Cours 2023-2024:
La perception des objets mathématiques élémentaires: Formes géométriques, motifs et graphiques
Perception of elementary mathematical objects: Geometric shapes, patterns, and graphics

Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

Cours n°6

Modèles de la perception des formes géométriques

Models for the perception of geometric shapes

Modeling geometrical shapes

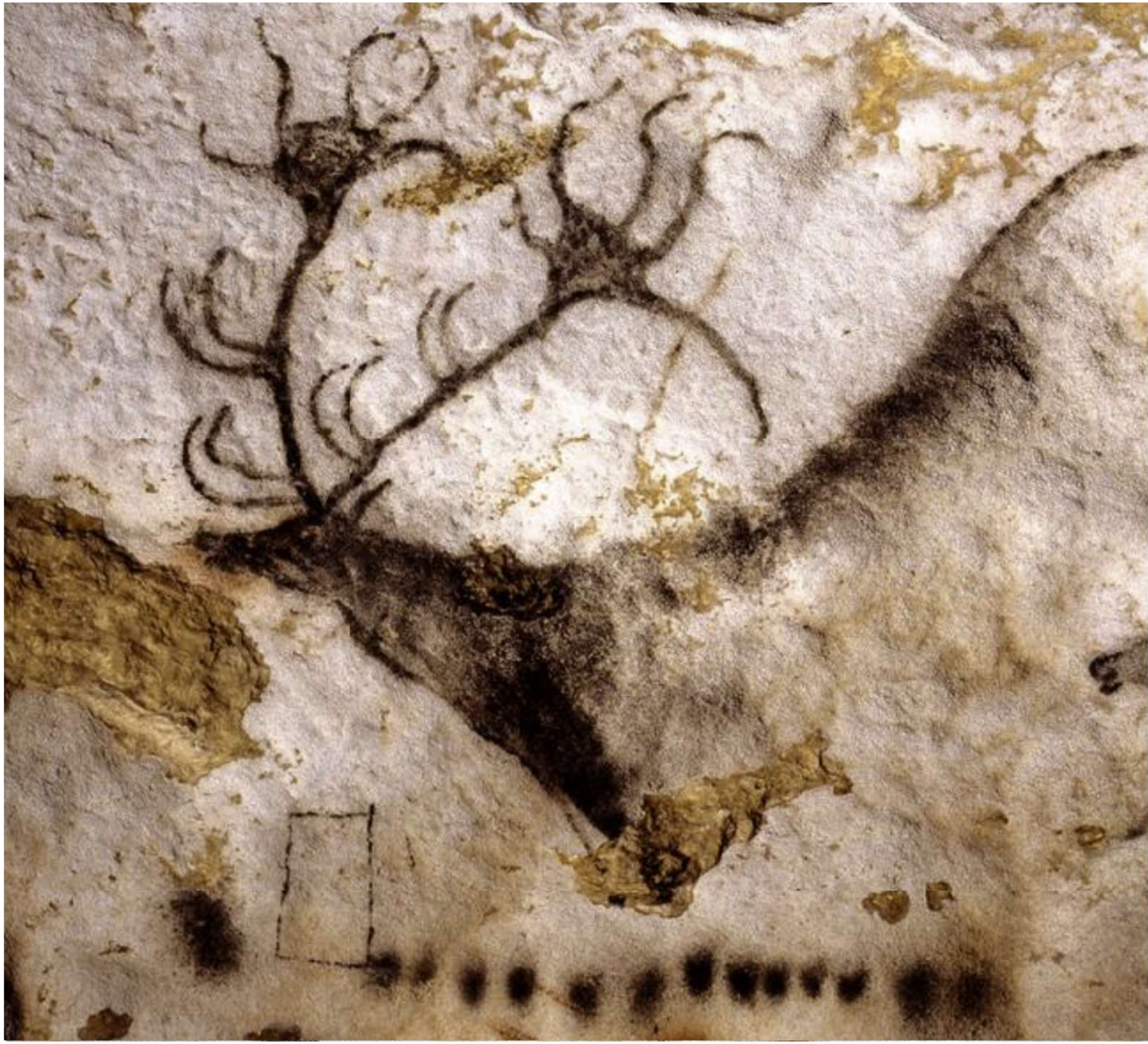
Part I. Modeling shapes with a language of thought.

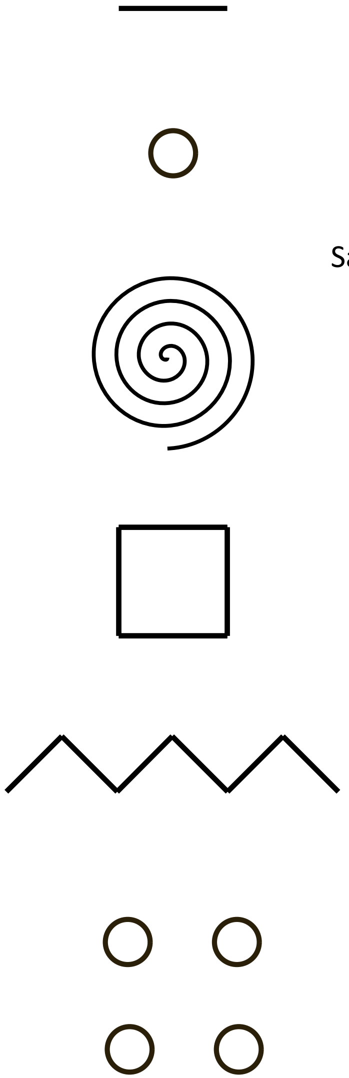
We will propose a **simple language** such that sequences and shapes with a lower **minimal description length** are precisely those that are universal across cultures, are judged as simpler, and are easier to recognize or to memorize.

Part II. Some challenges

- Can **neural networks** implement the language of thought? How?

- The importance of **principal axes** and the alternative theory of **medial axis coding**.





Can a "language of thought" account for all cross-culturally attested geometrical shapes?

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

Goal: propose a programming language that can account for the basic geometrical shapes used in human cultures throughout the world.
Test it as a candidate Language of Thought for geometrical shapes

The language contains a few key primitives:

- ***Number:*** `1, successor, fraction`
- ***Geometry:*** `Move, Turn Trace`
- ***Control:*** `Repeat, Concatenate, Subprogram`

For instance, a square is:

```
Repeat (4)
   { Concatenate ( Trace() , Turn() }
```

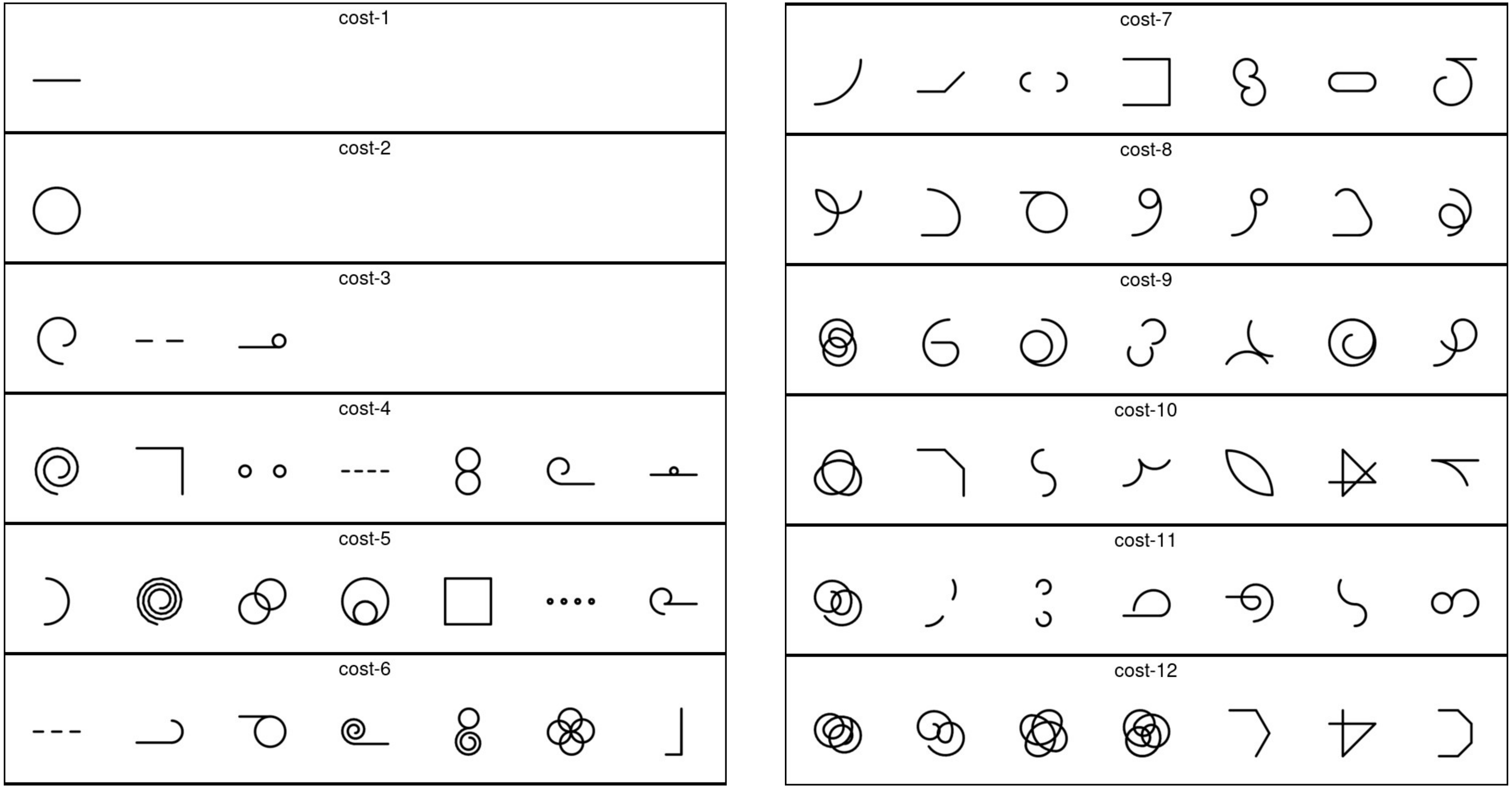
The full instructions set in our "language of geometry"

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

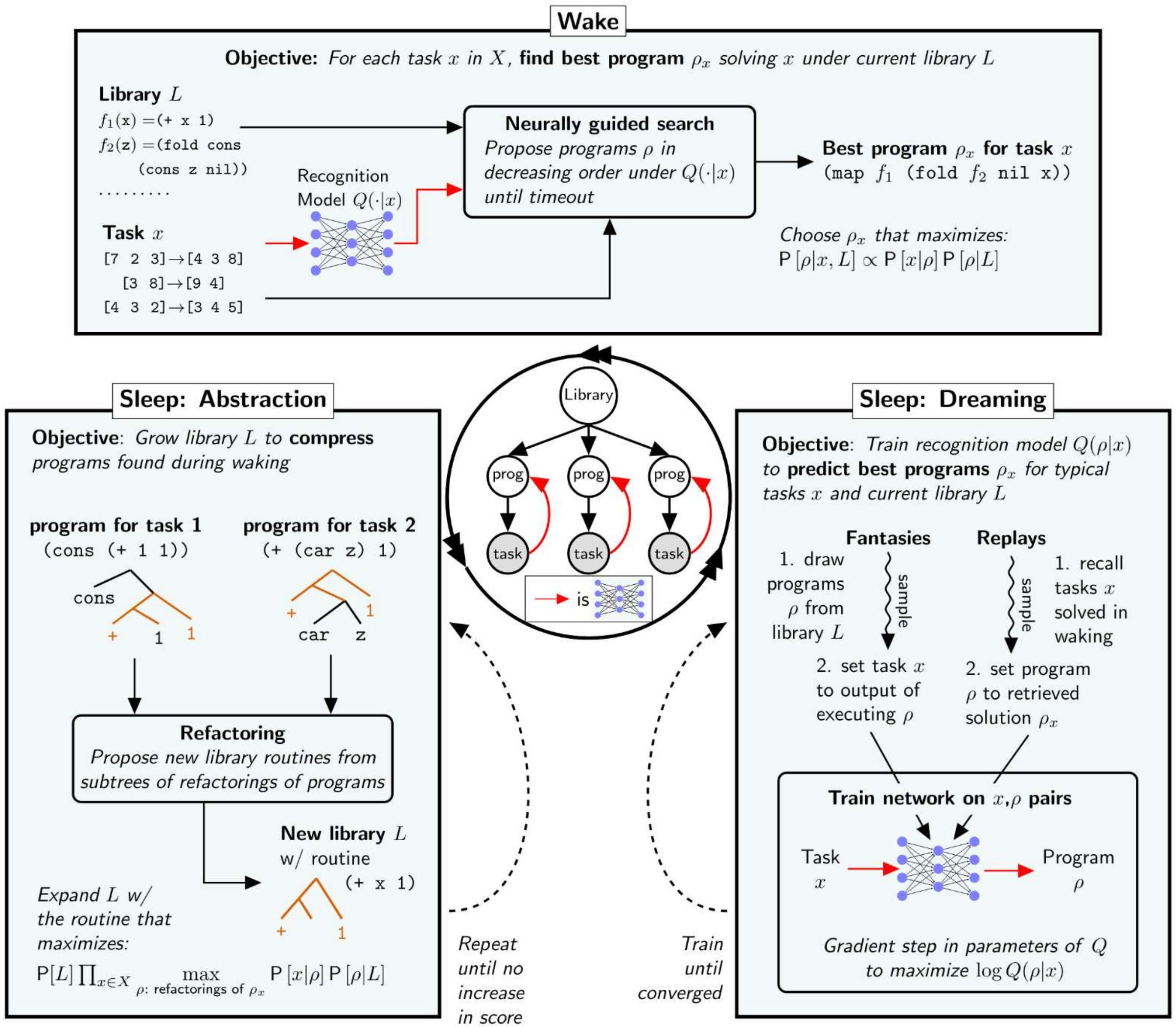
<table>
<tr><td>Program :=</td><td></td><td></td></tr>
<tr><td>| Program ; Program</td><td>Concatenate : run one program and then another</td><td rowspan="3">control</td></tr>
<tr><td>| Repeat([Int=2]) { Program }</td><td>Repeat a program a certain number of times</td></tr>
<tr><td>| Subprogram { Program }</td><td>Execute a program, then restore the original state</td></tr>
<tr><td>| Trace([t=Int=1],
[speed=Num=+1],
[acceleration=Num=+0],
[turningSpeed=Num=+0])</td><td>Trace a curve by moving according to the parameters</td><td rowspan="3">drawing</td></tr>
<tr><td>| Move([t=Num=+1])</td><td>Move a certain distance without tracing anything</td></tr>
<tr><td>| Turn(angle=Num)</td><td>Rotate the current heading</td></tr>
<tr><td>Int :=</td><td></td><td rowspan="6">arithmetic</td></tr>
<tr><td>| one</td><td>Number 1</td></tr>
<tr><td>| Next(Int)</td><td>Successor function</td></tr>
<tr><td>Num :=</td><td></td></tr>
<tr><td>| +Int | -Int</td><td>Return a signed number</td></tr>
<tr><td>| +Int/Int | -Int/Int</td><td>Return the signed fraction of two integers</td></tr>
</table>

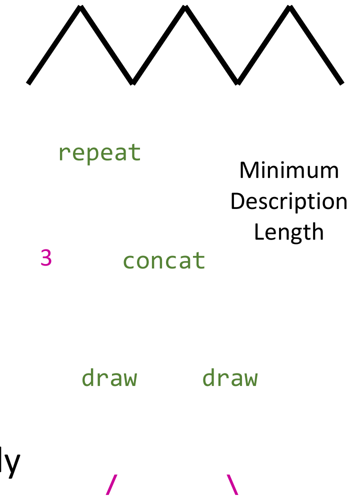
Minimal description length explains which shapes are simple and universal

The language generates all common shapes (e.g. square, spiral) with very short programs.
As programs get longer, more sophisticated shapes are produced.



Shape perception as program induction: A proof of concept





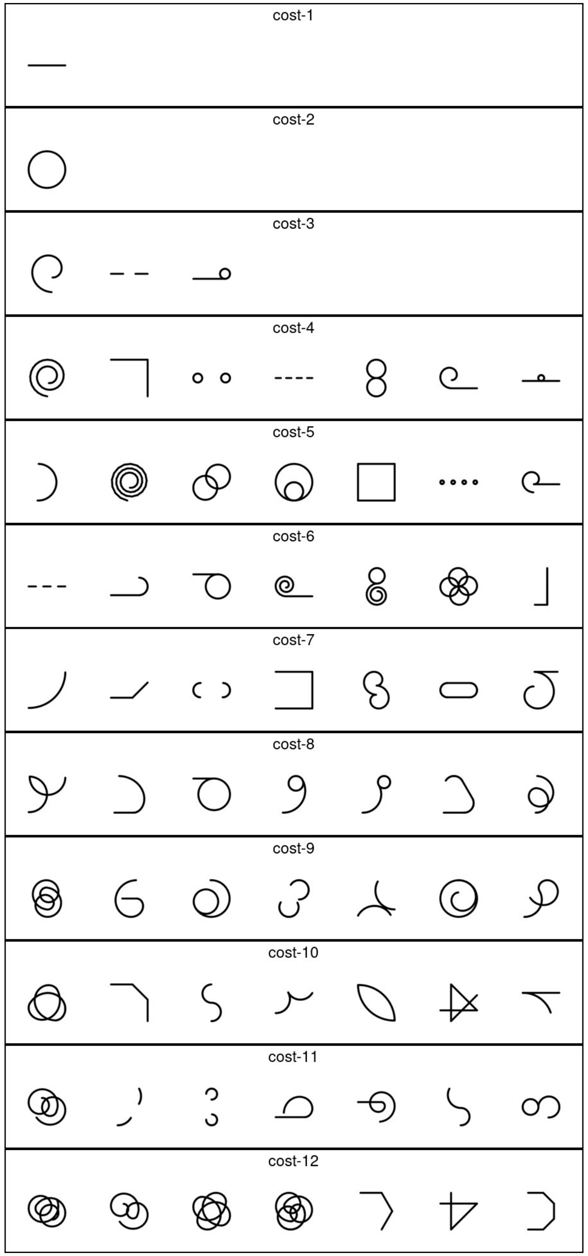
Program inference is, in general, a formidable task because the space of programs is vast and cannot be efficiently searched e.g. by gradient descent.

In DreamCoder, three tricks are used:

- programs generate shapes in a top-down manner, but a neural network makes suggestions in a bottom-up manner.
- a « dream » stage is used to train this neural network, both with (program, picture) pairs drawn either from random programs or from past pictures and their solutions.
- another « sleep » stage discovers a libary of subprograms that can efficiently compress previously found programs, and therefore reduce the search space.

Ellis, K., Wong, C., Nye, M., Sable-Meyer, M., Cary, L., Morales, L., ... & Tenenbaum, J. B. (2020). Dreamcoder: Growing generalizable, interpretable knowledge with wake-sleep bayesian program learning. arXiv preprint arXiv:2006.08381.

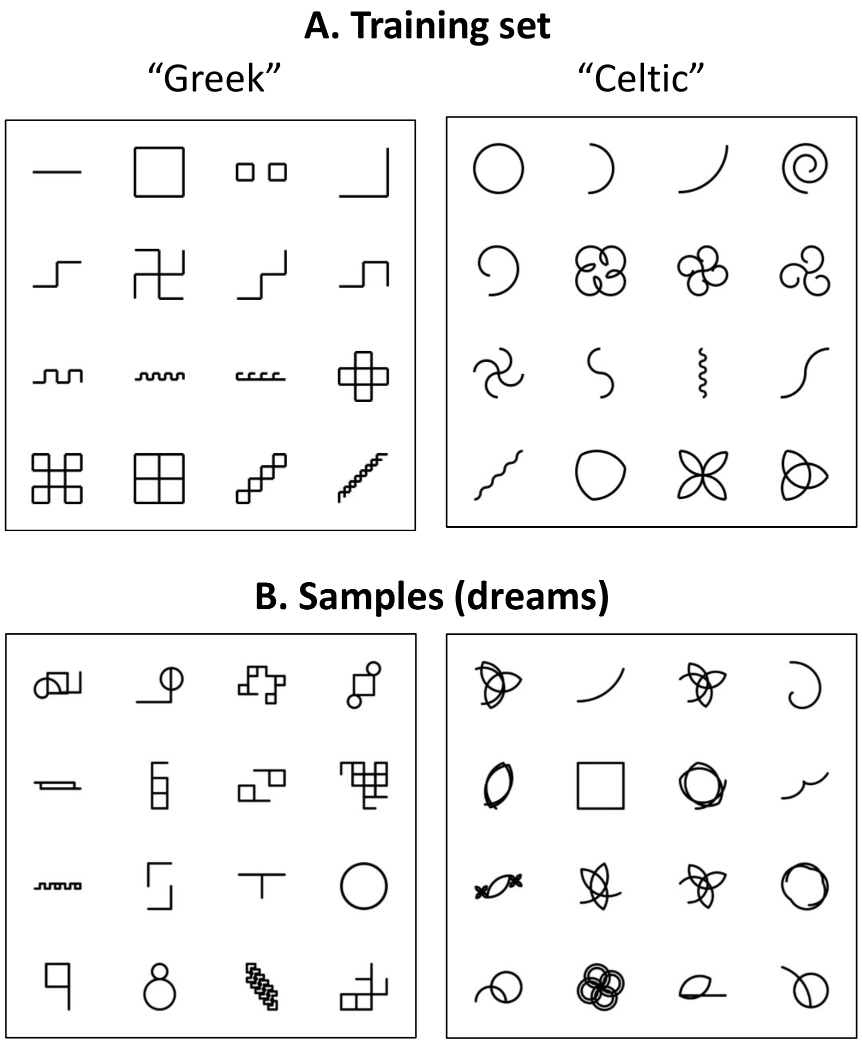
Explaining cultural universals as well as diversity



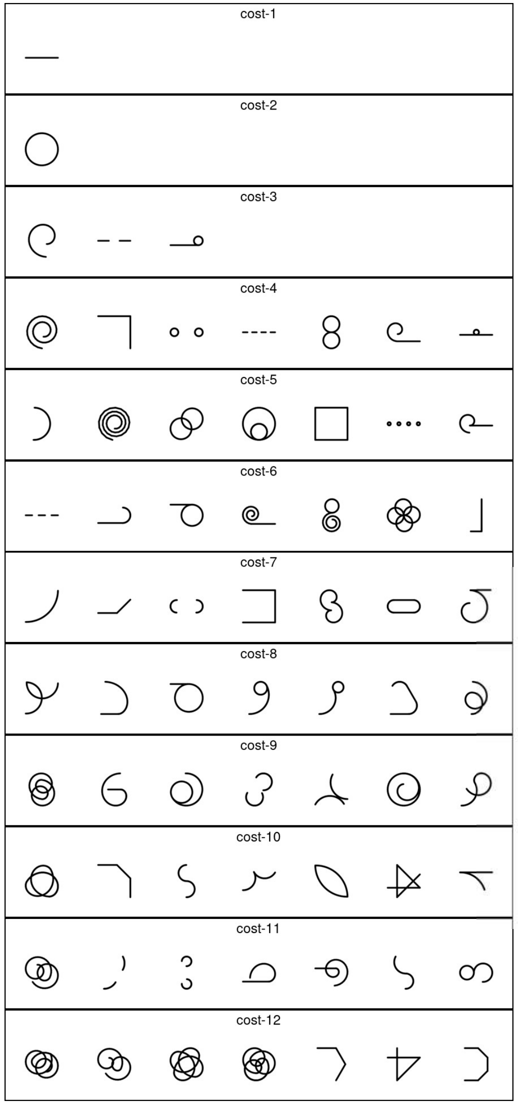
The language generates all common shapes (e.g. square, spiral) with very short programs.

As programs get longer, more sophisticated shapes are produced.

The language can adapt to a specific “culture”, e.g. the grammar can be biased to produce more right-angles or, on the contrary, more curves.



Testing the specific language proposed

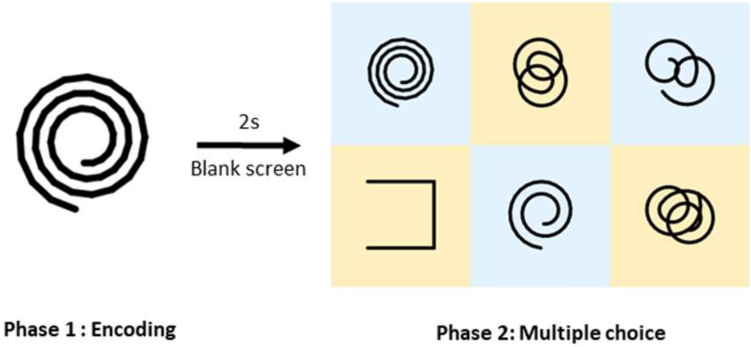


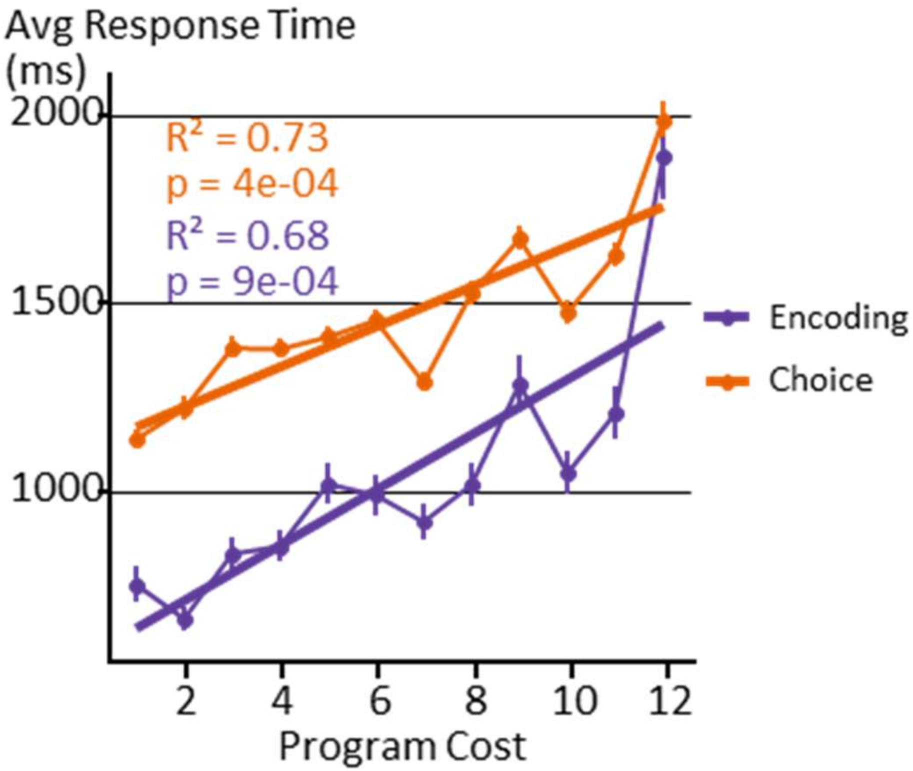
Sablé-Meyer, Ellis, Tenenbaum & Dehaene. A language of thought for the mental representation of geometric shapes. *Cognitive Psychology* (2022)

Experiment 1. Does program length predict psychological complexity?

Task =

1. Encode a shape in memory
2. After a delay, select it amongst distractors, chosen to be similar in both gray level and IT encoding.



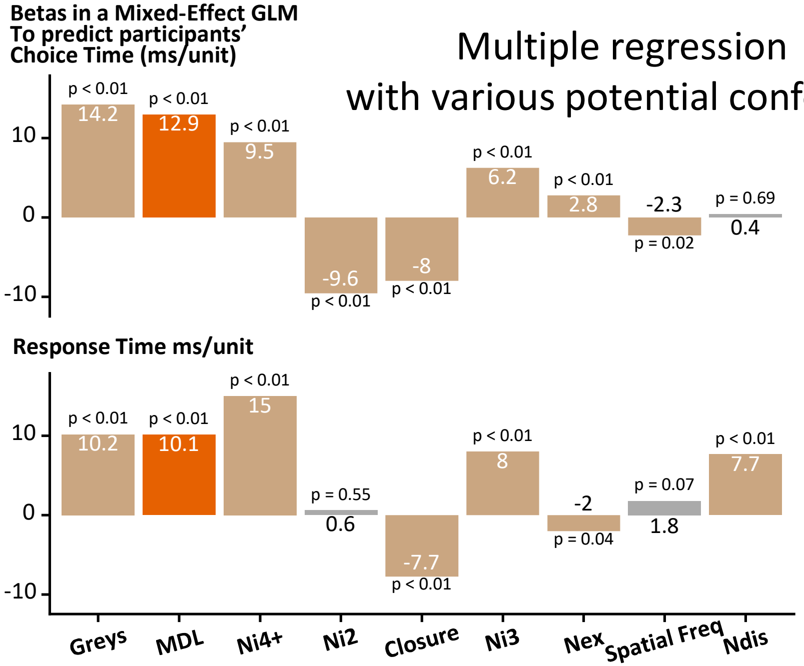


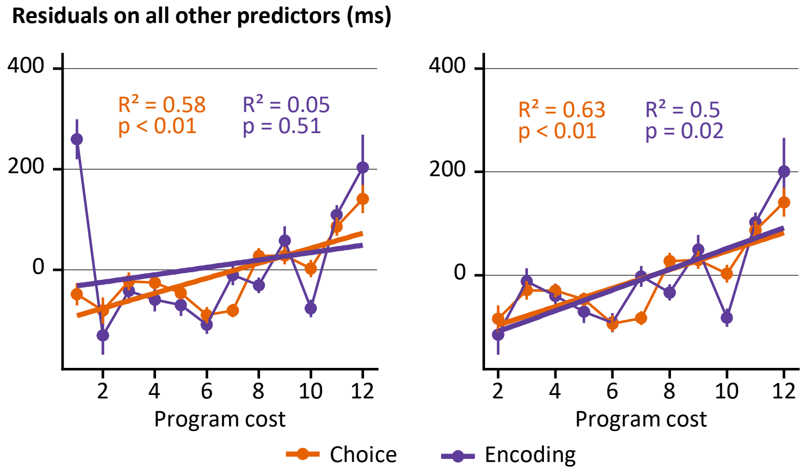
= Minimal description length (MDL)

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

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

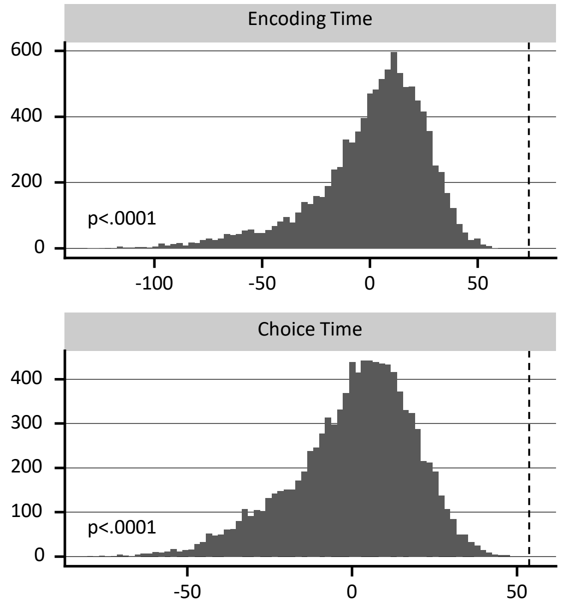
Multiple regression with various potential confounds





Various statistical controls

Bootstrap over alternative theories



Experiment 2. Testing a generic prediction about shape complexity

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

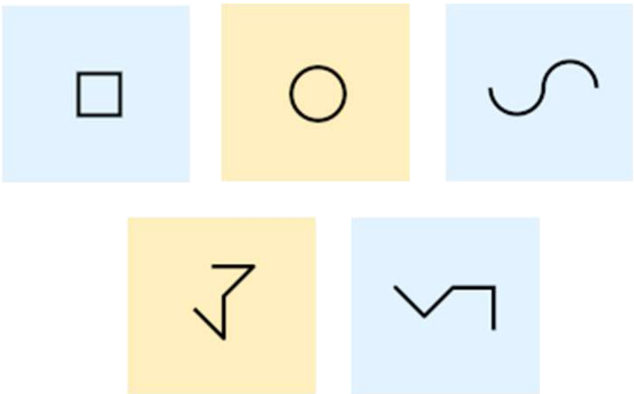
Prediction: Shape complexity should be determined by the **length of the shortest program capable of reproducing it.**

Perceptually rich drawings can be generated by **a single instruction**: *repeat*, *concat*, or *embed.*

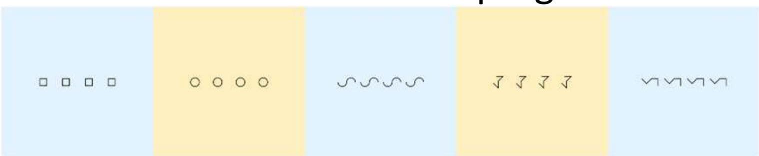
Complexity should follow **additive** rules:

Complexity (***Repeat*(x)**) = Complexity (x) + constant
Complexity (***Concat*(x,y)**) = Complexity (x) + Complexity (y) + constant'
Complexity (***Embed*(x,y)**) = Complexity (x) + Complexity (y) + constant''

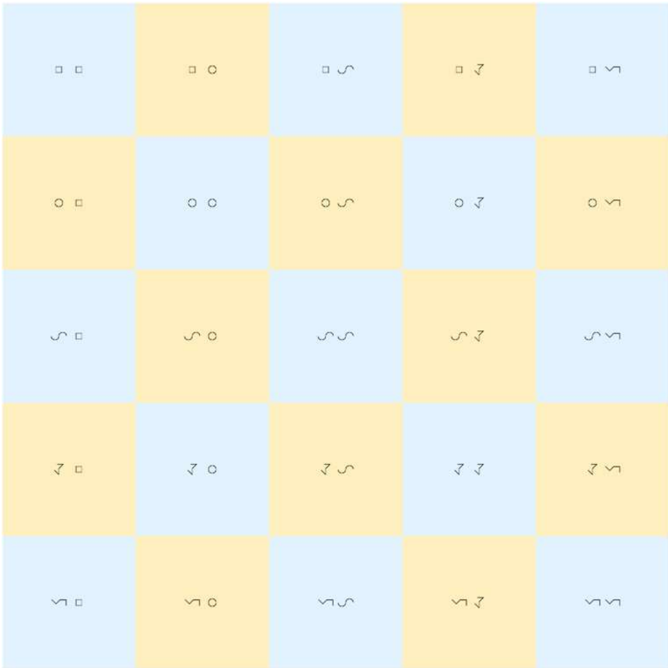
We selected 5 base shapes with increasing complexities



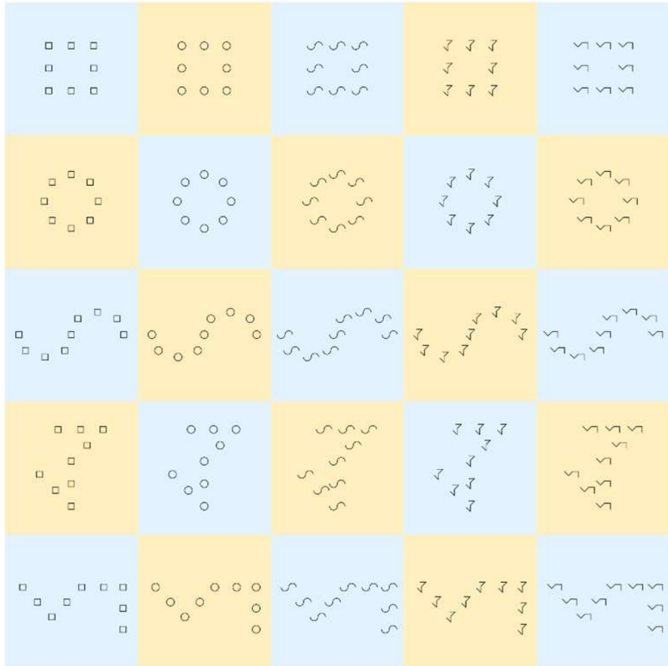
… and used them into programs:



Repeat



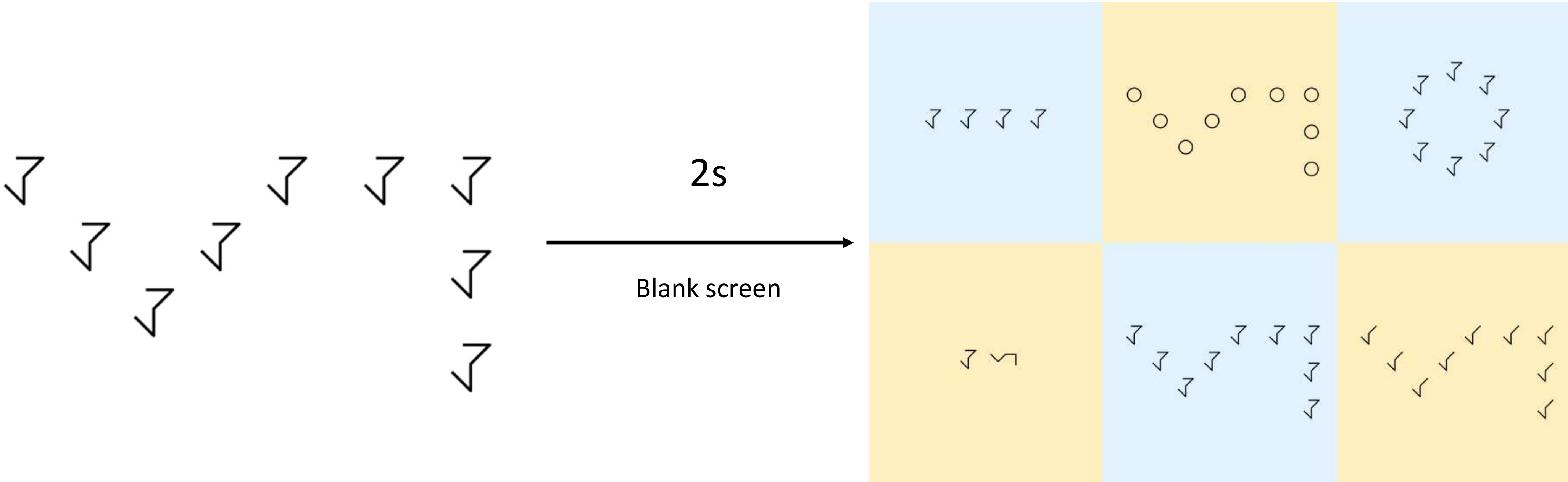
Concatenate



Embed

Two behavioral measures of shape complexity in humans

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



Phase 1 : Encoding

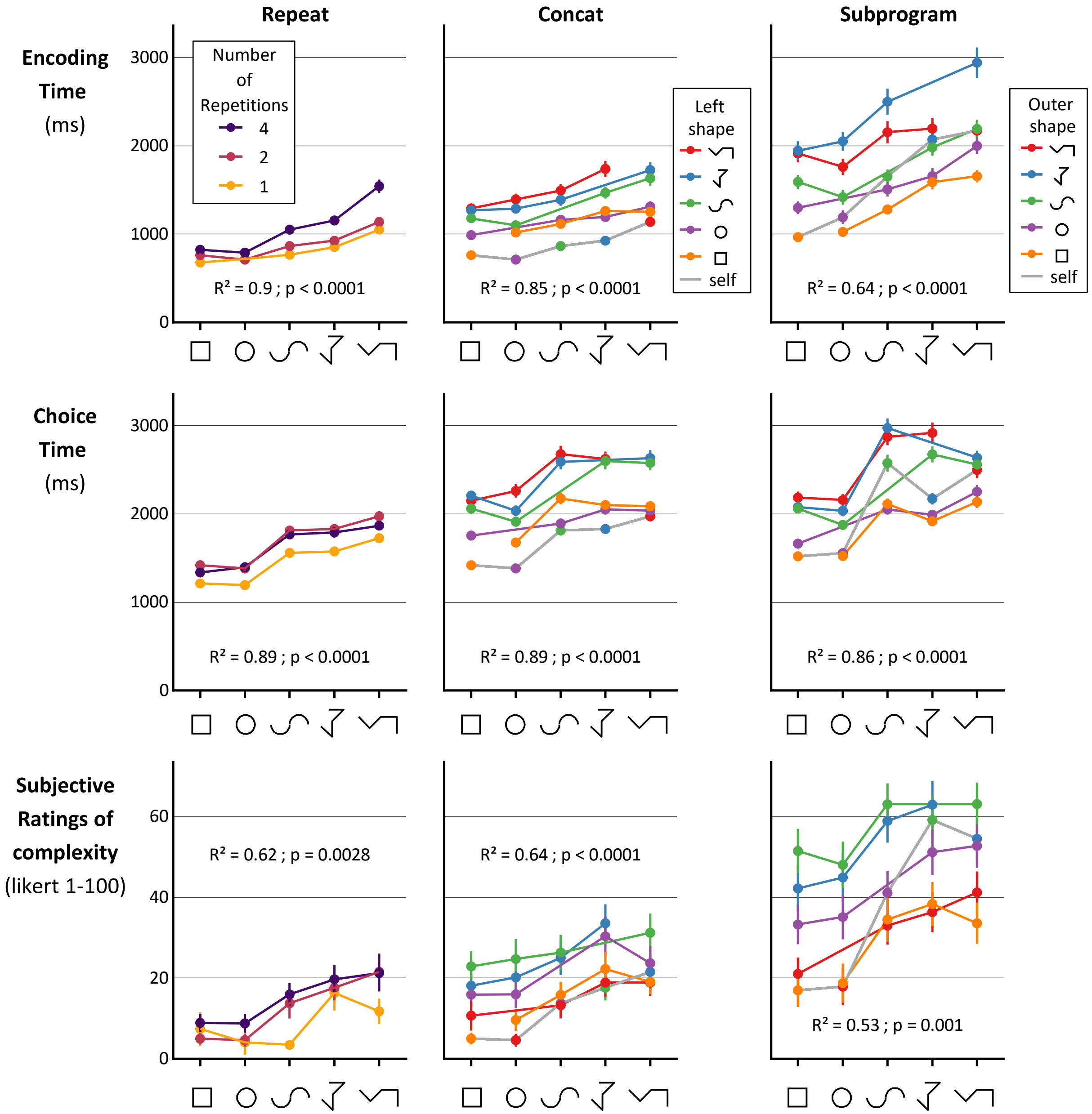
Subjects press a bar, then lift it when they feel they remember the pattern

measure = **encoding time**

Phase 2: Multiple choice

Subjects select the previous pattern among many mutants

measure = **choice time** (and errors)



Testing the predicted additive relationships

There is an effect of shape complexity even for individual shapes → different "programs"

This effect predicts what happens in other conditions:

- **Repetition** of a shape n times
  = addition of a term roughly proportional to log(n)

- **Concatenation** of two shapes
  = addition of the two complexities

no interaction term, once we remove the special case of two identical shapes

- **Embedding** of two shapes (e.g. a circle of squares)
  = addition of the two complexities, with steeper slopes

Again, no interaction term, but a special savings when the same program is used twice (e.g. a circle of circles)

Beyond geometry:
Is recursive symbolic compositionality unique to humans?

I speculate that only the human species possesses c**ompositional languages** that can produce an infinity of new expressions or « mental programs », based on the same small repertoire of basic concepts.

We probably possess the same **core knowledge** as other animals (objects, people, colors, numbers, probabilities, etc), but we **recombine these concepts using « languages of thought »**, which allows us to form an infinite pyramid or **coral** of nested thoughts.

Those languages are universal – all humans can think the same thoughts.

However, the space of mental expressions is so vast that different cultures may not make the same choices – linguistic communication and education orient attention to the branches that a given culture judges as most relevant.

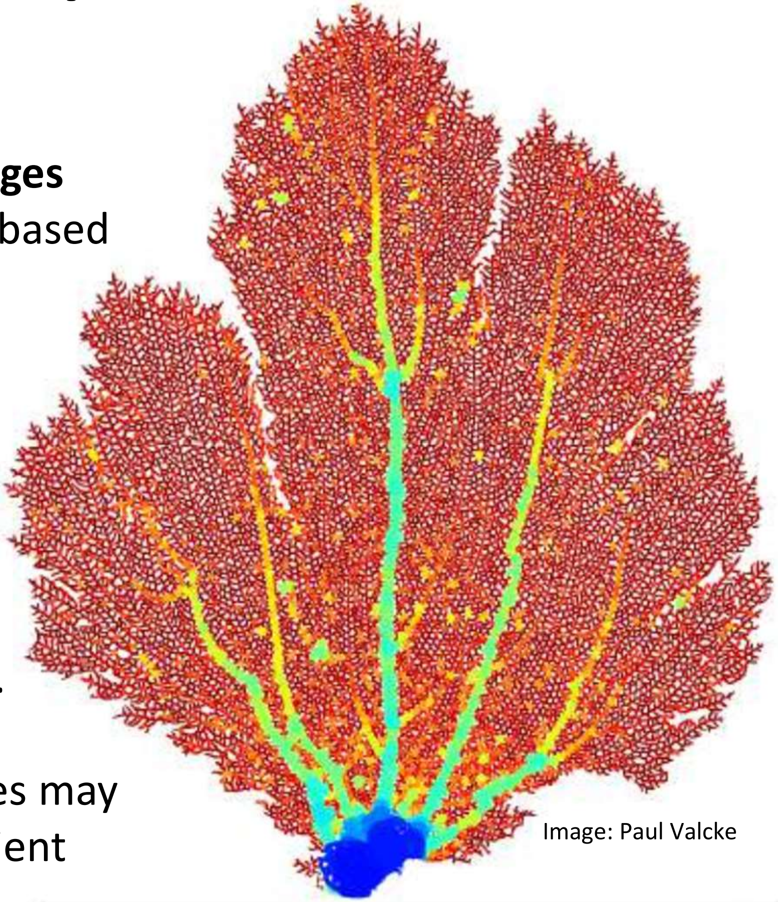
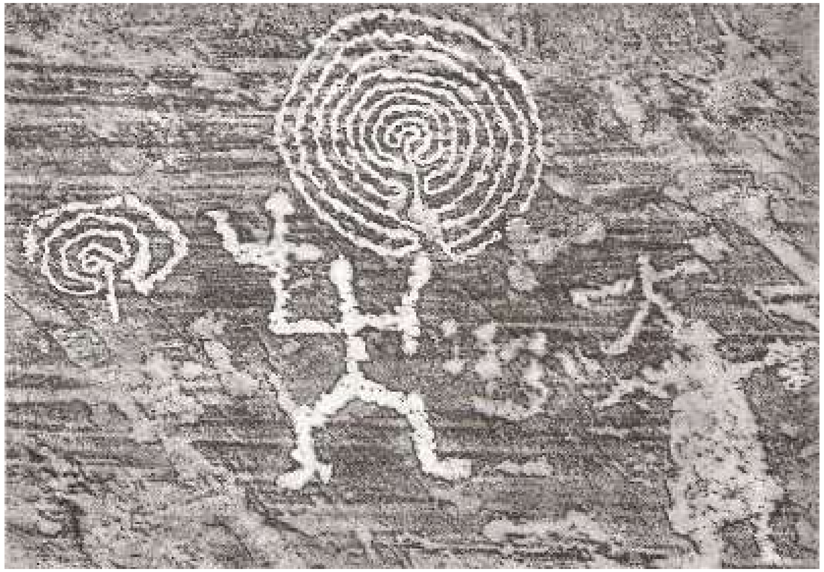


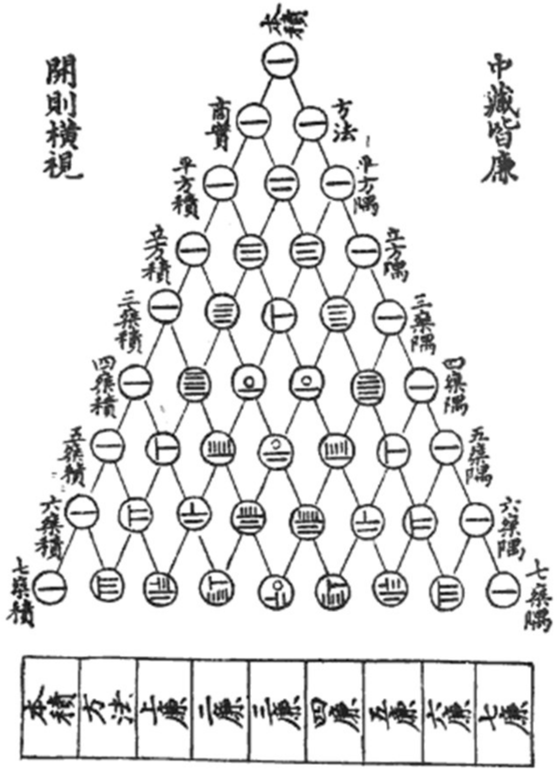
Image: Paul Valcke

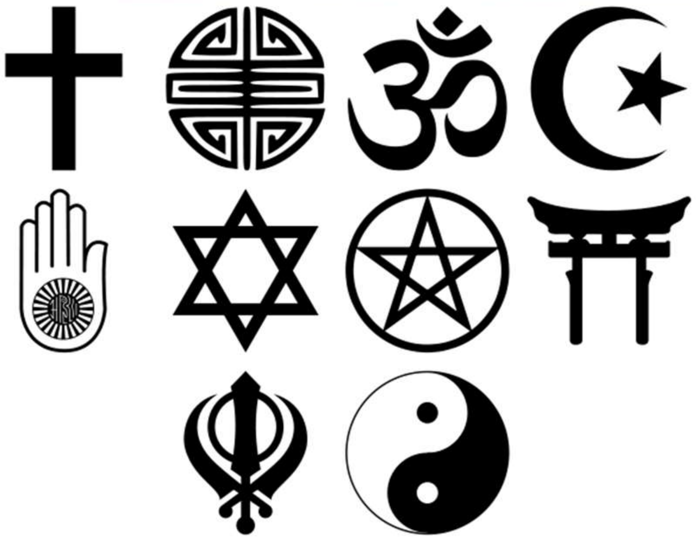
Consequences (1)

1. There is **a universal set** of shapes, patterns, expressions and concepts that **all humans find simple**

→ **cross-cultural convergence** towards the same ideas







Consequences (2)

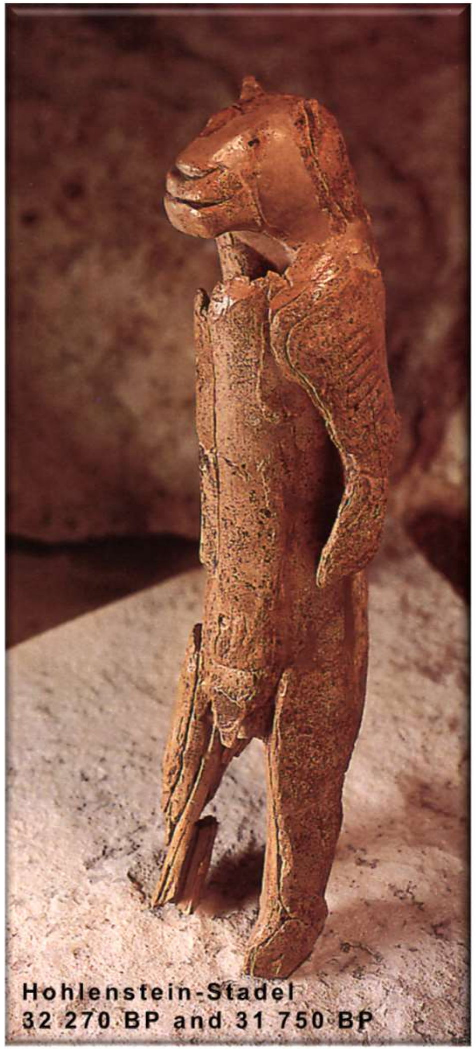
2. The **space of possible concepts** is exponentially large

→ human capacity to generate **infinitely many concepts**

→ **Extraordinary expansion** of representational abilities

“infinite use of finite means” : chimeras, imaginary ideas…

Chimeras as the reflections of human singularity



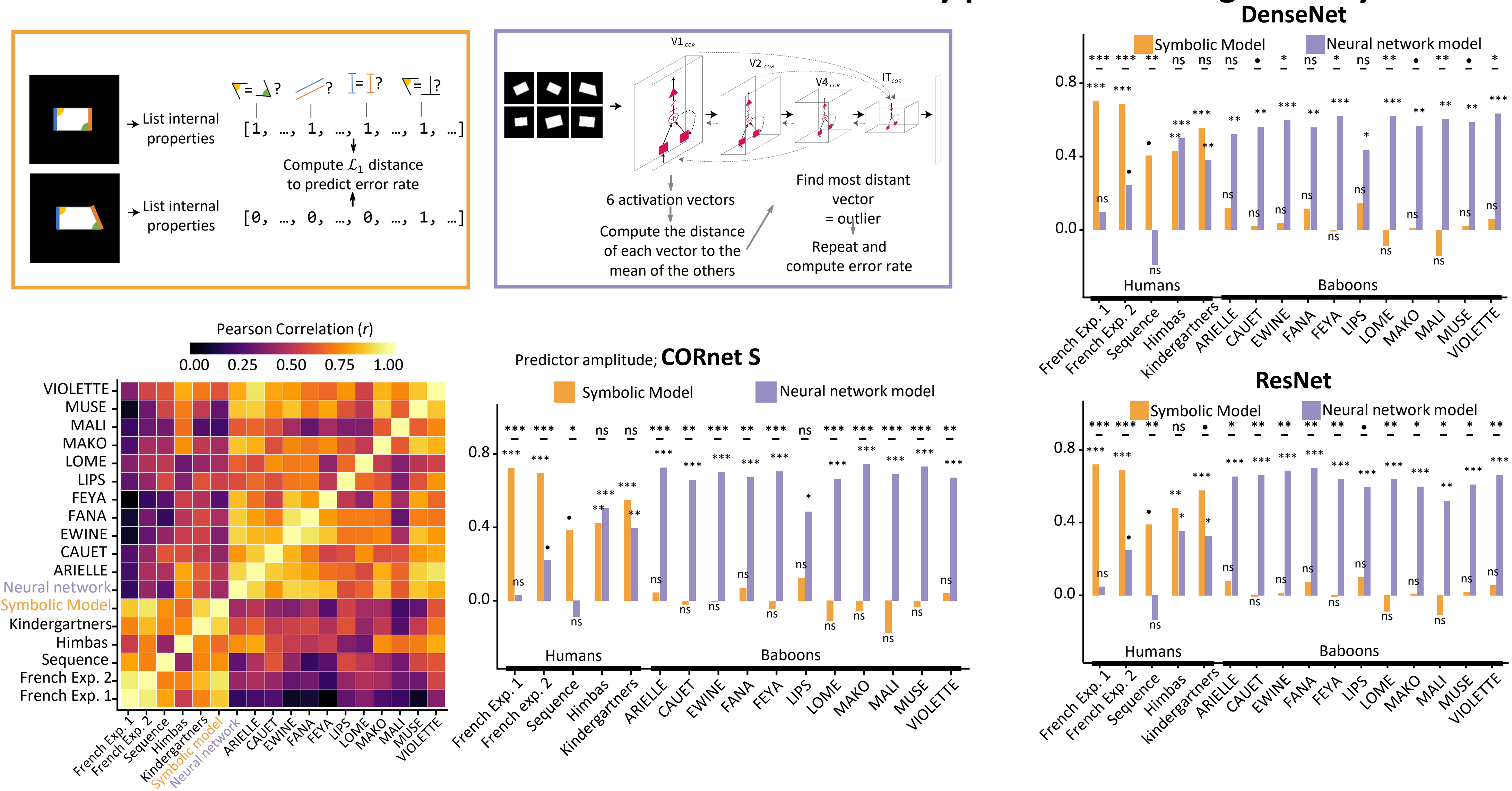
The Lascaux "unicorn"

The Dragon

The feathered snake (Quetzalcoatl)

Chimera of Arezzo (Etruscan)

Challenge 1. How is the language of thought implemented at the neural level?
Current convolutional neural networks are very poor models of geometry



Deep convolutional neural networks fall short of explaining human vision

Jacob, G., Pramod, R. T., Katti, H., & Arun, S. P. (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. Nature Communications, 12(1), Article 1. https://doi.org/10.1038/s41467-021-22078-3
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
Bowers, J. S., Malhotra, G., Dujmović, M., Montero, M. L., Tsvetkov, C., Biscione, V., Puebla, G., Adolfi, F., Hummel, J. E., Heaton, R. F., Evans, B. D., Mitchell, J., & Blything, R. (2022). Deep Problems with Neural Network Models of Human Vision. Behavioral and Brain Sciences, 1-74. https://doi.org/10.1017/S0140525X22002813

Many examples:

Jacob et al. 2021: “phenomena were absent in trained networks, such as 3D shape processing, surface invariance, occlusion, natural parts and the global advantage”

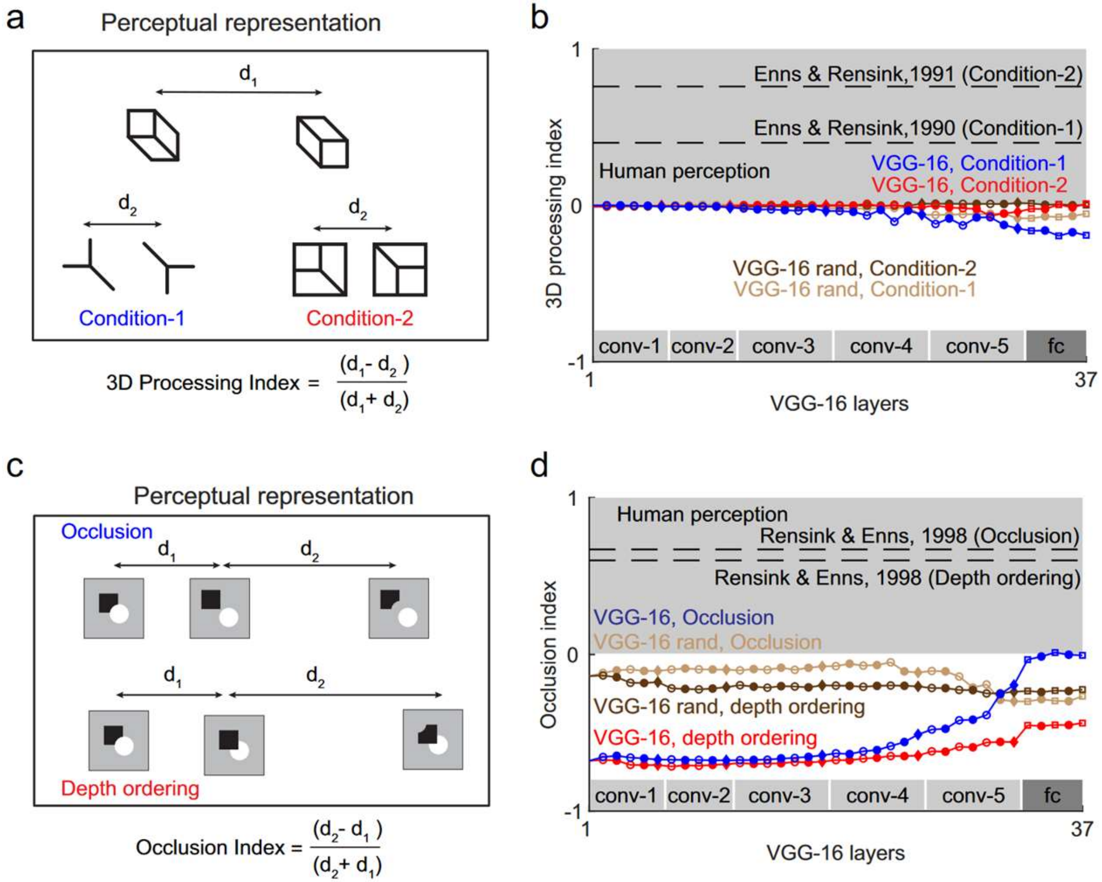
Ullman et al. 2016 show a non-linear collapse in human image recognition that does not exist in machines.

Bowers et al. 2022 go as far as to claim that “DNNs account for almost no results from psychological research” in shape representation.

(a) Texture image
81.4% **Indian elephant**
10.3% indri
8.2% black swan

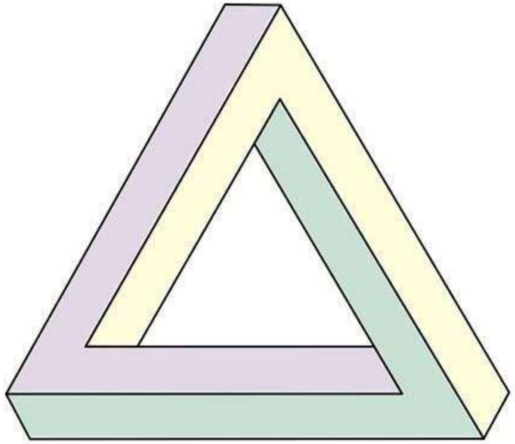
(b) Content image
71.1% **tabby cat**
17.3% grey fox
3.3% Siamese cat

(c) Texture-shape cue conflict
63.9% **Indian elephant**
26.4% indri
9.6% black swan



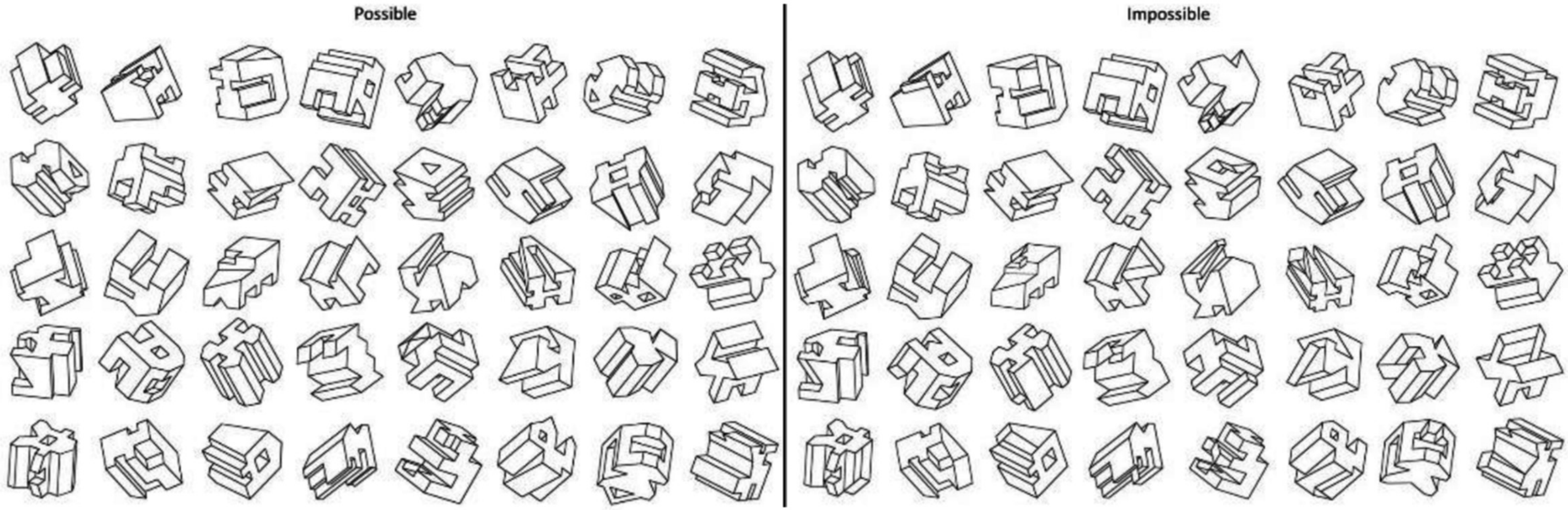
Convolutional neural networks are severely limited in accounting for geometric shape perception

Heinke, D., Wachman, P., Van Zoest, W., & Leek, E. C. (2021). A failure to learn object shape geometry : Implications for convolutional neural networks as plausible models of biological vision. *Vision Research*, *189*, 81-92. https://doi.org/10.1016/j.visres.2021.09.004



Impossible figures are a class of visual stimuli that can only be recognized at the global level. Thus, they offer a nice opportunity to analyze whether CNNs can reconstruct global and not just local features.

Even 4 month-old infants can detect impossible figures, and no training is required. Thus, it seems important to compare with both untrained and trained networks.



Convolutional neural networks are severely limited in accounting for geometric shape perception

Heinke, D., Wachman, P., Van Zoest, W., & Leek, E. C. (2021). A failure to learn object shape geometry : Implications for convolutional neural networks as plausible models of biological vision. *Vision Research*, *189*, 81-92. https://doi.org/10.1016/j.visres.2021.09.004

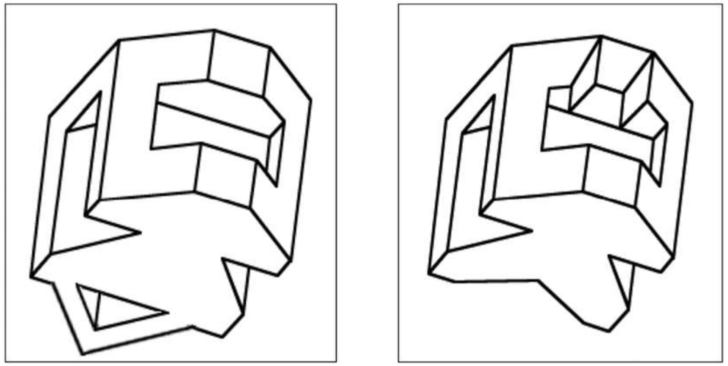
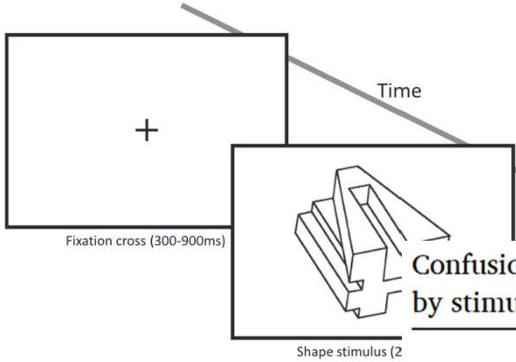


Fig. 3. Example of how a possible shape (left) was turned into an impossible shape (right).



Confusion matrix for human performance showing the percentage of responses by stimulus category (possible/impossible) and response.

		Response [%]	
		impossible	possible
Stimulus Category	impossible	42.6	7.4
	possible	5.9	44.1
	TOTAL	48.5	51.5

Result: all networks, trained or not, perform much worse than human adults.
The best network achieves 67% correct, as compared to 87% for humans.
The confusion matrix is quite different:

Table 3
Confusion matrix for four convolution neural networks using the original dataset. The confusion matrix shows the percentage of the network's response by stimulus category (possible/impossible) and response.

		GoogLeNet (trained) Response [%]		AlexNet (pre-trained) Response [%]		VGG-16 (pre-trained) Response [%]		ResNet-18 (pre-trained) Response [%]	
		impossible	possible	impossible	possible	impossible	possible	impossible	possible
Stimulus Category	impossible	34	16	12	38	25	29	21	30
	possible	12	38	1	49	3	47	9	41
	Total	46	54	13	88	26	74	29	71
	SDT	d':0.94; c: 0		d': −1.35; c: 0.03		d': 0.05; c: −0.03		d': −0.46; c: −0.03	

Symmetry : an ecologically valid prior in human object perception

McBeath, M. K., Schiano, D. J., & Tversky, B. (1997). Three-dimensional bilateral symmetry bias in judgments of figural identity and orientation. *Psychological Science*, *8*(3), 217-223.

Most biological objects have at least bilateral symmetry, often around the vertical. Some have more symmetries (e.g. rotational symmetry for trees).

Animals, including humans, exhibit a sexual preference for symmetrical bodies and faces.

Symmetry is frequently assumed, for instance when inferring the shape of an occluded object.

Viewers detect symmetrical objects faster and, in their memory encoding impose more symmetry than was actually present

McBeath et al. (1997) presented 2-D shapes that are either random polygons (0%), or symmetrical around the vertical axis (100%), or in between, and asked for a description.

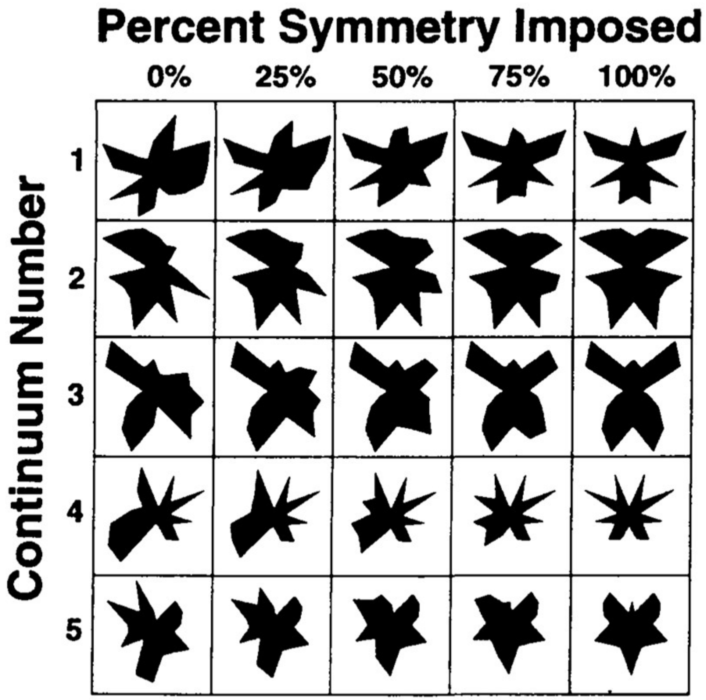
Results:

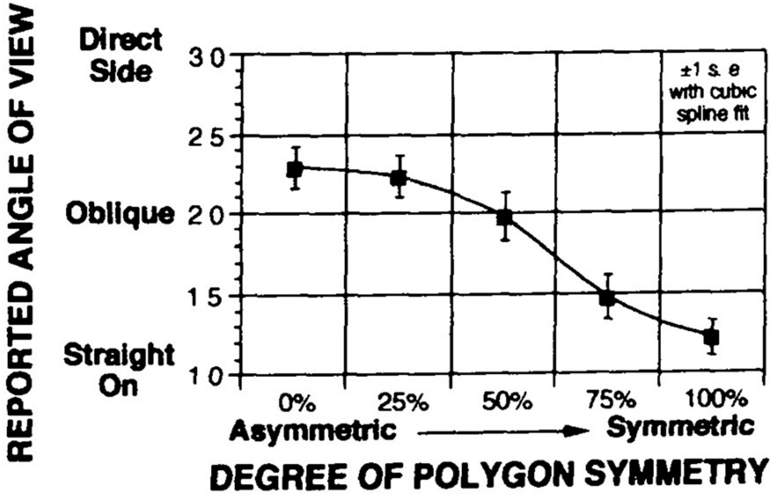
- nearly all stimuli (90%) are interpreted as 3D symmetrical, regardless of their 2D symmetry

e.g. A typical cartoon dog head, turned to the right

- the 3D orientation is systematically influenced by the degree of symmetry: symmetrical 2D shapes are considered as “head on”, and asymmetry is interpreted as an indication that the symmetrical object is viewed from the side.

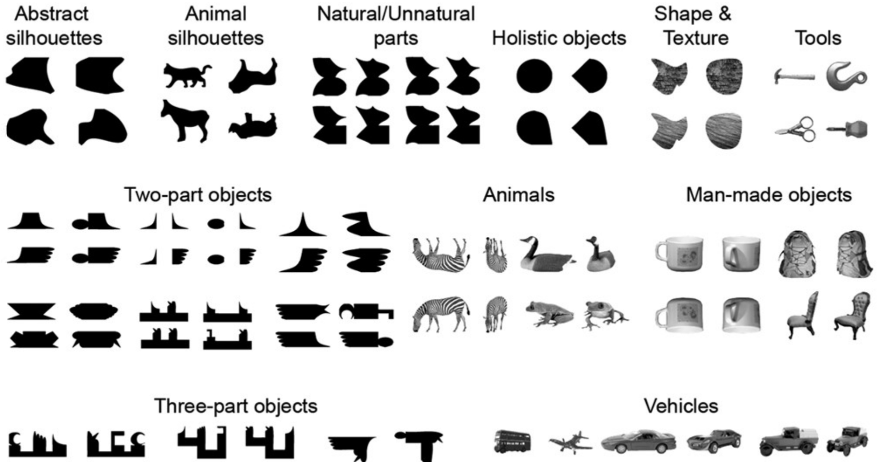
- Identifying the rotated objects took more time, suggest a stage of mental rotation.



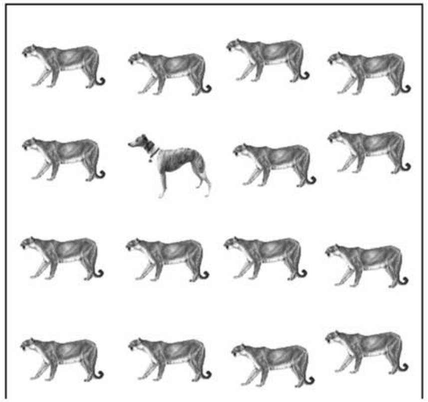


Connectionist models of visual recognition miss a concept of symmetry

Pramod, R., & Arun, SP. (2022). Improving Machine Vision Using Human Perceptual Representations : The Case of Planar Reflection Symmetry for Object Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1), 228-241.



Humans estimated the similarity between various object pairs, using a visual search task.



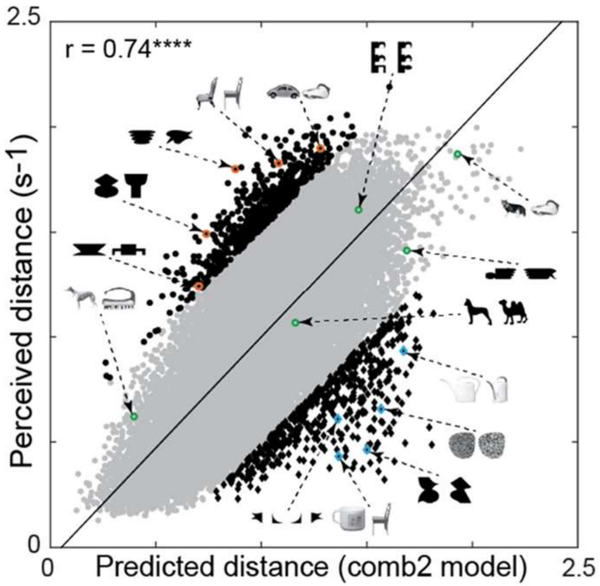
The data was compared to a variety of convolutional neural networks.

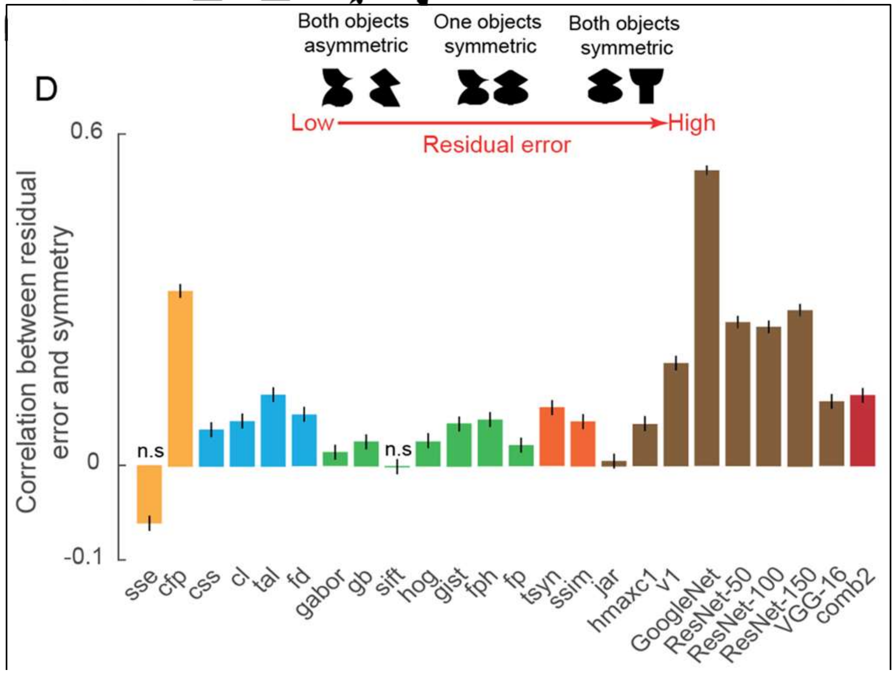
For object pairs that share a feature of symmetry, dissimilarity is underestimated.

The authors show that neural networks can be improved by adding extra features of symmetry at the input stage.

However, this is just a patch, because the human system is able to detect symmetries at various nested levels.

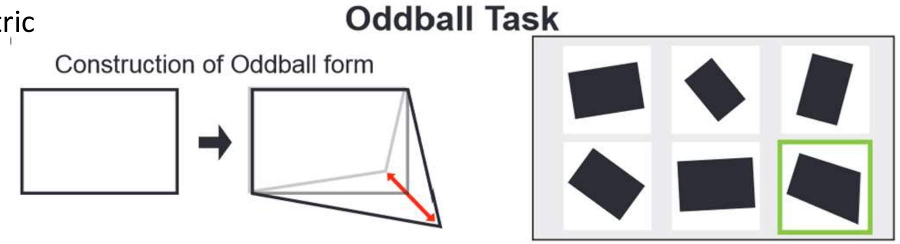
Conclusion: **current CNNs vastly underestimate the sophistication of human vision.**





Can advanced artificial neural networks solve our geometric tasks?

Campbell, D., Kumar, S., Giallanza, T., Griffiths, T. L., & Cohen, J. D. (2024). Human-Like Geometric Abstraction in Large Pre-trained Neural Networks (arXiv:2402.04203). arXiv. https://doi.org/10.48550/arXiv.2402.04203



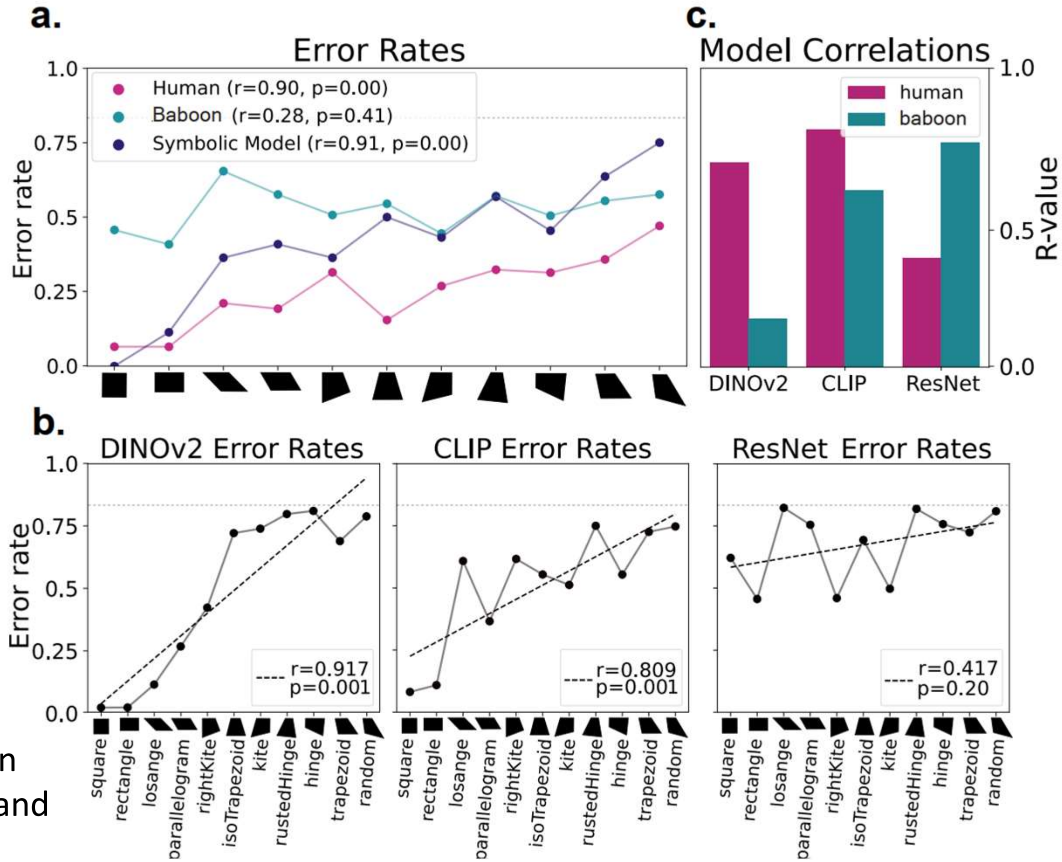
3 recent AI vision models are challenged with our test :
- ResNet : a convolutional neural network
- CLIP : a transformer for vision and language
- DinoV2 : a large, 1-billion parameter vision transformer, trained both to identify image similarity up to affine augmentations, and to complete image patches.

Like us, they extract the embeddings evoked by each of the six shapes (in the last layer?), and define the outlier as the one which is maximally different from the mean of the others.

They replicate our findings: ResNet, a simple convolutional network, does not predict the human geometrical regularity effect – but only the baboon data.

However, they also find that DinoV2 and, to a lesser extent, CLIP can predict human behavior relatively well (though not with perfect linearity, as the symbolic model does).

Interestingly, CLIP is in between, captures a bit of both human and baboon data, and resembles the profile of human preschoolers (it treats squares and rectangles as radically different).



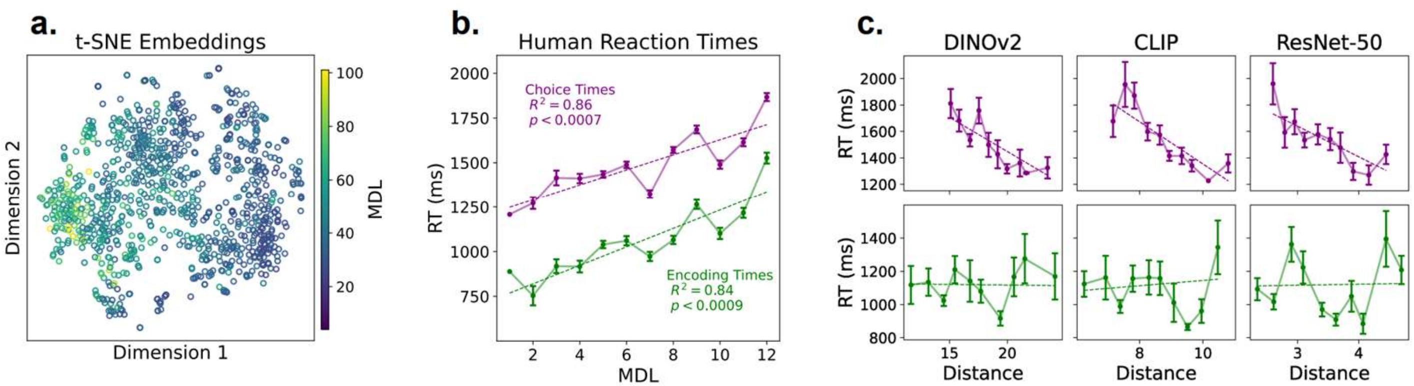
Can advanced Large Language models (LLMs) implement a language of geometry?

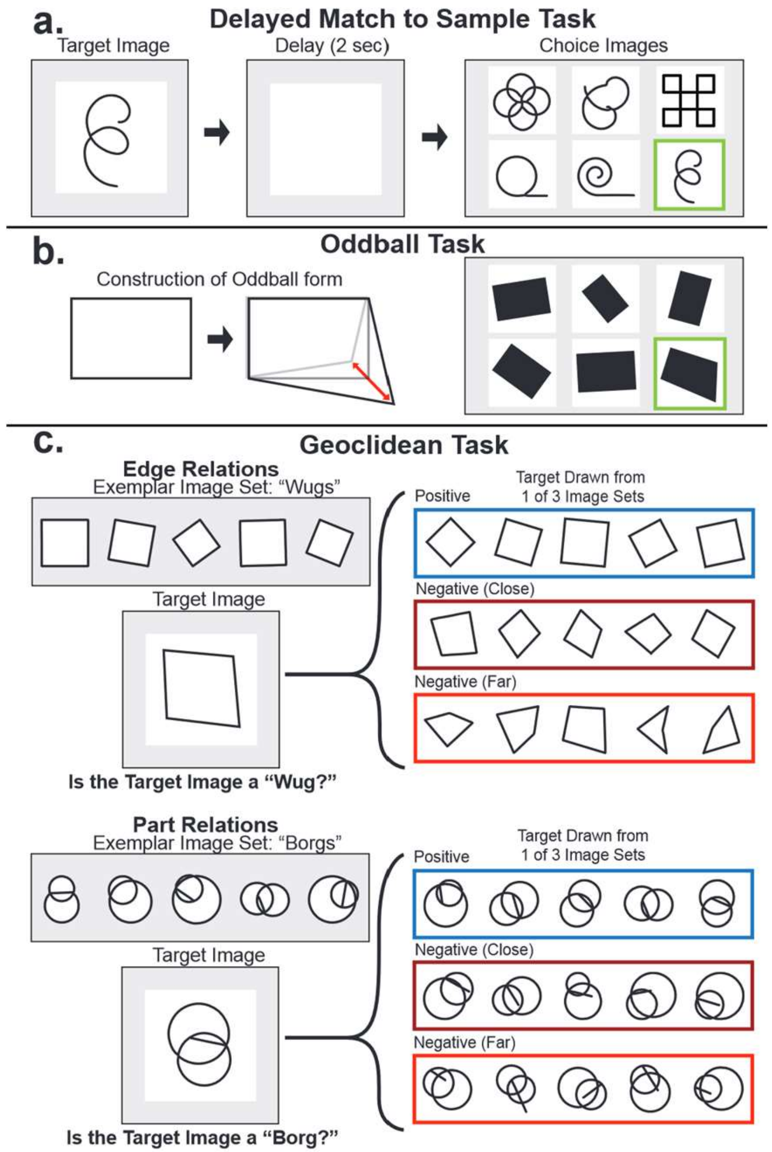
Campbell, D., Kumar, S., Giallanza, T., Griffiths, T. L., & Cohen, J. D. (2024). Human-Like Geometric Abstraction in Large Pre-trained Neural Networks (arXiv:2402.04203). arXiv. https://doi.org/10.48550/arXiv.2402.04203

The authors identify three signatures of human geometric processing:
- **geometric complexity** (captured by minimal description length),
- **geometric regularity** (based on parallelism, right angles, symmetries…)
- **decomposition** into geometric parts and relations when learning new categories.

They test whether those human phenomena emerge with training in 3 current AI models: ResNet (a convolutional neural network), CLIP (a transformer for vision and language), and DinoV2 (a large, 1-billion parameter vision transformer, trained both to identify image similarity up to affine augmentations, and to complete image patches).

For **geometric complexity,** they show that low vs high MDL shapes tend to be encoded in different parts of the embedding space. Thus, they can account for choice time with the distance between target and distractors – but not for encoding time (or barely with dinoV2)

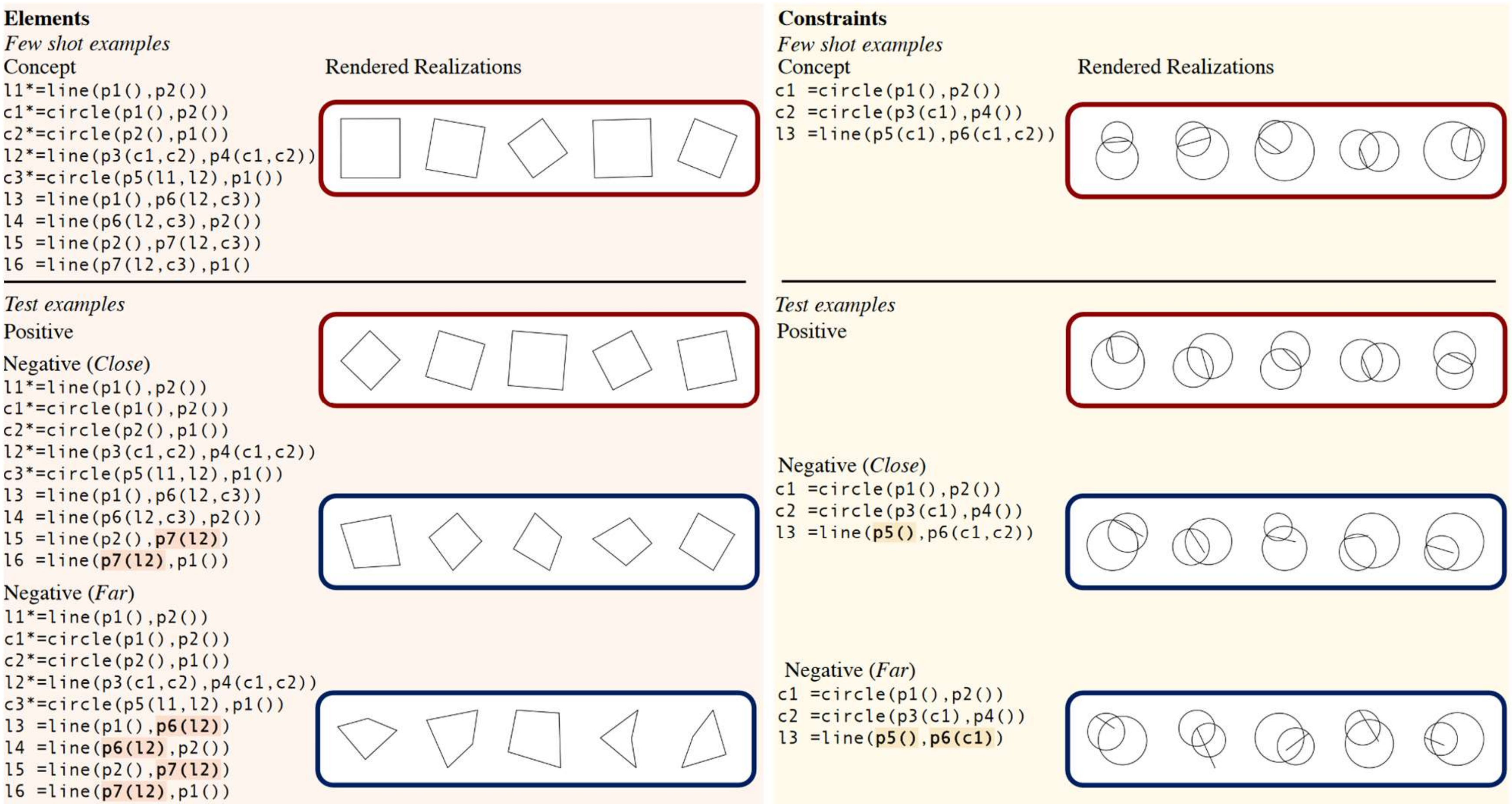




The Geoclidean benchmark: more evidence for human geometrical intuitions

Hsu, J., Wu, J., & Goodman, N. (2022). Geoclidean : Few-Shot Generalization in Euclidean Geometry. *Advances in Neural Information Processing Systems*, *35*, 39007-39019.

For **decomposition** into geometric parts and relations when learning new categories: the Geoclidean benchmark (Hsu et al.) asks subjects to learn a new concept ("dork") from five positive examples, and then distinguish new instances from close and far negative ones.
The items are generated by a **domain specific language (DSL)** which incorporates Euclid's axioms for point, line and circle.



The Geoclidean benchmark: more evidence for human geometrical intuitions

Hsu, J., Wu, J., & Goodman, N. (2022). Geoclidean : Few-Shot Generalization in Euclidean Geometry. *Advances in Neural Information Processing Systems*, *35*, 39007-39019.

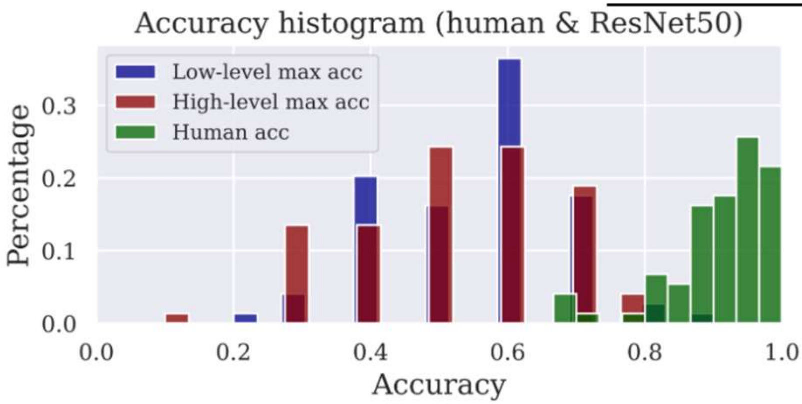
Hsu et al. test human adults on 37 concepts
- 17 concepts that are constructed in Euclid's Elements (e.g. triangle)
- 20 concepts that are pure combinations of primitives (Line, Circle, Triangle)

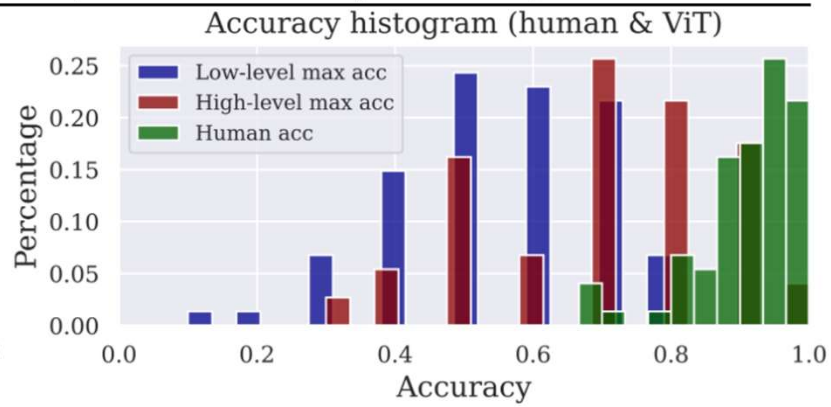
The results show
- a very high accuracy for human subjects for all concepts, with a slight advantage for the regular ones in Euclid's book
- with slightly better rejection of 'far' than of 'close' negative examples (distance effect)
- Current AI neural networks are unable to match this performance (although they are above chance, especially in high-level layers, and VisualTransformer fares better).

Campbell et al. replicate those findings entirely: neither CLIP nor DinoV2 do better than ViT, and the benchmark remains unsolved by machines.

Table 2: Human accuracy across all 74 tasks in Geoclidean.

CONCEPT	*Close*	*Far*	CONCEPT	*Close*	*Far*
ANGLE	0.9767	0.9833	LLL	0.9700	0.9667
PERP BISECTOR	0.9367	0.9833	CLL	0.9467	0.9667
ANG BISECTOR	0.9433	0.9533	LLC	0.6767	0.9233
SIXTY ANG	0.8233	0.9533	CCL	0.8700	0.8833
RADII	0.9233	0.9600	LCC	0.8867	0.9633
DIAMETER	0.9567	1.0000	CCC	0.6667	0.8767
SEGMENT	0.9300	0.9833	LLLL	0.8833	0.9767
RECTILINEAR	0.9000	0.9033	LLLC	0.6667	0.8867
TRIANGLE	0.9633	0.9767	CLLL	0.8367	0.9033
QUADRILATERAL	0.9167	0.9267	CLCL	0.8700	0.8567
EQ T	0.9533	0.9800	LLCC	0.8867	0.9333
RIGHT ANG T	0.7200	0.8133	CCCL	0.9233	0.9333
SQUARE	0.8933	0.9867	CLCC	0.8633	0.9000
RHOMBUS	0.9367	0.9667	CCCC	0.8167	0.8800
OBLONG	0.9666	0.9900	TLL	0.9467	0.9800
RHOMBOID	0.9700	0.9300	LLT	0.9267	0.9400
PARALLEL L	0.9500	0.9567	TCL	0.9533	0.9633
			CLT	0.9533	0.9633
			TCC	0.9533	0.9633
			CCT	0.9533	0.9633

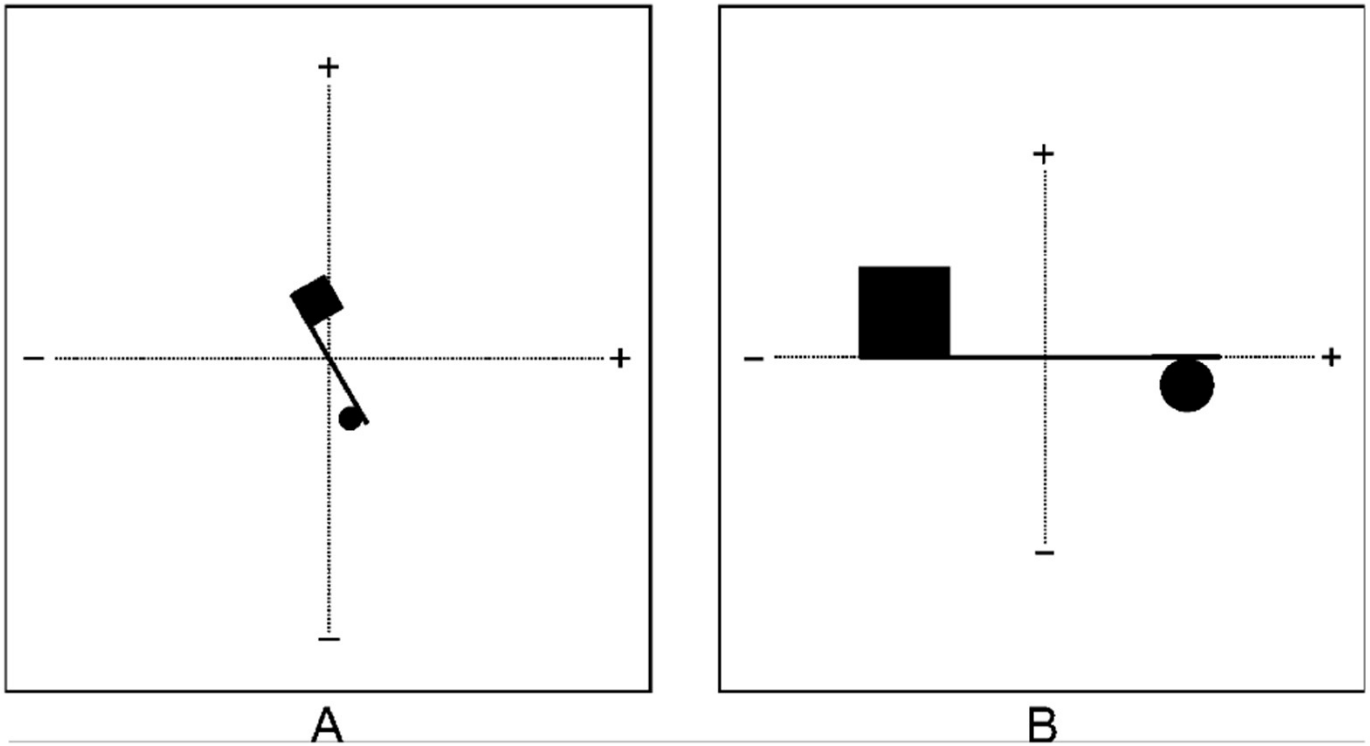




A challenge for both language-of-thought and connectionist models: Evidence for a compositional representation of objects, axes, and orientation

Harris, I. M. (2024). Interpreting the orientation of objects : A cross-disciplinary review. Psychonomic Bulletin & Review.
McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation : A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680-713. https://doi.org/10.1080/02643290500538356

It is essential to represent the object in an invariant, object-centered frame of reference, regardless of its orientation.
The visual system seems to assign a coordinate system of orthogonal axes centered on the object, based on its principal axis.



A compositional representation of objects, axes, and orientation is needed to explain the errors of many patients

McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation : A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680-713. https://doi.org/10.1080/02643290500538356

Patient B.C.'s orientation errors:
A vast majority of left-right inversion
along the vertical axis.

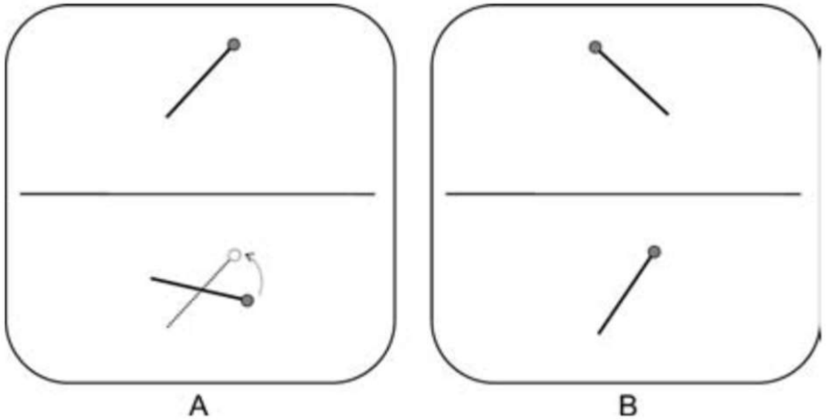
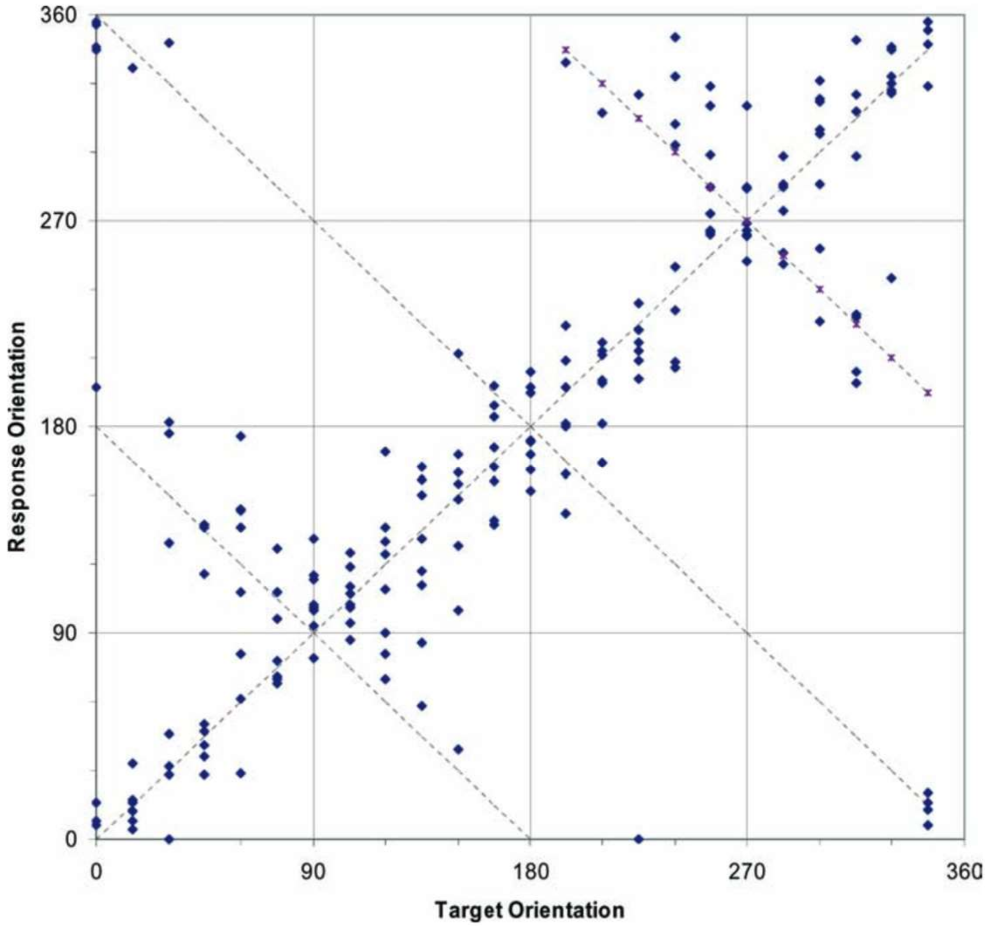


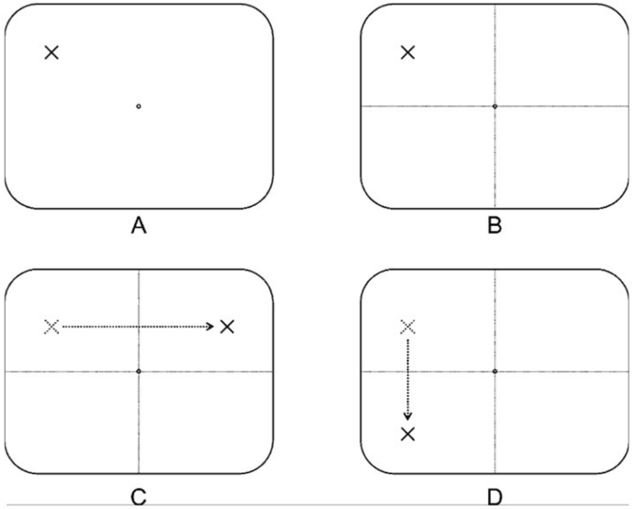
Figure 14. *The line orientation reproduction task in which stimulus lines had a red tip at one end (indicated by the grey dots in the figure).* ***A.*** *Target and response lines displayed on a computer monitor.* ***B.*** *Example of a left–right reflection error made by B.C.*



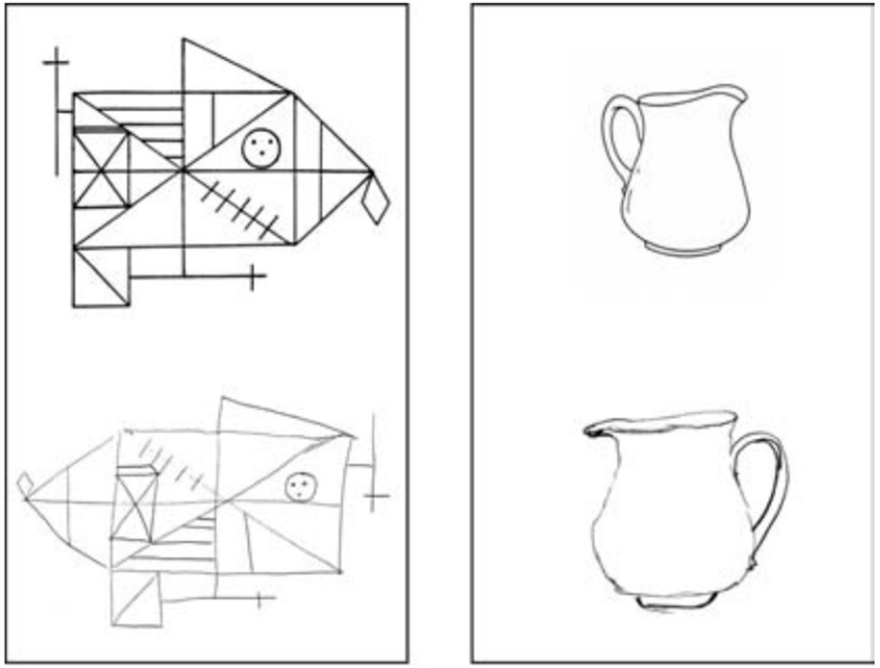
A compositional representation of objects, axes, and orientation is needed to explain the errors of many patients

McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation : A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680-713. https://doi.org/10.1080/02643290500538356

Patient A.H.'s mirror errors in pointing to a cross just after it vanished



Patient A.H.'s copying of drawings:



Note that such errors are ambiguous:
They could refer a left-right inversion along the vertical axis, but also a polarity inversion around one of the object's axis.

A crucial test: Patient A.H.'s copying of chiral, tilted figures

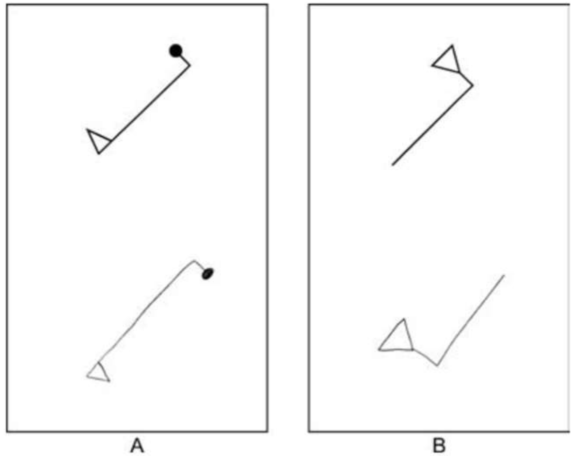


Figure 23. *A.H.'s direct copies of two asymmetric figures. A. Reflection across the principal object axis. B. Reflection across the secondary object axis.*

The errors are primarily reflections along the object's own axes, implying an impaired object-centered representation.

The need for geometrical models of object perception and orientation

Vannuscorps, G., Galaburda, A., & Caramazza, A. (2021). Shape-centered representations of bounded regions of space mediate the perception of objects. *Cognitive Neuropsychology*, 1-50.
Vannuscorps, G., Galaburda, A., Falk, E., & Caramazza, A. (2017). A developmental deficit in seeing the orientation of typical 2D objects. *Journal of Vision*, *17*(10), 28-28. https://doi.org/10.1167/17.10.28

Through more than 100 experiments, Patient Davida appears to have a very specific deficit of conscious vision : she assigns the wrong orientation to objects.
« Davida perceives 2D regions of space bounded by some types of edges (sharp luminance and chromatic edges) alternating between their correct orientation and all the other orientations that would result from their mirroring across one or both axes of their own "**shape-centered**" **coordinate system**, their rotation by 90, 180 or 270 degrees around their center, or both".
The axes are perpendicular or symmetrical to the true one, which implies that she has extracted it! And also, she perceives the correct orientation *less often* than chance, which remains unexplained.
"all the types of errors she made corresponded to a specific failure to specify the correct axis correspondence and axis polarity correspondence necessary to map ISCRs [intermediate shape-centered representations] onto higher frames"

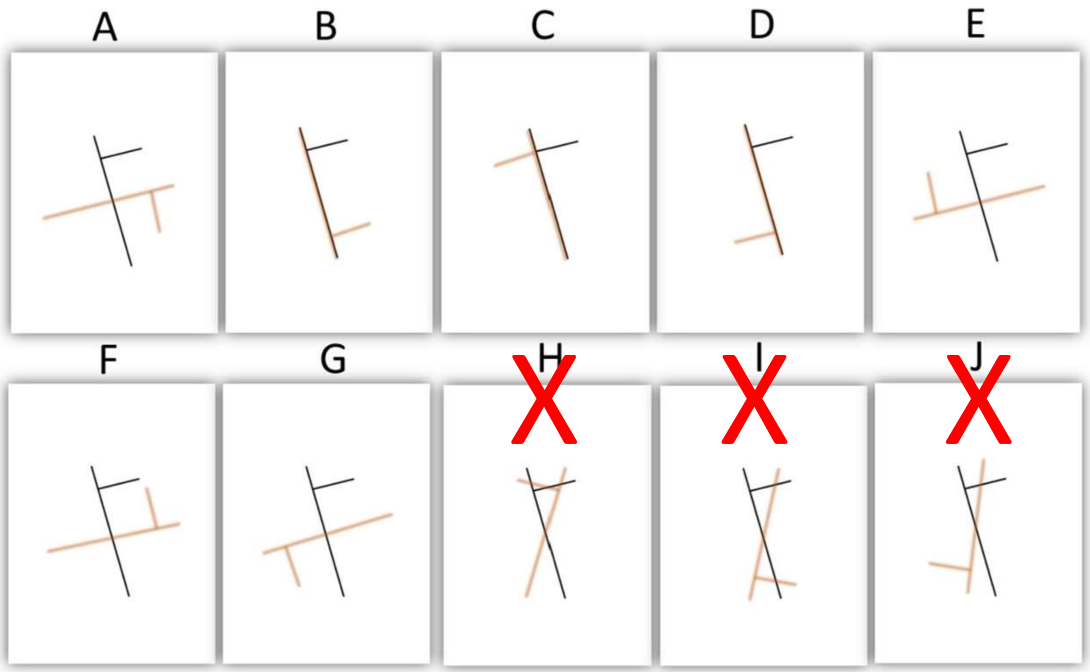


Figure 1. Example of stimuli (in black ink) and corresponding potential types of errors (in red ink). Davida's errors consist in seeing the stimulus as if it were rotated by 90 degrees around its center (A), reflected across one (B–C) or both (D) of its own shape-centered axes, or resulted from a combination of these types of errors (E–G). However, she never reported perceiving tilted stimuli as if it were flipped vertically (H), horizontally (I) or both (J).

A challenge for both language-of-thought and connectionist models: Evidence for a compositional representation of objects, axes, and orientation

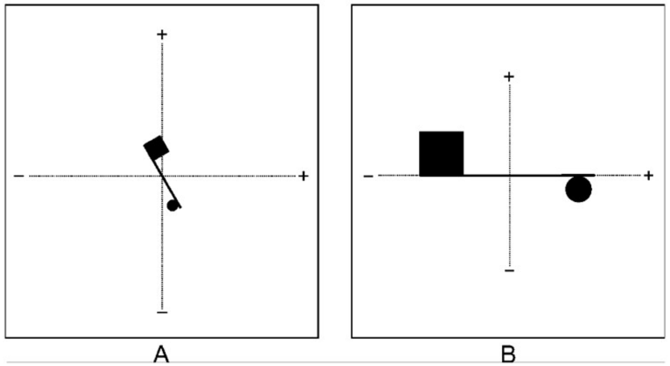
Harris, I. M. (2024). Interpreting the orientation of objects : A cross-disciplinary review. Psychonomic Bulletin & Review.
McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation : A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680-713. https://doi.org/10.1080/02643290500538356

It is essential to represent the object in an invariant, object-centered frame of reference, regardless of its orientation.
The visual system seems to assign a coordinate system of orthogonal axes centered on the object, based on its principal axis.

Key assumptions of the McCloskey et al. model:
Mental representations of objects and their orientation comprise

1. an **object-centered** (orientation-invariant) representation
2. a **frame of reference** external to the object.
3. a representation of the **axis correspondence**, consisting of several independent components:

- One component specifies which object axes map onto which external axes.
- A second component specifies, for each object axis, how the poles of that axis map onto the poles of the corresponding external axis.
- A third component specifies the tilt of the object reference frame relative to the external frame.
- The tilt component separately represents direction and magnitude of tilt.



Relation to frame of reference:	**Object-centered representation :**
Axis correspondence:	BAR
Principal-Vertical	LOCATION (0, 0)
Secondary-Horizontal	TILT (0°)
Polarity correspondence:	CIRCLE
-/+ and +/+	LOCATION (-20, – 5)
Axis Tilt:	SQUARE
Direction –	LOCATION (–20, -10)
Magnitude 30°	TILT (0°)

The shape skeleton: a mathematical representation of the axes of shapes

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

Blum (1973) develops a sophisticated, yet still highly intuitive, mathematical description of shapes.
First, he defines the notion of **distance from a point to an object** (or to its boundary) – the shortest Euclidean distance to any point in the object (or its boundary). Distance is measured along a **normal** or **radial** line.

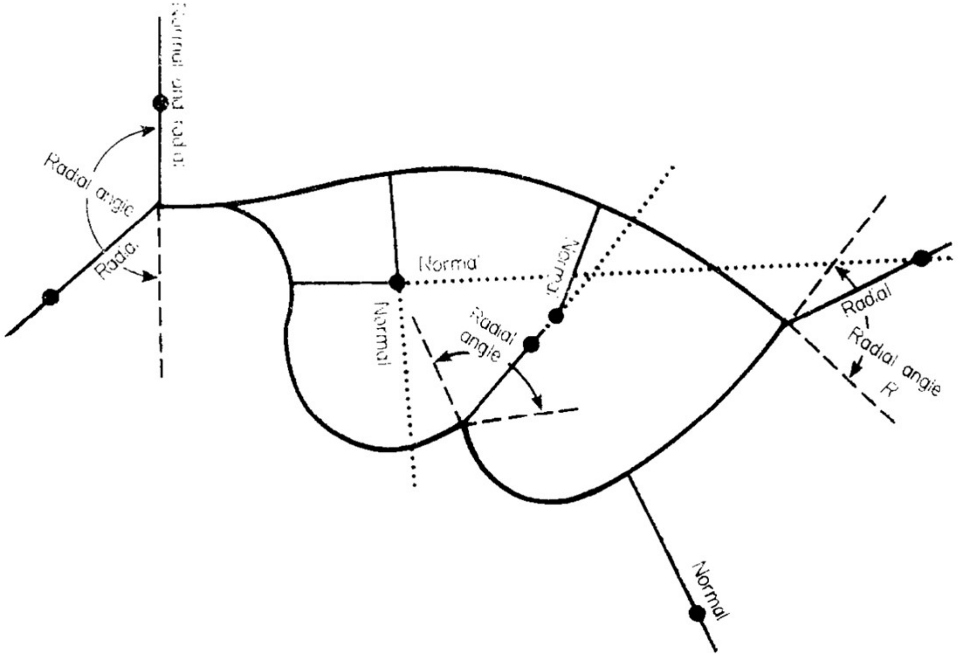
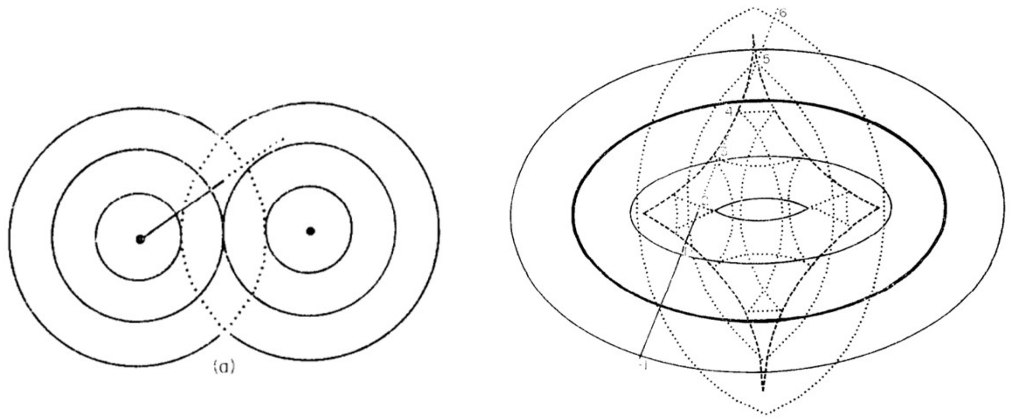
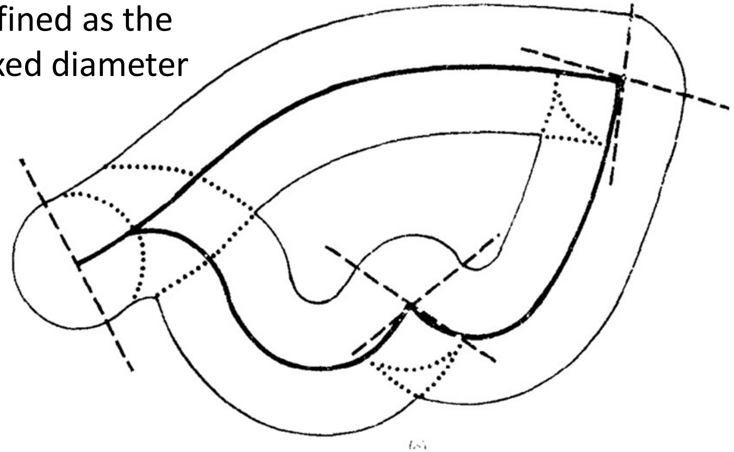


FIG. 4. Pannormals, lines along which distance between a point and object is measured. Normals issue from smooth points, radials from non-smooth points. Solid lines show nearest pannormals, ones using nearest distance constraint. Dotted lines show global pannormal extensions. These have no such constraint.



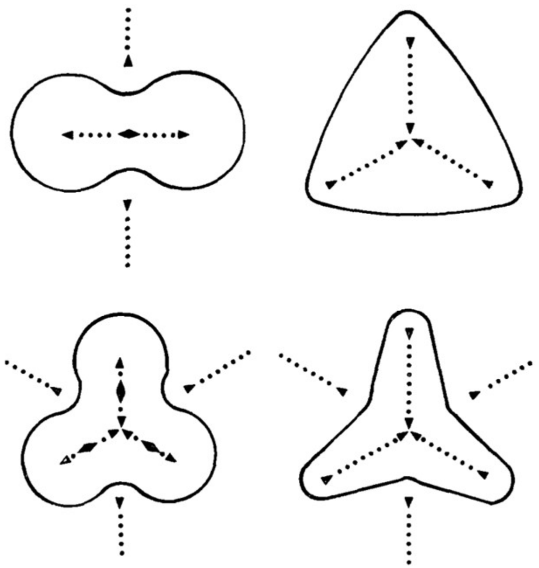
From this distance, one can define **parallels to a given curve.**
There are already some interesting observations, for instance non-commutativity: C’ can be a parallel to C, but C is not a parallel to C’.
Note: a parallel can be defined as the **envelope** of circles of a fixed diameter moving along the curve.



The shape skeleton: a mathematical representation of biological shapes

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

Moving along a given parallel, **sym-points** are special points beyond which the normal "jumps" discontinuously to another point with minimal distance. At a **Sym-point**, one can fit a circle that touches the curve at 2 places.
The **symmetric axis** or **medial axis** or **skeleton** is the set of all sym-points.
The **sym-distance** is the value of the minimal distance at each point on the skeleton.
Fundamental property : the locus of the skeleton + its associated sym-distance function suffices to fully describe the shape !
The object is the union of disks of *sym-distance* radius, covering the entire symmetric axis (this is called the *inverse transform*).



Here are some examples of shape skeletons. Symmetric axes can also exist outside the object (for non-convex shapes). One gets a feeling that the axes defined "what happened" (the object was squeezed, a protuberance grew, etc). This view is developed by Leyton (1989).

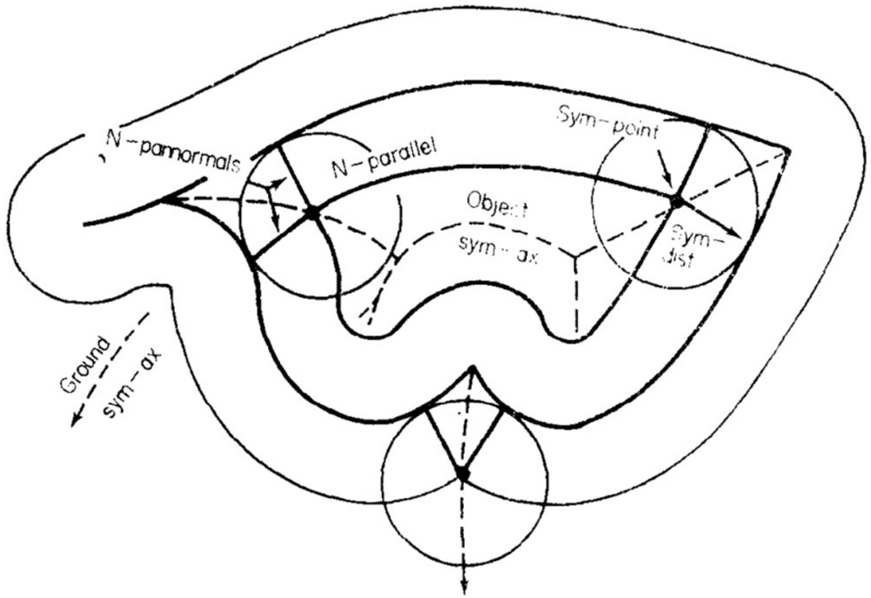


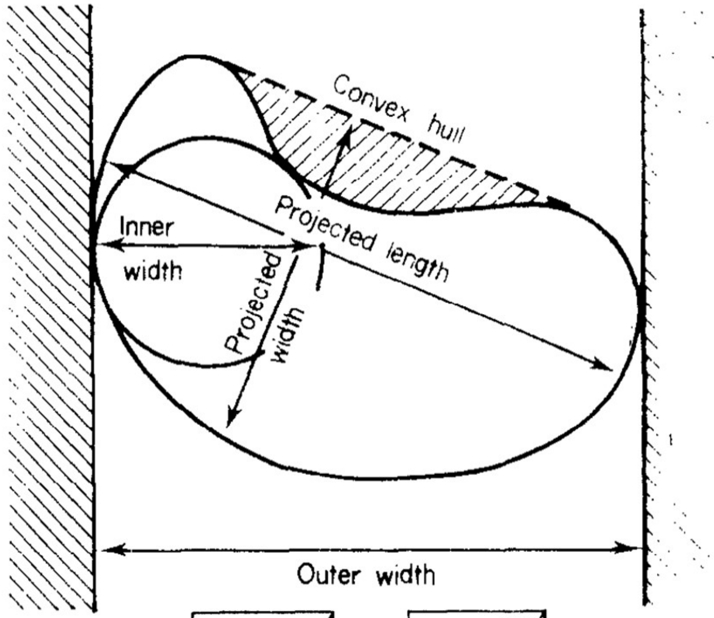
FIG. 6. The sym-transform. The sym-function consists of the locus of sym-points and their associated sym-dist. Note that the separated angles of the boundary at the pannormal feet become real angles in the parallel at this sym-point. Note also that the touching disc at each sym-point touches only at the feet of its pannormals. The remainder of the disc lies entirely in the object or ground. This property makes the sym-function a descriptor of the object that is equivalent to the boundary.

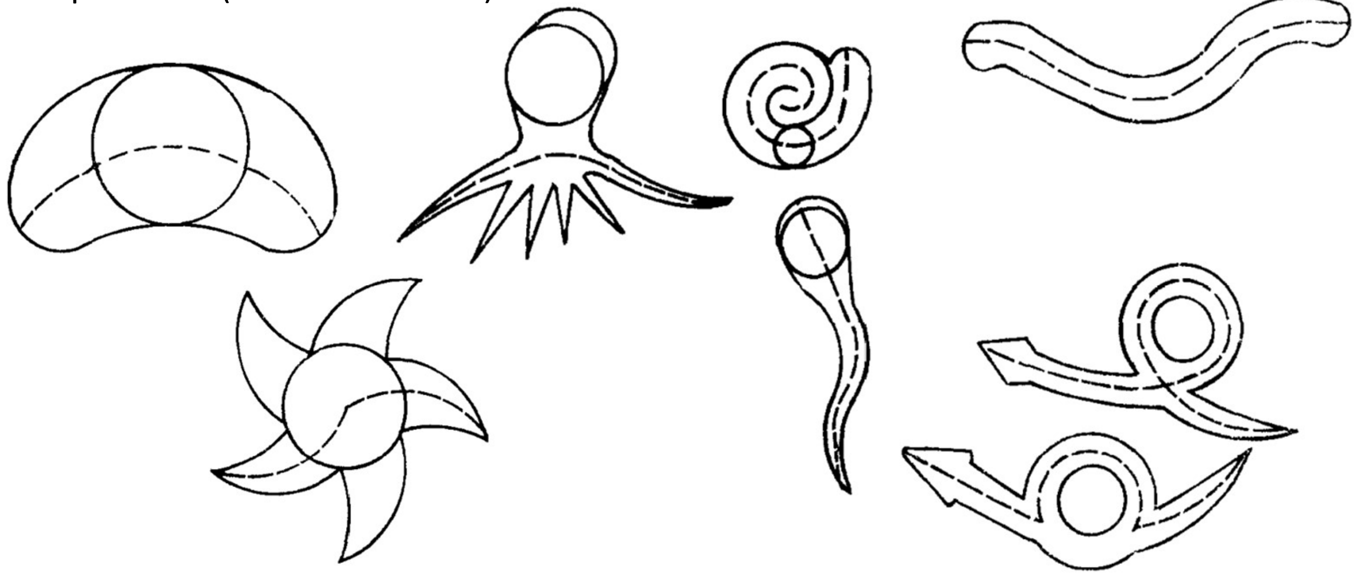
Leyton, M. (1989). Inferring Causal History froms Shape. *Cognitive Science*, *13*(3), 357-387. https://doi.org/10.1207/s15516709cog1303_2

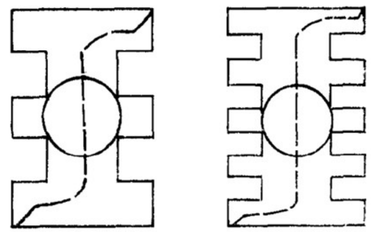
The shape skeleton: a mathematical representation of biological shapes

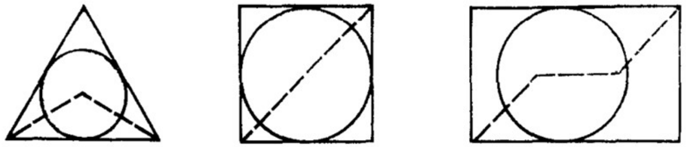
Blum, H. (1973). Biological shape and visual science (part I). Journal of Theoretical Biology, 38(2), 205-287. https://doi.org/10.1016/0022-5193(73)90175-6

One can now develop a better theory of **length** and **width.**
Projected or **outer** width (the classical definition) is usually defined, in a certain direction, as the distance between two parallel lines that touch the object on each side. **Projected length** is then the maximal width.
However, for non-convex shapes, better definitions are obtained from the shape skeleton.
Inner width is the maximal value of 2 x sym-distance along the skeleton – i.e. the diameter of the largest circle that fits inside the shape.
Length is the maximal value of the sum of skeleton path lengths + the sym-distance at ends.
These quantities (and the skeleton) are invariant to **flexion**.









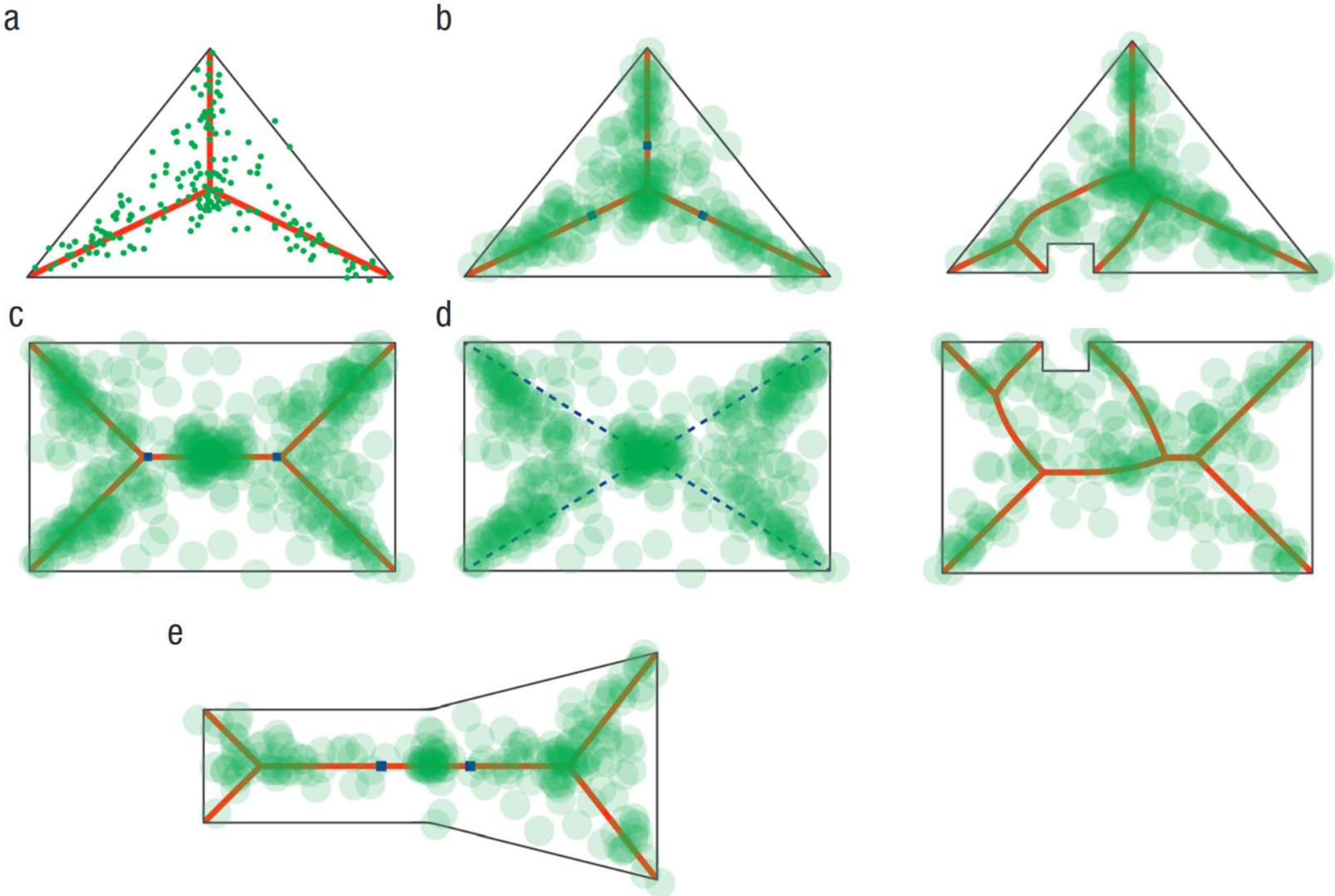
The theory works best for biological shapes, not so well for purely geometrical ones.

The shape skeleton contributes to the perception of geometric shapes

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

"The world's fastest psychology experiment": a single touch anywhere inside the shape, one per participant (n=1480 pedestrians).

Results: Taps tend to fall on the medial axes (the shape skeleton)
with a striking influence of relatively minor border perturbations.



The shape skeleton contributes to the perception of geometric shapes

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

“The world’s fastest psychology experiment”: a single touch anywhere inside the shape, one per participant (n=1480 pedestrians).

Interestingly, the results are counter-intuitive and not consciously predicted.

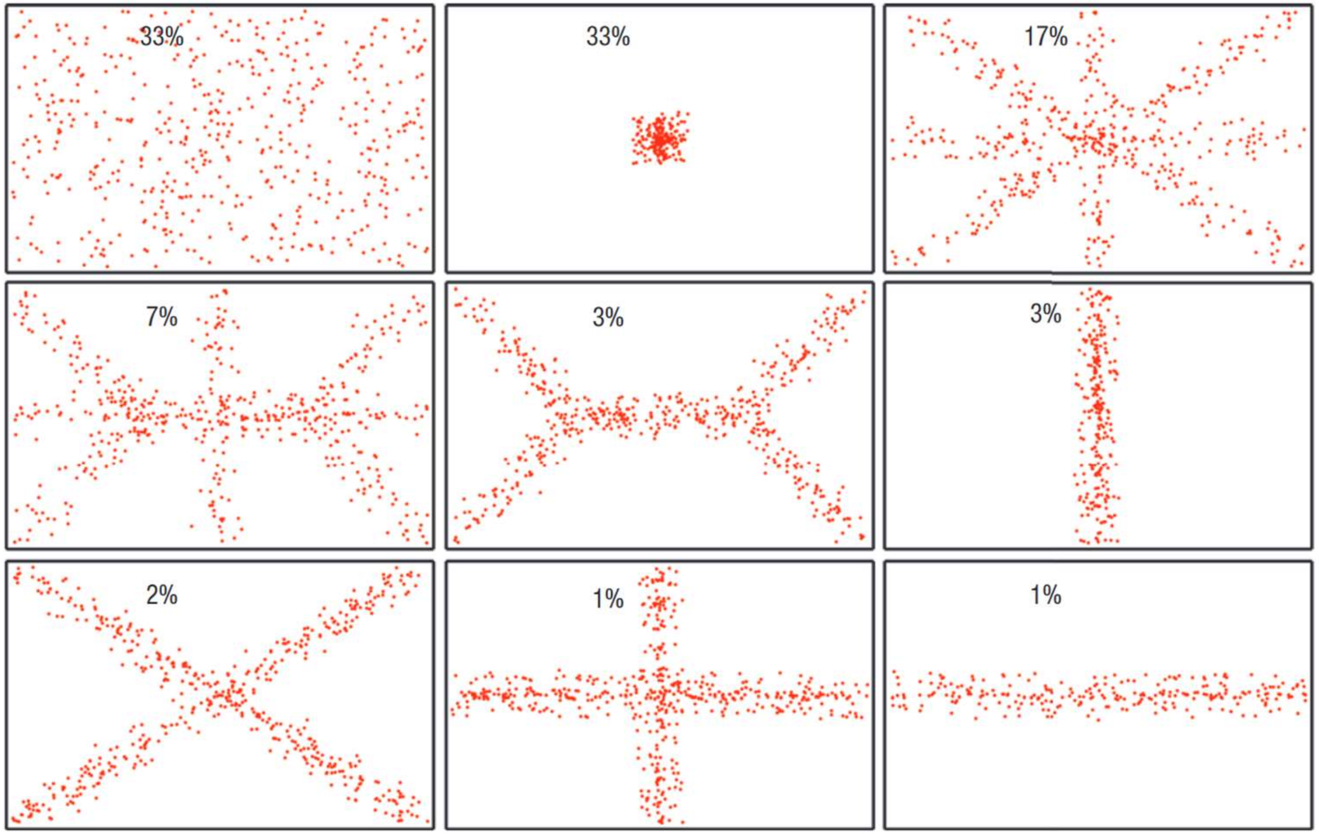
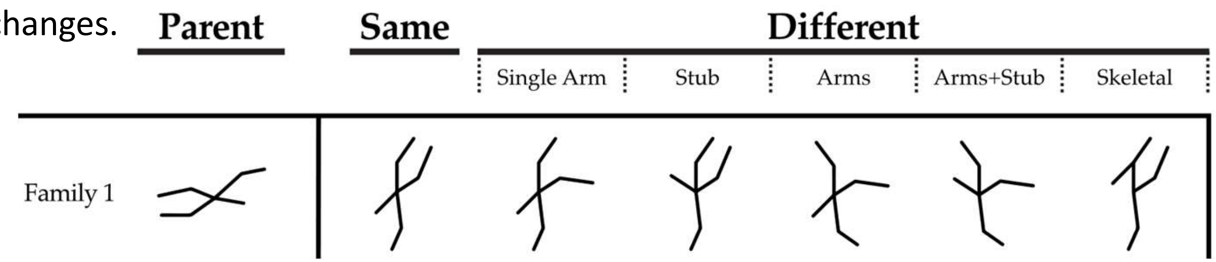


Fig. 5. The nine hypothetical distributions of touches used in Experiment 8. Naive participants selected among these distributions to indicate their prediction of the most likely outcome of Experiment 2. The number at the top of each panel (not displayed to participants) indicates the percentage of the participants who chose that option.

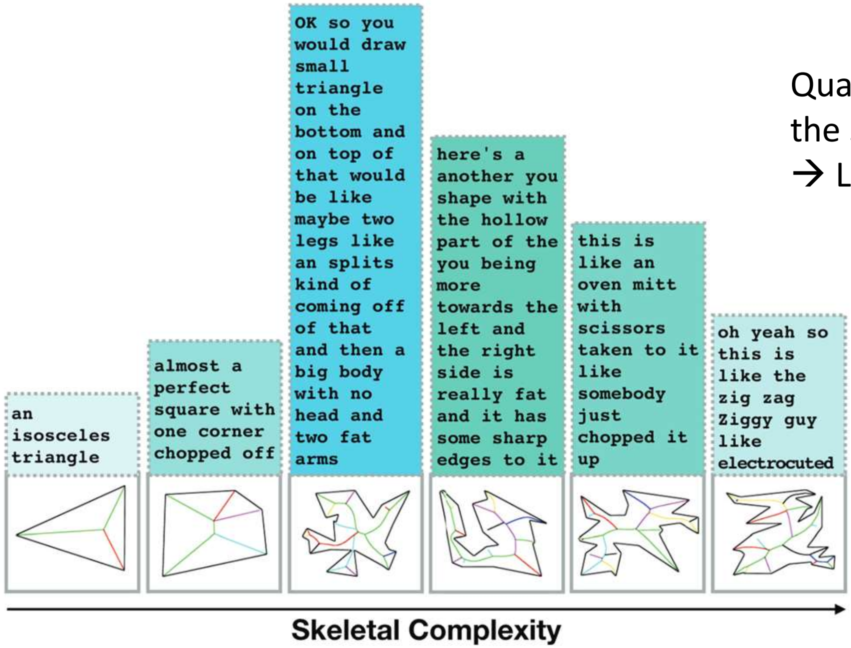
A growing set of empirical results in support of the shape skeleton

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

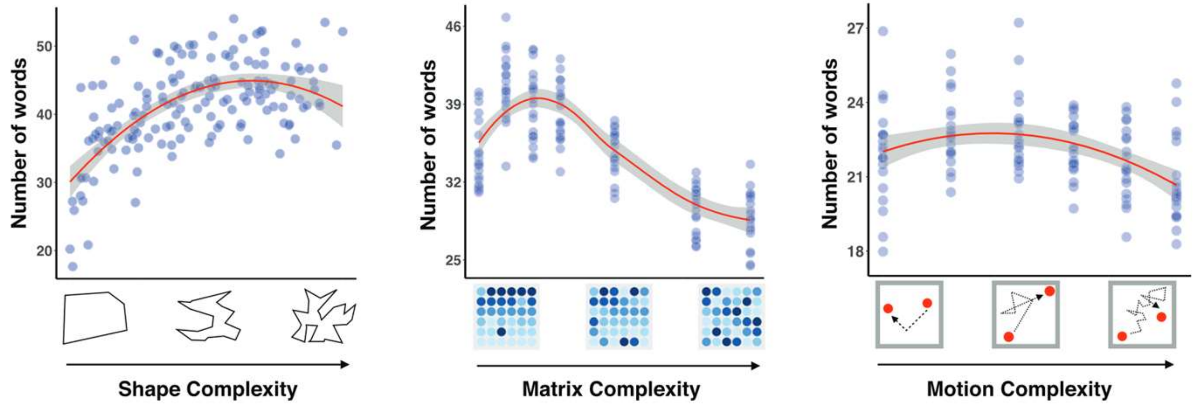
Skeletal changes are better discriminated than other shape changes.



Sun, Z., & Firestone, C. (2021). Seeing and speaking : How verbal “description length” encodes visual complexity. Journal of Experimental Psychology: General, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xge0001076



Quadratic relationship between the length of verbal descriptions and the complexity of the shape skeleton.
→ Link to minimal description length



Some infero-temporal neurons seem to encode the shape skeleton, or a part of it :
Hung, C.-C., Carlson, E. T., & Connor, C. E. (2012). Medial Axis Shape Coding in Macaque Inferotemporal Cortex. Neuron, 74(6), 1099-1113.

The importance of the shape skeleton in object perception

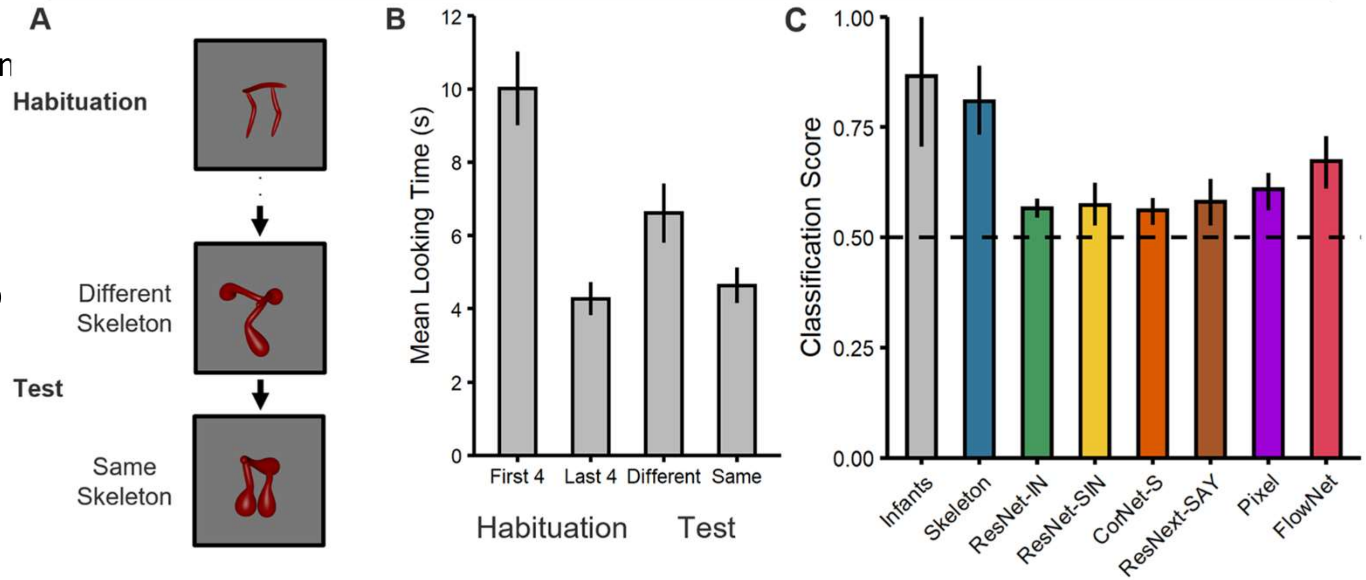
Ayzenberg, V., & Lourenco, S. (2022). Perception of an object’s global shape is best described by a model of skeletal structure in human infants. *eLife*, *11*, e74943. https://doi.org/10.7554/eLife.74943

What is the mental representation of object shape in infancy?

« Using [habituation / dishabituation], researchers have shown that newborns can already discriminate between simple 2D shapes (Slater et al., 1983) and display shape constancy, such that they recognize a shape from a novel orientation (Slater and Morison, 1985). By 6 months of age, infants’ shape representations are also robust to variations among category exemplars, such that they can categorize objects using only the stimulus’ shape silhouette (Quinn et al., 1993; Quinn et al., 2001a), as well as extend category membership to objects with varying local contours, but the same global shape (Quinn et al., 2002; Quinn et al., 2001b; Turati et al., 2003).”

Here, 6-12 month-old infants habituated to a given 3-D shape, generalized to another shape with the same skeleton (but distinct sym-dist function), but dishabituated to a shape with a topologically distinct skeleton.

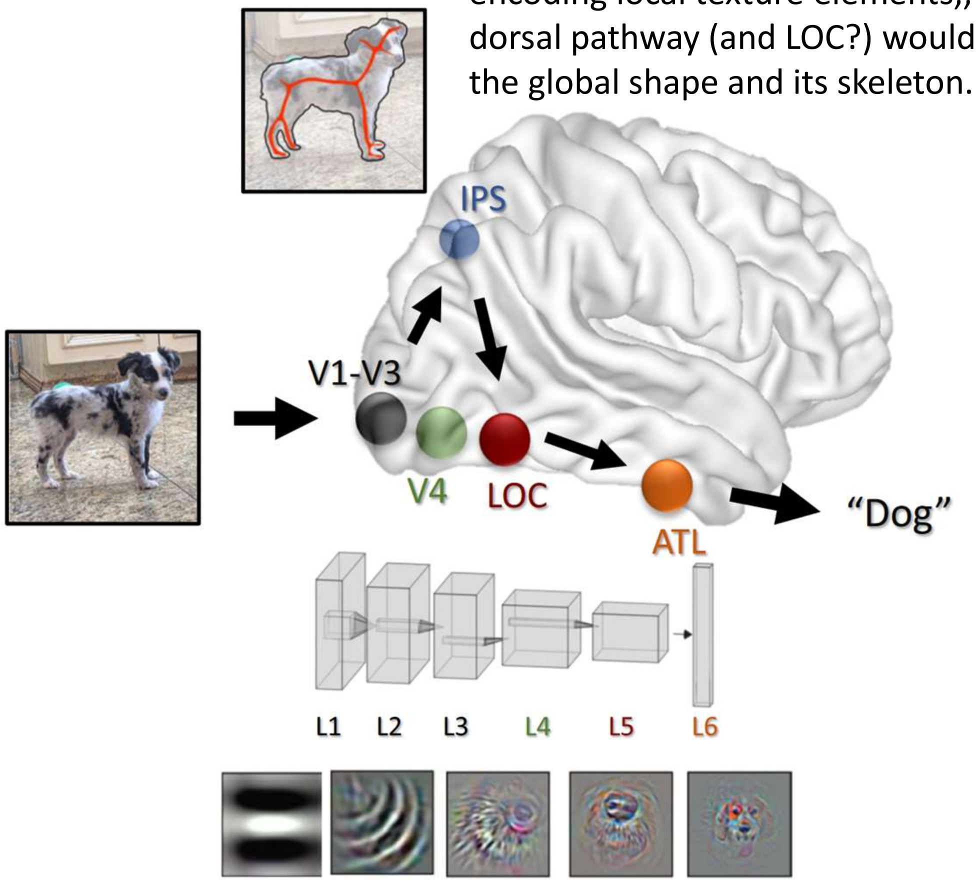
Once again, CNN models fail in this task.

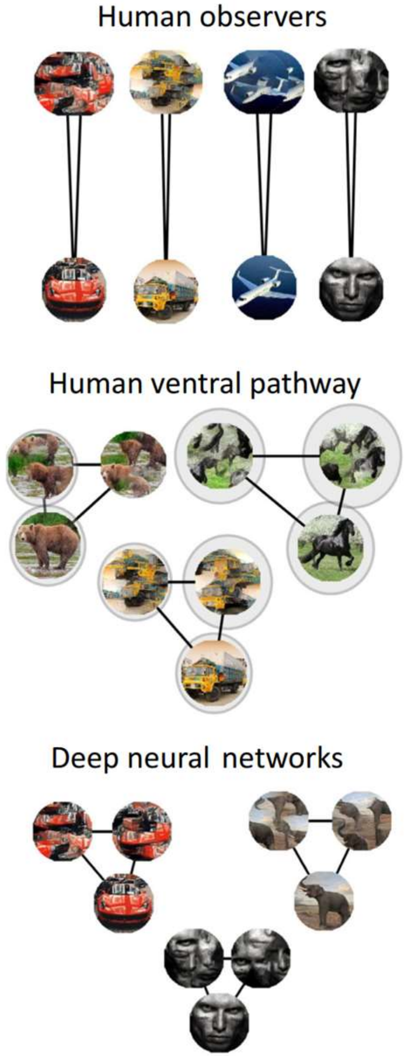


A dorsal network for global shape perception?

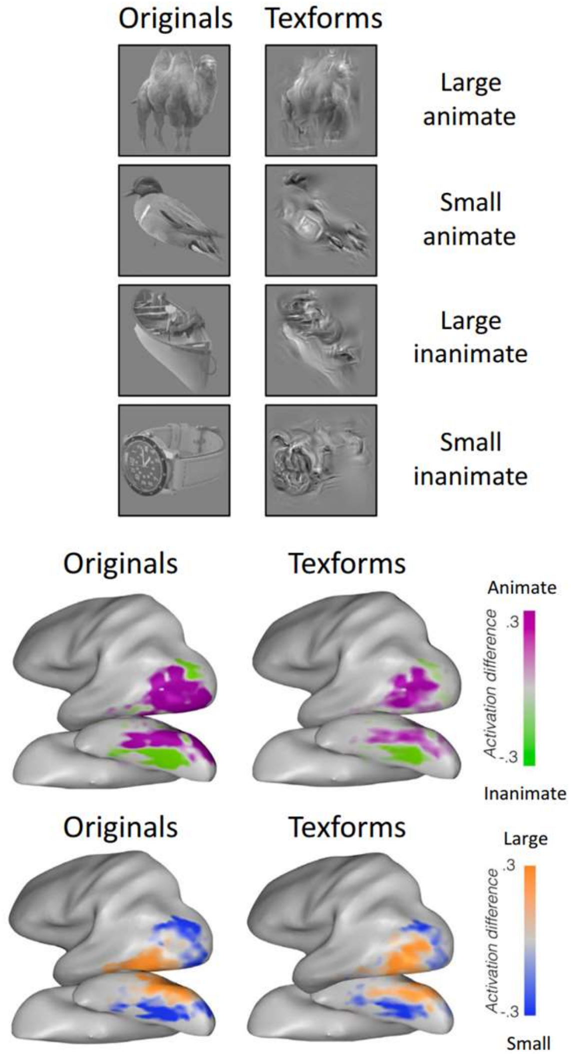
Ayzenberg, V., & Behrmann, M. (2022). Does the brain's ventral visual pathway compute object shape? *Trends in Cognitive Sciences*, *26*(12), 1119-1132. https://doi.org/10.1016/j.tics.2022.09.019

The ventral pathway may be involved in encoding local texture elements,, while the dorsal pathway (and LOC?) would encode the global shape and its skeleton.





Jagadeesh, A. V., & Gardner, J. L. (2022). Texture-like representation of objects in human visual cortex. *PNAS 119*(17), e2115302119.



Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *PNAS*, *115*(38), E9015-E9024.

Some final thoughts :
A complementarity of the shape skeleton and the ”langage of thought” approaches

- 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. https://doi.org/10.1016/j.cogpsych.2022.101527
- Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. Proceedings of the National Academy of Sciences, 103(47), 18014-18019. https://doi.org/10.1073/pnas.0608811103

Current convolutional neural networks are definitely **insufficient** to account for the human sense of shape, including geometric shapes.

Medial axis theory (skeleton) and language-of-thought theories are likely to be **complementary**.

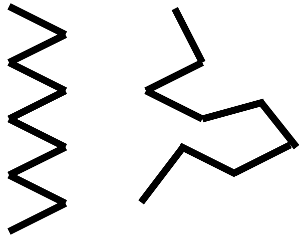
The dorsal pathway may implement **both** (1) a sense of the **axes** of objects, **and** (2) a **mathematical language** for their regularities.

1. There are many visual domains for which our language is not well suited and the skeleton-based approach is clearly superior, for instance to predict the complexity of natural shapes (animals, trees).
2. Our proposal, meanwhile, focuses entirely on the specific domain of **abstract geometric shapes**, and identifies the core tools required to account for their perception and production in humans.

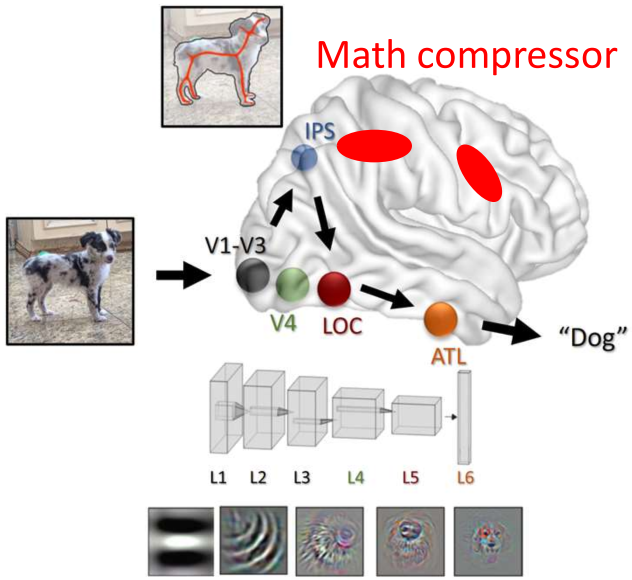
Speculation : skeleton extraction is evolutionarily ancient visual process, deeply entrenched in the visual cortex of all primates.

The language-of-geometry, however, is unique to humans.

Humans found a new way to **compress** the shape skeleton.



Same skeletal complexity for Feldman & Singh
Very different MDL for the language-of-thought



Merci de votre attention!