Appendix A

Table 1: List of all 43 concepts and 7 classes in (Dehaene et al. 2006)

Class	Concept
Topology	Holes
Topology	Inside
Topology	Closure
Topology	Connectedness
Euclidean geometry	Alignment of points in lines
Euclidean geometry	Curve
Geometrical figures	Convex Shape
Euclidean geometry	Straight line
Euclidean geometry	Alignment of points in lines
Geometrical figures	Quadrilateral
Geometrical figures	Right angled triangle
Euclidean geometry	Right angle
Euclidean geometry	Right angle
Metric properties	Distance
Geometrical figures	Circle
Metric properties	Center of circle
Metric properties	Middle of segment
Geometrical figures	Equilateral triangle
Metric properties	Fixed proportion
Metric properties	Center of quadrilateral
Geometrical figures	Square
Geometrical figures	Rectangle
Geometrical figures	Parallelogram
Geometrical figures	Trapezoid
Geometrical transformations	Vertical symmetry
Symmetrical figures	Vertical axis
Symmetrical figures	Horizontal axis
Symmetrical figures	Oblique axis
Geometrical transformations	Translation
Geometrical transformations	Point symmetry
Geometrical transformations	Horizontal symmetry
Geometrical transformations	Rotation
Geometrical transformations	Oblique symmetry
Geometrical transformations	Homothecy (fixed orientation)
Euclidean geometry	Parallel lines
Chiral figures	Oblique axis
Geometrical transformations	Homothecy (fixed size)
Euclidean geometry	Secant lines
Chiral figures	Vertical axis
Chiral figures	Vertical axis
Metric properties	Equidistance
Chiral figures	Oblique axis
Metric properties	Increasing distance
Training	Color
Training	Orientation

Appendix B

We also evaluated the CNN models against data collected from the 2-AFC version of the odd-one-out task developed by Marupudi and Varma (2023). In this version, participants are shown a target image and two alternative images, only one which embodies the same GT concept as the target image. Participants must judge which of the two alternative images is most "similar" to the target image.

We used a similar, though simpler procedure for the 2-AFC task. For a given model and a given layer, we presented the three images of the stimulus – the standard image and the correct alternative, both of which embody the GT concept, and the incorrect alternative, which does not – and recorded the three activation vectors. We then computed the cosine similarity between every pair and ranked them in decreasing order of average similarity. Correct performance was signaled when the vector corresponding to the incorrect alternative was ranked least similar (rank = 3). We again quantified performance in two ways: by the average rank of the incorrect alternative image and, more coarsely, by whether this image was correctly identified (i.e., had rank = 3) or not. Thus, the 2-AFC task can be viewed as a 3-panel version of the 6-panel odd-one-out task (Dehaene et al. 2006).

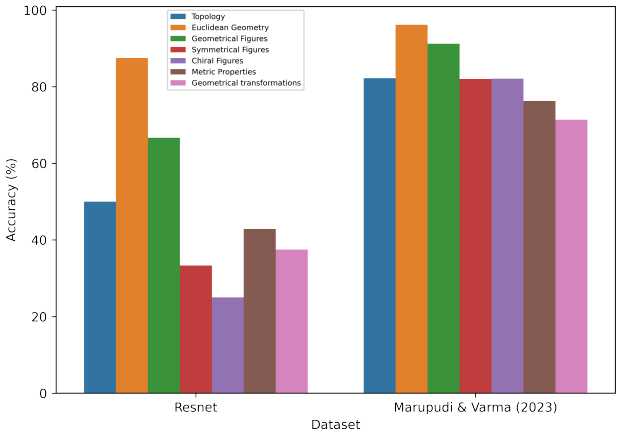

Figure 1: Absolute accuracies across the 7 classes for ResNet and the Marupudi and Varma (2023) participants on the 2-AFC task. The Pearson correlation is 0.79.

Table 2 summarizes the performance of the CNN models, again as measured on the finally, fully-connected layer and aggregated to the level of the 7 classes of GT concepts. ResNet is still the best performing model (accuracy = 53.49%). Recall that chance in the 2-AFC task is a much higher bar for participants to exceed to demonstrate sensitivity than in the odd-one-out task (50% vs. 16.67%). Interestingly, GoogLeNet (avg. rank = 2.26) now achieves better performance than ResNest (avg. rank = 2.2) on the finer-grain average rank measure, assigning the higher average rank to the incorrect alternative image.

Figure 1 presents the profile analysis for ResNet and for the American adults who completed the 2-AFC task of Marupudi and Varma (2023). The two profiles are highly correlated ($r = .79$), indicating that the model finds the same classes of GT concepts to be relatively easy and relatively difficult as humans.

Table 2: Overall performance of the CNN models on the 2-AFC task. Note that all the entries that differ significantly from the chance level of 16.67% (i.e., where p ¡ 0.05 on a binomial test) have been marked with an "*". The results of the overall best performing model, Resnet, are shown in bold.

Model	Topology	Euclidean Geometry	Geometrical Figures	Symmetrical Figures	Chiral Figures	Metric Properties	Geometrical Transformations	Overall Accuracy	Avg. Rank
VGG-19	50.0	87.5*	44.44	0.0	25.0	57.14	25.0	46.51*	2.13
AlexNet	75.0	75.0*	44.44	33.33	25.0	14.29	50.0	46.51*	2.12
ResNet	**50.0**	**87.5***	**66.67***	**33.33**	**25.0**	**42.86**	**37.5**	**53.49***	**2.2**
DenseNet	50.0	62.5	55.56	33.33	25.0	57.14	37.5	48.84*	2.15
GoogLeNet	50.0	75.0*	55.56	0.0	50.0	57.14	25.0	48.84*	2.26

References

Dehaene, S.; Izard, V.; Pica, P.; and Spelke, E. S. 2006. Core Knowledge of Geometry in an Amazonian Indigene Group.

Marupudi, V.; and Varma, S. 2023. Graded Human Sensitivity to Geometric and Topological Concepts. *Cognition*, 232: 105331.