluer le développement cognitif d'un enfant en comptant le nombre de propriétés abstraites correctes ou incorrectes, par exemple lorsqu'on lui demande de dessiner une personne (Goodenough, 1926; Harris, 1963; Long et al., 2019; Prewett et al., 1988; Reynolds & Hickman, 2004). Il existe certaines preuves, bien que limitées, que cette capacité pourrait être spécifiquement humaine : lorsqu'on leur donne des crayons ou une tablette, d'autres primates non humains ne dessinent aucune forme abstraite ou figure reconnaissable, mais génèrent surtout des gribouillages informes (Saito et al., 2014; Tanaka et al., 2003).

Les programmes comme candidats aux représentations mentales : Une prise sur l'hypothèse du langage de la pensée

> "Les humains ont une capacité et une propension multi-domaines à déduire des structures arborescentes à partir de chaînes de caractères, à un degré qui est difficile ou impossible pour la plupart des espèces animales non humaines"
>
> W. Tecumseh Fitch(Fitch, 2014)

Primitives géométriques de la cognition

En sciences cognitives, la première application de la théorie de l'information à la perception visuelle provient des travaux de At-

tneave (1954). Il a observé qu'une perception visuelle est souvent redondante : de nombreuses parties pourraient être cachées et pourtant reconstruite avec succès par un observateur humain. Il en déduit que la représentation mentale d'un percept visuel doit tirer parti de cette redondance pour compresser l'information visuelle, et à l'inverse en reconstruire et en inférer des parties manquantes. Pour les formes géométriques spécifiquement, la plus ancienne théorie de « programmes » représentant des formes géométriques vient de Leeuwenberg et ses collègues (E. L. Leeuwenberg, 1969, 1971; Boselie & Leeuwenberg, 1986), qui ont proposé un langage de programmation formel pour les formes bidimensionnelles et tridimensionnelles. Ils défendent l'idée que la représentation mentale internet d'une forme est d'autant plus complexe que le plus petit programme qui génère cette forme est long, de ce langage. C'est une propriété que je défends également au chapitre 5. En fait, Leeuwenberg observe déjà que certains éléments de leur langage sont très généraux, et qu'ils pourraient être appliqués à la compression de séquences auditives également, une propriété qui a été retrouvée dans des travaux récents sur les séquences géométriques dans le domaine visuel puis auditif (Amalric et al., 2017; Planton et al., 2021). En plus des avancées théoriques, ils fournissent un certain soutien empirique à leur proposition, mais concluent en disant que "[...] pour l'instant [la procédure de codage proposée] ne se prêtera guère à la programmation informatique". De nos jours, cette limitation ne s'applique plus, et j'aborde ce sujet au chapitre 5. Je compare cette approche à d'autres, notamment celle de Leyton (Leyton, 1984, 2003) dans le corps de la thèse.

Complexité de Kolmogorov et LMD

Selon une hypothèse cognitive de longue date, le cerveau excelle dans la compression des informations. En effet, en présence de structure dans les stimuli, qu'ils soient visuels, auditifs ou autre, le score de participants s'améliore dans une grande variété de tâches. La première observation de ce phénomène provient de Miller (1956), qui déclare "J'ai pris l'habitude de faire la distinction entre les unités d'information et les morceaux d'information. [...] L'étendue de la mémoire immédiate semble être presque indépendante du nombre d'unités par morceau, du moins sur la plage qui a été examinée jusqu'à présent." On trouve des exemples de ce phénomène dans la vie courante par exemple dans la mémorisation de mots, où le facteur principal est le nombre de mots et non le nombre de lettres, mais des observations similaires sont om-

263

niprésentes en psychologie.

Une version forte de cette hypothèse affirme que le cerveau trouve la structure d'une manière plus riche que le regroupement par morceaux, et s'appuie à la place sur des langages génératifs (de programmation) : la complexité d'un élément d'information donné est alors la longueur du programme le plus court qui génère cette information (Goodman et al., 2012; Tenenbaum et al., 2011). Cette hypothèse est liée à l'hypothèse du codage prédictif, car la capacité de prédire et de générer peut être liée aux langages génératifs. En outre, pour les langages de programmation probabilistes, les théories des langages génératifs ont la capacité de rendre compte à la fois des succès et des erreurs du comportement humain : une caractéristique cruciale d'une bonne théorie pour expliquer des phénomènes cognitifs. Qui plus est, de telles théories peuvent envisager la coexistence de plusieurs représentations mentales possibles avec des probabilités différentes, et donc rendre compte des représentations ambiguës ou sous-déterminées.

Le choix du langage de programmation a une grande importance, car différents choix peuvent faire des prédictions très différentes. Et il n'est pas évident de trouver une proposition unificatrice qui puisse rendre compte de stimuli de nature très différente (auditifs, visuels, physiques intuitifs, etc.). Heureusement, un domaine de recherche entier étudie un problème connexe : le domaine de la théorie de l'information. Ce domaine étudie des questions théoriques telles que "quelle est la manière la plus efficace en termes d'information pour représenter un ensemble donné et ses éléments". Dans l'introduction de ma thèse, je décris deux notions de ce domaine et je montre comment elles sont liées : La complexité de Kolmogorov, et la longueur minimale de description (LMD). Je discute de l'influence et de l'utilité qu'elles peuvent avoir lors de la spécification des théories cognitives.

Cognition et induction de programmes

Si les représentations mentales ont des propriétés qui les font ressembler à des programmes, il est alors crucial de proposer une théorie sur la manière dont ces programmes sont inférés : comment passe-t-on des entrées sensorielles à la représentation structurée ? Quels sont les bons modèles de "construction de programmes mentaux" ?

En informatique, le sous-domaine qui s'attaque à cette question est

celui de "l'induction de programme" : le problème de l'induction de programme est de trouver un programme, étant donné un ensemble de paires d'exemples d'entrée-sortie. Une observation immédiate est que c'est un problème sous-spécifié : pour une taille d'entrée finie, il existe un nombre infini de programmes qui peuvent fonctionner, mais ils se comporteront différemment sur une nouvelle entrée - en l'absence de quoi la notion de "programme correct" ne peut être décidée. Mais si l'on donne suffisamment d'exemples, la longueur du programme peut être utilisée comme stratégie de sélection dans un modèle probabiliste de génération de programmes. Je prétends que lorsqu'ils cherchent une forme géométrique, les humains cherchent, et trouvent, le programme le plus court, et donc que la complexité d'une forme sera prédite par sa LMD dans un langage bien choisi pour un domaine donné.

Une base de référence pour résoudre l'induction de programme est l'énumération de programmes : énumérer récursivement tous les programmes possibles dans la grammaire d'un langage de programmation jusqu'à trouver un programme qui satisfait les contraintes. Mais cette procédure est un piètre candidat pour une implémentation cognitive de l'induction de programme dans le cadre de l'hypothèse LMD : si nous considérons que la complexité d'un certain programme est fonction de sa LMD, alors dans le cadre de cette approche, la complexité de la *recherche de* ce programme croîtrait exponentiellement avec sa LMD. Cependant, plusieurs méthodes ont été conçues pour tenter de contenir cette explosion combinatoire, je les décris dans l'introduction de la thèse et je m'appuie sur DreamCoder (Ellis et al., 2021), que j'ai contribué à mettre en œuvre, au chapitre 5.

Structure des chapitres

Dans le chapitre 1, je montre que même dans leur façon d'effectuer une détection d'un intrus parmi des quadrilatères, les humains et les primates non humains diffèrent. En utilisant une tâche de détection d'intru qui utilise des quadrilatères de régularité différente, allant de carrés très réguliers à des quadrilatères irréguliers, je montre que des humains d'éducation, d'âge et de culture différentes partagent un sens commun de la complexité géométrique : certaines formes suscitent systématiquement de meilleures performances que d'autres. Par exemple, lorsqu'ils cherchent un intrus parmi des carrés, les participants sont rapides et précis, mais lorsqu'ils cherchent un intrus parmi des

Chapter 1: Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity

Task: pick the intruder

Populations:
- N≈800 Educated Adults
- N=44 Himbas
- N=28 Kindergartners
- N=156 First graders
- N≈26 Baboons

Several tasks tested in educated adults.

Models: Several Neural Networks, custom symbolic models

Modality: touchscreens, online or in the lab, measuring errors and Response Times

Result: strong dissociation between humans and Baboons

Chapter 2: M/EEG evidence for symbolic and non-symbolic neural mechanisms of geometric shape perception in adults and infants

Task: passive presentation with oddballs

Miniblock with 1st shape

SOA 1s, ISI 200ms

6 reference shapes   20 reference shapes, 4 oddballs

Modality and population

N=20 Adults, MEG & anatomical MRI

N=40 3-months-olds with EEG nets Still ongoing

Results: spatial & temporal dissociation of the models

Neural Network Model
Symbolic Model

Onset   400ms   time

Chapter 3: Geometric Shape Perception Activates Brain Regions Associated with Mathematical Cognition, an fMRI study

Task: Category Localizer with geometric shapes

BON
走
2 - 1

Modality and population

N=20 Adults 1h30 of fMRI

N=20 Six-years-old 1h00 of fMRI

Results: In both age groups, comparing single items to shapes reveals a lower activation in ventral pathways and a math-like activation in the IPS

Adults   6 y.o.

Contrast:

⬡ > (👤 i 🖼)

t   -10 0 10

Chapter 4: Categorical perception of right angles in humans and baboons

Task: delayed match-to-sample; **Method**: online with touchscreen

Population:
N≈153 educated adults
N=26 baboons

Stimuli:
Target angle in 0°, 30°, 60°, 90°, 120°, 150°, 180°
Distractors: random with ±10°, 20°, 30°, 40°, 50°

Results: categorical perception of right angle only in a subset of conditions in humans, no conclusion yet in baboons

P(chosen|presented)

Chapter 5: A Language of Thought for the Mental Representation of Geometric Shapes

Core proposition: shape perception is performing program induction, Thus shape complexity increases With Minimal Description Length

Tasks: delayed match to sample with shapes of ranging MDL

repeat
3   concat
draw   draw

Modeling: DreamCoder, State of the Art Program induction model

Modality & population: touchscreens, online

Exp. 1: N=57 humans, N=26 baboons
Exp. 2: N=170 humans

Results: shape complexity correlates with MDL Concatenation, repetition and embedding are pervasive in geometric shape perception

Residuals on all other visual predictors

Choice Time
R² = 0.63
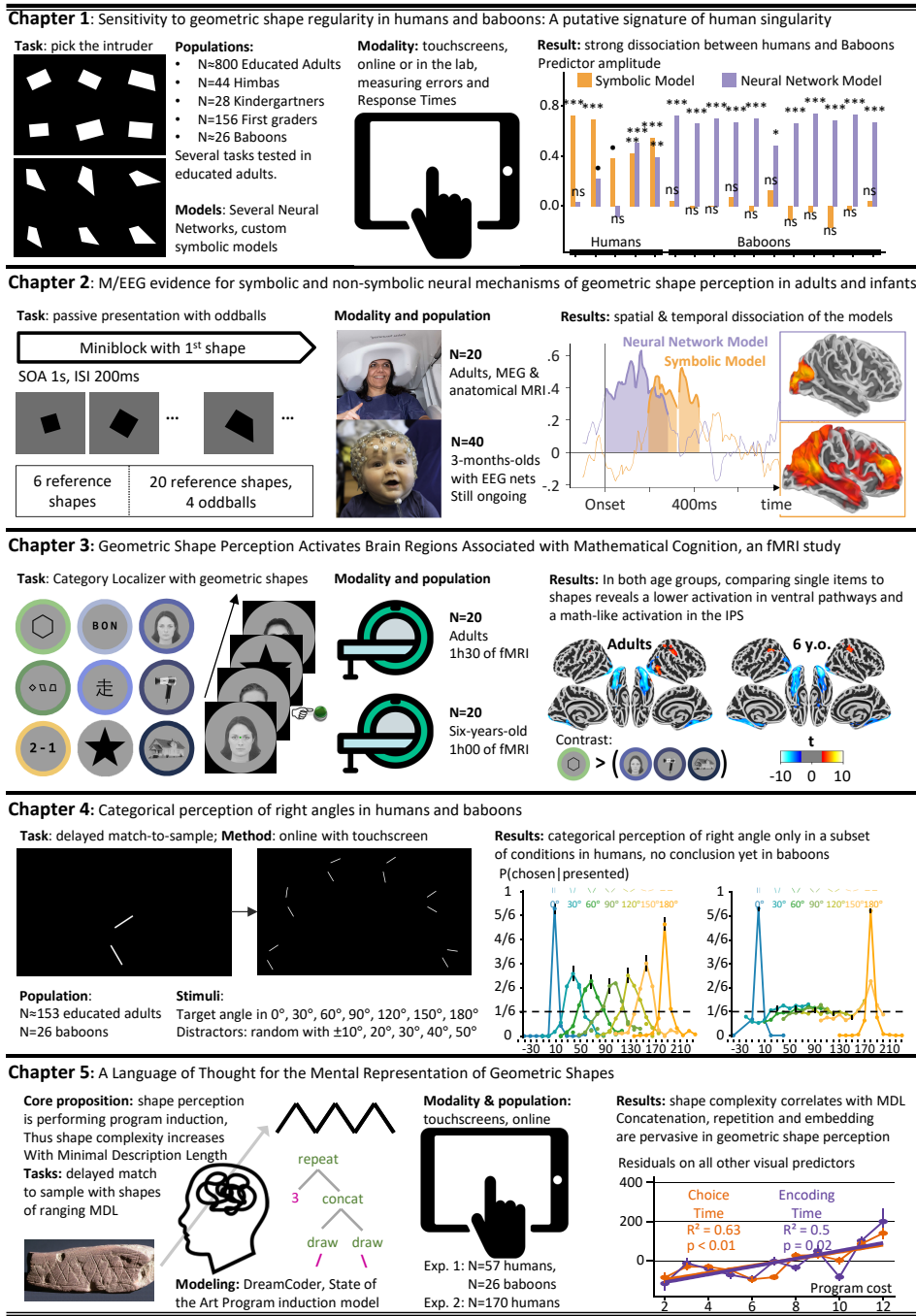p < 0.01

Encoding Time
R² = 0.5
p = 0.02

Program cost

Figure 16

formes irrégulières identiques, les participants sont lents et peu précis. Je montre que cette observation reste vraie avec différents paradigmes tels que la recherche visuelle ou un classement introspectif des formes, suggérant un effet fort et facile à reproduire.

En revanche, les babouins n'ont pas cette sensibilité aux formes, même après un entraînement intensif adéquat : sur une séquence de tâches de détection d'intru de plus en plus difficiles, je confirme que les babouins peuvent comprendre la tâche et généraliser entre les stimuli, mais qu'ils ne parviennent pas à généraliser aux formes quadrilatérales. Même avec un entraînement intensif sur les formes quadrilatérales, leur performance moyenne globale finit par atteindre le niveau d'enfants de 5 ans non entraînés, mais les babouins ne font pas preuve de l'effet de régularité géométrique détecté chez les humains : toutes les formes suscitent un comportement similaire.

De nombreux réseaux de neurones artificiels de reconnaissance d'objets, d'architecture différente, ont un comportement comparable à celui des babouins, mais expliquer le comportement des humains nécessite d'utiliser des propriétés symboliques supplémentaires telles que la présence de parallèles ou d'angles droits (je m'attarde sur cette propriété symbolique spécifique au chapitre 4). Les modèles symboliques du comportement des humains prédisent plusieurs tâches connexes, ce qui valide la robustesse du choix des primitives géométriques incluses dans le modèle. Cette forte dissociation entre les humains et les babouins suggère que deux stratégies sont disponibles pour encoder les formes géométriques : les humains et les primates non humains partagent une stratégie perceptive, bien capturée par les modèles de la voie visuelle ventrale, mais seuls les humains ont accès à une stratégie exacte, symbolique. Cela suggère une possible signature de la singularité humaine dans une tâche visuelle très simple, et constitue un défi pour les modèles non symboliques de la perception des formes humaines à l'avenir.

Ce chapitre correspond à un article publié dans PNAS sous la référence suivante : **Sablé-Meyer, M.**, Fagot, J., Caparos, S., Kerkoerle, T. van, 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).

Dans le chapitre 2, j'utilise des techniques de neuro-imagerie pour mettre en lumière l'implémentation neuronale des deux stratégies pro-

posées au chapitre 1.

Tout d'abord, j'utilise une nouvelle tâche comportementale de recherche visuelle pour mesurer la matrice de confusion entre les quadrilatères conçus au chapitre 1, et je la modélise à nouveau avec le modèle symbolique et le réseau neuronal. Je montre également qu'une décomposition de la matrice de complexité basée sur les données coïncide avec le modèle symbolique déjà proposé. Cela me fournit à la fois une réplication des résultats du chapitre 1, une façon guidée par les données de voir ce qui domine l'effet de régularité géométrique, et une matrice de confusion que je pourrai utiliser à la fois dans le chapitre 2 et dans le chapitre 3 pour modéliser les données de neuro-imagerie.

Ensuite, pour rendre une tâche compatible avec des adultes et des nourrissons, je conçois un paradigme d'« oddball » purement passif : les participants voient les quadrilatères du chapitre 1 centrées sur un écran avec une taille et une rotation aléatoires, une fois par seconde, avec de un déviant de temps en temps, à des moments tirés au hasard. Chez vingt adultes testés pendant 90 minutes en MEG, je peux décoder le signal cérébral associé aux déviants, et la performance du décodage reproduit l'effet de régularité géométrique. En utilisant une analyse en similarité représentationnelle (RSA), je montre que le signal cérébral ressemble d'abord au modèle du réseau neuronal, puis au modèle symbolique, ce qui indique que les deux modèles existent chez les adultes malgré le fait que leur comportement ne reflète que le modèle symbolique. En utilisant de la reconstruction de sources, je montre que le modèle de réseau neuronal correspond à un groupe de sources occipital bilatéral, tandis que le modèle symbolique est associé à un large groupe de sources, qui inclut des sources frontales et des sources dans la voie dorsale.

Cela permet de reproduire les données comportementales avec un paradigme de présentation passive, dans lequel les intrus sont d'autant plus détectables qu'il s'agit d'intrus de formes régulières. De plus, cela confirme la proposition du chapitre 1 selon laquelle les deux stratégies sont présentes chez les adultes éduqués, un résultat invisible à partir du comportement seul mais nécessaire pour comprendre les données du premier chapitre. Enfin, cela fournit une dissociation spatiale et temporelle de la voie visuelle ventrale « bottom-up » du traitement des formes, précoce et transitoire, située dans la zone occipitale, suivie d'un traitement symbolique « top-down », s'étendant

sur une zone cérébrale beaucoup plus large.

Dans deux groupes de 20 nourrissons de trois mois, j'essaie de reproduire cette expérience en mesurant l'activité cérébrale mesurée en EEG. Les résultats préliminaires ne sont pas concluants quant à la détection des intrus, mais il semble que les différentes formes soient représentées différemment les unes des autres d'une manière qui est compatible avec l'effet de régularité géométrique déjà observé chez les nourrissons. Les résultats obtenus chez les nourrissons sont préliminaires et je travaille encore sur ce projet.

Dans le chapitre 3, j'utilise l'IRMf 3T chez vingt adultes et vingt enfants de 6 ans pour localiser plus précisément les zones associées à la perception des formes géométriques à travers l'éducation.

À l'aide d'un localisateur de catégories visuelles, avec des catégories visuelles usuelles (visages, objets, etc.) ainsi que des mots, des nombres et des formes géométriques, j'ai pu constater que la perception de formes géométriques, par opposition aux autres objets visuels, sur-active un réseau cortical dans le sulcus intrapariétal, de façon bilatérale, dans des régions typiquement associées au traitement de nombres et d'énoncés mathématique. À l'inverse, les formes géométriques sous-activent les voies visuelles ventrales, lorsqu'elles sont comparées aux autres catégories visuelles telles que les visages, les objets et les maisons. Ce résultat est valable pour les deux groupes d'âge, ce qui indique que les mécanismes fondamentaux en jeu dans les formes géométriques sont déjà présents chez les enfants de six ans. Une exploration plus approfondie du groupe ventral sous-activé révèle que même si, intuitivement, on pourrait penser que le traitement des formes ressemble à la lecture des lettres, au traitement visuel des visages, au traitement visuel des objets ou à d'autres traitements visuels généraux, aucune des zones typiquement associées à ces catégories n'est fortement activée par les formes géométriques.

À l'intérieur de l'IRMf, les participants ont aussi effectué une variante du test de l'intrus du chapitre 1 où ils devaient indiquer l'emplacement d'une forme déviante sur l'écran. Les deux groupes d'âge ont effectué une version simple de la tâche, et les adultes ont en plus effectué une version plus difficile de la tâche. Le comportement à l'intérieur du scanner reproduit très bien les expériences précédentes, y compris un classement général des performances entre les groupes d'âge et la difficulté, ce qui indique que même avec un temps limité, l'effet de régu-

larité géométrique est présent.

Les données provenant d'adultes dans la condition facile ont donné lieu à de larges clusters bilatéraux qui montrent une activité accrue en fonction de la complexité de la forme. En plus de cela, une analyse en similarité représentationnelle a détecté des clusters significatifs associés à la fois au modèle de réseau de neurones, principalement à l'intérieur des aires visuelles, et au modèle symbolique, principalement autour de l'IPS. Les deux modèles sont ceux définis dans le chapitre 1 et le chapitre 2. Il y avait également quelques zones associées aux deux modèles, principalement dans les zones de traitement de la décision, ce qui suggère un processus d'intégration pondérant les deux sources d'information, peut-être lorsque les participants doivent répondre.

Dans le chapitre 4, je me concentre sur une seule propriété géométrique : les angles droits. Je compare le comportement d'adultes éduqués et de babouins dans une tâche d'appariement différé impliquant différents angles et leurs distracteurs relatifs. La question est de savoir si des angles spécifiques, à savoir 0°, 90° et 180° (correspondant à des lignes alignées, des lignes perpendiculaires et des lignes parallèles) suscitent un comportement catégoriquement différent des autres angles, comme l'indique principalement une sensibilité beaucoup plus forte à la déviation de l'angle de référence. La tâche teste des angles également espacés : 0°, 30°, 60°, 90°, 120°, 150° et 180°. À chaque essai, des distracteurs sont échantillonnés à partir d'angles voisins distants de 10°, 20°, 30°, 40°, 50° de chaque côté. Ensuite, la comparaison des performances entre un angle et ses voisins (par exemple, 90° par rapport à 80° et 100°, par rapport à 30° par rapport à 20° et 40°) nous informe sur la sensibilité des participants à des angles spécifiques.

En essayant de reproduire des résultats classiques sur la perception catégorielle des angles droits chez les adultes, j'ai mis en avant le fait que plusieurs propriétés sont nécessaires pour que les angles droits suscitent un comportement différent des autres angles voisins : (i) les stimuli doivent être affichés suffisamment longtemps, et (ii) aucune autre propriété de bas niveau ne peut être utilisée pour réaliser la tâche. Ces deux résultats coïncident avec l'idée que les propriétés symboliques requièrent une certaine attention de la part des participants, comme énoncé au chapitre 1, mais il s'agit de nouveaux résul-

tats qui n'étaient pas évidents au vu de la littérature sur la perception catégorielle des angles droits.

Je présente également les premières données recueillies chez les babouins où les performances sont très bonnes pour 0° et 180°, et médiocres partout ailleurs. Malheureusement, les données collectées jusqu'à présent ne correspondent qu'à un sous-ensemble de conditions où les humains ne présentent pas non plus de perception catégorielle des formes géométriques, et donc son absence chez les babouins n'est pas très informative sur les théories de la perception des angles chez les humains et les babouins.

Dans le chapitre 5, Je vais au-delà du petit ensemble de quadrilatères hautement contrôlés utilisé précédemment, et j'essaye de rendre compte de *toutes les formes géométriques* produites par les humains. Pour ce faire, je fais une proposition concrète pour un langage mental génératif de formes géométriques inspiré par des productions géométriques humaines attestées. Ma proposition satisfait deux contraintes : les programmes courts doivent générer des formes simples (pour les humains), et les formes simples doivent avoir des programmes courts.

C'est dans ce chapitre que je développe pleinement l'argument selon lequel la perception d'une forme donne lieu à de l'induction de programme : pour une forme géométrique donnée, trouver le programme mental le plus court dans le langage interne qui génère cette forme. Tout d'abord, pour le langage que je propose, je montre que l'induction de programme est en principe un problème qui peut être résolu : à cette fin, j'utilise l'algorithme DreamCoder que j'ai aidé à développer avant mon doctorat. Je montre qu'il peut trouver avec succès le programme le plus court pour une forme donnée. Ensuite, j'utilise ce langage pour générer des formes de plus en plus complexes, et je montre que la performance des humains dans une tâche de correspondance avec l'échantillon pour une forme est corrélée avec la longueur de son programme le plus court, au-delà de nombreuses autres caractéristiques perceptives qui sont par ailleurs typiquement attestées dans la littérature.

En plus de cela, je fais remarquer que DreamCoder met en œuvre une théorie plausible de la façon dont différentes cultures se mettent à produire des dessins qui ont l'air très différents, par exemple les productions curvilignes celtiques et les productions rectilignes

grecques. Bien que DreamCoder dispose au départ d'un langage unique pour la géométrie, plus il est entraîné, plus il apprend à la fois de nouvelles primitives et de nouveaux biais. Ce faisant, DreamCoder dérive lentement dans des directions spécifiques tout au long de sa procédure d'entraînement. Il pourrait en être de même pour les dérives culturelles, où les formes de base sont partagées par tous (cercles et carrés par exemple), mais les formes complexes réutilisent des blocs de construction intermédiaires qui sont appris à partir des primitives de base, et au fil du temps cela constitue une dérive visible dans les productions des différentes cultures.

Enfin, pour découpler mon résultat de ma proposition concrète de langage, je dérive également des règles additives générales auxquelles tout langage alternatif se doit d'obéir. Je conçois une nouvelle expérience avec un nouvel ensemble de formes pour tester cette hypothèse, et confirmer que la concaténation, la répétition et l'emboîtement sont essentiels pour capturer la nature compositionnelle de la complexité des formes géométriques. Pour rendre cela plus concret, je montre que la représentation mentale d'une forme qui est « un carré composé de carrés plus petits » induit une optimisation, qui vient du fait que les formes locales et globales sont identiques. Ce qui indique la nécessité de la récursion dans la perception des formes géométriques par les humains.

Ce chapitre correspond à un article publié dans Cognitive Psychology sous la référence suivante : **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.

Conclusion

Le langage naturel n'est pas le seul signe distinctif des capacités cognitives singulières des humains. Dans le travail présenté ici, et dans la lignée de la littérature sur le langage de la pensée, je défends l'idée qu'il pourrait exister plusieurs "langages de la pensée" internes, et que les formes géométriques constituent un levier très efficace pour étudier des processus cognitifs non linguistiques structurés. Pour ce faire, je montre que la cognition impliquant des formes géométriques nécessite un ensemble de représentations mentales discrètes et symboliques qui agissent comme un langage mental interne, doté de propriétés de composition. Grâce à la méthodologie de la psychologie

culturelle et de la psychologie du développement, je démontre que cette capacité ne dépend pas de l'éducation, de l'âge ou de la culture. À l'aide de méthodes de cognition comparative, j'apporte la preuve que les primates non humains ne peuvent accéder à ces représentations symboliques. Partant de ce constat, j'affirme que la perception d'une forme est comparable au processus d'induction de programme : trouver la représentation la plus courte de la forme dans le langage mental interne. Je présente des preuves neuronales de ce phénomène chez l'humain, en utilisant à la fois la MEG et l'IRMf, et je modélise les données avec des modèles inspirés de plusieurs domaines : les réseaux neuronaux issus de la littérature sur l'intelligence artificielle, les modèles neurosymboliques, et les modèles symboliques discrets couramment utilisés en psychologie.

Other Publications

As mentioned in the introduction, in addition to the work presented in this manuscript, two additional first-author articles were published, one from continued collaboration after an internship supervised by Pr. Salvador Mascarenhas at ENS and the other one in a collaboration with Lorenzo Ciccione, a friend and colleague. I provide the abstract of these two articles hereafter and will happily share the full articles upon request, but decided against including them in the manuscript because they are not central to the overall argument presented here, and to save paper when this manuscript is printed.

In addition to these two articles, I was a contributor (but not main author) of two additional articles during that time period. The abstracts are provided below.

First Author Publications

Indirect Illusory Inferences from Disjunction: A New Bridge Between Deductive Inference and Representativeness

Published in Review of Philosophy and Psychology in 2021, authored by myself and Pr. Salvador Mascarenhas.

Abstract. We provide a new link between deductive and probabilistic reasoning fallacies. Illusory inferences from disjunction are a broad

class of deductive fallacies traditionally explained by recourse to a matching procedure that looks for content overlap between premises. In two behavioral experiments, we show that this phenomenon is instead sensitive to real-world causal dependencies and not to exact content overlap. A group of participants rated the strength of the causal dependence between pairs of sentences. This measure is a near perfect predictor of fallacious reasoning by an independent group of participants in illusory inference tasks with the same materials. In light of these results, we argue that all extant accounts of these deductive fallacies require non-trivial adjustments. Crucially, these novel indirect illusory inferences from disjunction bear a structural similarity to seemingly unrelated probabilistic reasoning problems, in particular the conjunction fallacy from the heuristics and biases literature. This structural connection was entirely obscure in previous work on these deductive problems, due to the theoretical and empirical focus on content overlap. We argue that this structural parallelism provides arguments against the need for rich descriptions and individuating information in the conjunction fallacy, and we outline a unified theory of deductive illusory inferences from disjunction and the conjunction fallacy, in terms of Bayesian confirmation theory.

Analyzing the Misperception of Exponential Growth in Graphs

Published in Cognition in 2022, authored by Lorenzo Ciccione and myself (shared 1st authorship) and Stanislas Dehaene.

Abstract. Exponential growth is frequently underestimated, an error that can have a heavy social cost in the context of epidemics. To clarify its origins, we measured the human capacity (N = 521) to extrapolate linear and exponential trends in scatterplots. Four factors were manipulated: the function underlying the data (linear or exponential), the response modality (pointing or venturing a number), the scale on the y axis (linear or logarithmic), and the amount of noise in the data. While linear extrapolation was precise and largely unbiased, we observed a consistent underestimation of noisy exponential growth, present for both pointing and numerical responses. A biased ideal-observer model could explain these data as an occasional misperception of noisy exponential graphs as quadratic curves. Importantly, this underestimation bias was mitigated by participants' math knowledge, by using a logarithmic scale, and by

presenting a noiseless exponential curve rather than a noisy data plot, thus suggesting concrete avenues for interventions.

Contributing Author Publications

Symbols and Mental Programs: A Hypothesis about Human Singularity

Published in Trends In Cognitive Science in 2022, authored by Stanislas Dehaene, Fosca Al Roumi, Yair Lakretz, Samuel Planton myself.

Abstract. Natural language is often seen as the single factor that explains the cognitive singularity of the human species. Instead, we propose that humans possess multiple internal languages of thought, akin to computer languages, which encode and compress structures in various domains (mathematics, music, shape…). These languages rely on cortical circuits distinct from classical language areas. Each is characterized by: (i) the discretization of a domain using a small set of symbols, and (ii) their recursive composition into mental programs that encode nested repetitions with variations. In various tasks of elementary shape or sequence perception, minimum description length in the proposed languages captures human behavior and brain activity, whereas non-human primate data are captured by simpler nonsymbolic models. Our research argues in favor of discrete symbolic models of human thought.

DreamCoder: Bootstrapping Inductive Program Synthesis with Wake-Sleep Library Learning

Published in Proceedings of the International Conference on Programming Language Design and Implementation (PLDI), authored by Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sablé-Meyer, Lucas Morales, Luke Hewitt, Luc Cary, Armando Solar-Lezama, Joshua B. Tenenbaum.

Abstract. We present a system for inductive program synthesis called DreamCoder, which inputs a corpus of synthesis problems each specified by one or a few examples, and automatically derives a library of program components and a neural search policy that can be used to efficiently solve other similar synthesis problems. The library and search policy bootstrap each other iteratively through a variant of

"wake-sleep" approximate Bayesian learning. A new refactoring algorithm based on E-graph matching identifies common sub-components across synthesized programs, building a progressively deepening library of abstractions capturing the structure of the input domain. We evaluate on eight domains including classic program synthesis areas and AI tasks such as planning, inverse graphics, and equation discovery. We show that jointly learning the library and neural search policy leads to solving more problems, and solving them more quickly.

RÉSUMÉ

Le langage naturel n'est pas la seule capacité cognitive qui distingue les humains. Dans cette thèse, je défends l'idée que la cognition humaine des formes géométriques passe par un langage mental. Dans une tâche de détection d'intrus, le comportement d'humains est homogène et se distingue de celui des babouins dans son utilisation de propriétés symboliques comme la présence d'angles droits. Grâce à une dissociation visuelle/symbolique, je rends compte de cette différence, et je modélise les processus neuronaux de la perception de formes obtenus en imagerie cérébrale chez l'humain. Je fournis aussi des preuves préliminaires de l'existence de la stratégie symbolique chez le nourrisson. Enfin, je propose une version explicite de langage mental de la géométrie.

MOTS CLÉS

Psychologie Cognitive Expérimentale ; Neurosciences Computationnelles ; Formes Géométriques ; Langage de la Pensée ; Abstraction in Cognition ; Structures in Cognition

ABSTRACT

Natural language is not the only hallmark of humans' singular cognitive abilities: I propose that cognition involving geometric shapes requires a set of discrete, symbolic mental representations that act as a mental language. First, I show that all humans share a sense of geometric complexity that baboons lack even after adequate training. Artificial neural networks fit baboons' data, but explaining humans' behavior requires using symbolic properties such as the presence of right angles. Then, I identify the neural dynamics of both a visual and a symbolic strategy of shape perception using brain imaging methods, and I provide preliminary evidence for the existence of the symbolic strategy in infants. Finally, I propose and test an explicit mental language of geometry.

KEYWORDS

Experimental Cognitive Psychology; Computational Neurosciences; Geometric Shapes; Language of Thought; Abstraction in Cognition; Structure in Cognition