



**HAL**  
open science

## Visual object categorization in infancy

Céline Spriet, Etienne Abassi, Jean-Rémy Hochmann, Liuba Papeo

► **To cite this version:**

Céline Spriet, Etienne Abassi, Jean-Rémy Hochmann, Liuba Papeo. Visual object categorization in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 2022, 119 (8), 10.1073/pnas.2105866119 . halshs-03929941

**HAL Id: halshs-03929941**

**<https://shs.hal.science/halshs-03929941>**

Submitted on 9 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Visual object categorization in infancy

Céline Spriet<sup>a</sup>, Etienne Abassi<sup>a</sup>, Jean-Rémy Hochmann<sup>a</sup>, and Liuba Papeo<sup>a,1</sup>

<sup>a</sup>Institute of Cognitive Sciences Marc Jeannerod, UMR5229, CNRS, University Claude Bernard Lyon1, Bron 69675, France

Edited by Gergely Csibra, Central European University, Budapest, Hungary; received March 26, 2021; accepted December 30, 2021 by Editorial Board Member Renée Baillargeon

**Humans make sense of the world by organizing things into categories. When and how does this process begin? We investigated whether real-world object categories that spontaneously emerge in the first months of life match categorical representations of objects in the human visual cortex. Using eye tracking, we measured the differential looking time of 4-, 10-, and 19-mo-olds as they looked at pairs of pictures belonging to eight animate or inanimate categories (human/nonhuman, faces/bodies, real-world size big/small, natural/artificial). Taking infants' looking times as a measure of similarity, for each age group, we defined a representational space where each object was defined in relation to others of the same or of a different category. This space was compared with hypothesis-based and functional MRI-based models of visual object categorization in the adults' visual cortex. Analyses across different age groups showed that, as infants grow older, their looking behavior matches neural representations in ever-larger portions of the adult visual cortex, suggesting progressive recruitment and integration of more and more feature spaces distributed over the visual cortex. Moreover, the results characterize infants' visual categorization as an incremental process with two milestones. Between 4 and 10 mo, visual exploration guided by saliency gives way to an organization according to the animate–inanimate distinction. Between 10 and 19 mo, a category spurt leads toward a mature organization. We propose that these changes underlie the coupling between seeing and thinking in the developing mind.**

cognitive development | categorization | visual system | looking times | fMRI

Objects are the units of attention and perception; categories are the units of thought. We see objects (e.g., a rounded spongy red and white-dotted shape on an elongated support), but we think about objects primarily in terms of categories (e.g., the mushroom *Amanita muscaria*). By recognizing an object as member of a category, we understand what that object is and retrieve its visible (e.g., it is red with white spots) as well as its invisible properties (e.g., it is hallucinogenic). Categorization is thus the basis of inference and decision, although not all inferences and decisions require categorization.

Objects can be categorized according to a virtually infinite number of perceptual and nonperceptual dimensions (1, 2). Insight on the most basic and general dimensions for object categorization in humans has been gained by studying how information is organized in the vast brain territory for visual object representation, which forms the occipitotemporal visual ventral stream.

Here, categories emerge from the topography of responses to visual objects, resolving into a large-scale organization that distinguishes between animate and inanimate objects, and crumbles in finer-grained distinctions between human vs. nonhuman animals, small vs. big (in terms of real-world size) (3, 4), and natural vs. artificial objects (3, 5–11). Underneath this organization lies a mosaic of local hot spots of strong selectivity for stimuli, such as faces, bodies, and scenes (12–15). Because of its organization and role in object recognition, the visual ventral stream is regarded as the interface between perception and cognition, forming the backbone for semantic categorization

and representation of object and action knowledge in the rest of the brain (16).

Besides the topography, categorical distinctions in the visual cortex also emerge from dissimilarity between distributed patterns of neural activity evoked by individual objects (7, 17, 18). Thus, in visual areas, activity patterns recorded with functional MRI (fMRI) are more similar (i.e., less discriminable) for two animate objects (e.g., parrot and camel) than between an animate and an inanimate object (e.g., parrot and car). Visual object categories represented in the visual cortex prove behaviorally relevant, predicting the way in which individuals parse the visual world. For example, in a visual search for a target–object among a set of distractors, people are faster to discriminate and find a target among objects of a different visual category (e.g., a cat among artificial objects) than among objects of the same visual category: search times increase as neural similarity between target and distractors increases (5).

The organization of the human visual cortex by object categories appears to be a hallmark in the evolution of the primate brain: it is replicated in the visual cortex of monkeys (7, 19) and is resistant to variations of individual visual experience (20–23). A similar organization across species and conspecifics with different environment and life-long visual experience suggests a neural code optimized by evolution. This line of thinking encourages the hypothesis that object representation in the visual cortex reflects biological constraints and dispositions (24); as such, it would emerge early in life or even be present at birth.

## Significance

**Categorization is the basis of thinking and reasoning. Through the analysis of infants' gaze, we describe the trajectory through which visual object representations in infancy incrementally match categorical object representations as mapped onto adults' visual cortex. Using a methodological approach that allows for a comparison of findings obtained with behavioral and brain measures in infants and adults, we identify the transition from visual exploration guided by perceptual salience to an organization of objects by categories, which begins with the animate–inanimate distinction in the first months of life and continues with a spurt of biologically relevant categories (human bodies, nonhuman bodies, nonhuman faces, small natural objects) through the second year of life.**

Author contributions: C.S., J.-R.H., and L.P. designed research; C.S. and E.A. performed research; C.S., E.A., J.-R.H., and L.P. analyzed data; and C.S., J.-R.H., and L.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.C. is a guest editor invited by the Editorial Board.

This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [liuba.papeo@isc.cnrs.fr](mailto:liuba.papeo@isc.cnrs.fr).

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2105866119/-/DCSupplemental>.

Published February 15, 2022.

There is initial evidence for signatures, or precursors, of neural specialization to object categories (faces, bodies, animals, and scenes) in the visual cortex of newborns or young infants, based on electroencephalography (25–30) or fMRI (31, 32). Behavioral counterparts of those neural effects include early preference for faces or face-like stimuli over inverted faces (33–35), for biological over nonbiological motion (36, 37), and for canonical over distorted bodies (38–40).

While preference implies discrimination between two objects, visual categorization entails the ability to use the visual properties of a category (e.g., shape) to identify its members and keep them separate from other categories. By 4 mo, infants are already able to do so: exposed to various exemplars of a category (e.g., cats), they exhibit a novelty effect, looking longer at an object of a new category than at a novel object of the same category (41–43).

But when do infants begin to see the visual world as adults do? Here, we investigate whether the categorical dimensions that drive the large-scale organization of the human visual cortex could account for the spontaneous emergence and development of real-world object categories in infancy. In particular, under the hypothesis that the structuring of visual object information toward an adult-like organization begins at birth (27, 31, 32), we asked when such organization becomes functional so as to account for how infants explore the visual world.

We examined the development of visual object categorization in infancy, considering, in one experimental design, objects that have highlighted categorical representations in the visual cortex of human adults (and monkeys): animate vs. inanimate, human vs. nonhuman (animate), faces vs. bodies, natural vs. artificial inanimate, and real-world big vs. small (inanimate) (7). Each of the above distinctions defines a categorization model, whereby a given (behavioral or physiological) correlate of object perception would be more similar for two objects of the same category than for two objects of different categories.

Using eye tracking, we recorded the most reliable and informative measure of infants' cognition thus far: the looking behavior (44, 45). Infants of 4, 10, and 19 mo viewed two objects at a time on a screen, while we measured the looking time toward either object. We took the looking time difference between two stimuli as a measure of dissimilarity, under the assumption that looking times for two objects seen for the first time would be more similar, the closer their visual representation is (see also ref. 46). Since two stimuli of the same visual category are, normally, more similar than two stimuli from different categories, we expected the variations in differential looking times (DLTs) to reflect variations in representational similarity, uncovering categorical distinctions.

In classic categorization studies, infants' looking times are used to capture differences in novelty/familiarity, created ad hoc within the experimental session [e.g., through the presentation of multiple exemplars of a category during familiarization (41–43, 47, 48)]. Thus, a methodological challenge (and innovation) of the current work was to use looking times to capture differences in the perceived (dis)similarity between two objects, in the absence of any controlled unbalance in the exposure to a given category (at least within the experimental session). As a result, this approach defined a model where each object was represented in relation to the others (i.e., how similar/dissimilar it was from exemplars of the same and different categories). A model based on a relative measurement can be quantitatively compared with any model based on another relative measurement, whatever the source of the measurements (e.g., reaction times, neural activity) (49). We compared the model of visual object representation emerging from the infants' looking behavior, with synthetic (i.e., hypothesis-driven) and data-driven (i.e., fMRI-based) models reflecting visual object representation in the mature visual cortex. This approach had previously allowed

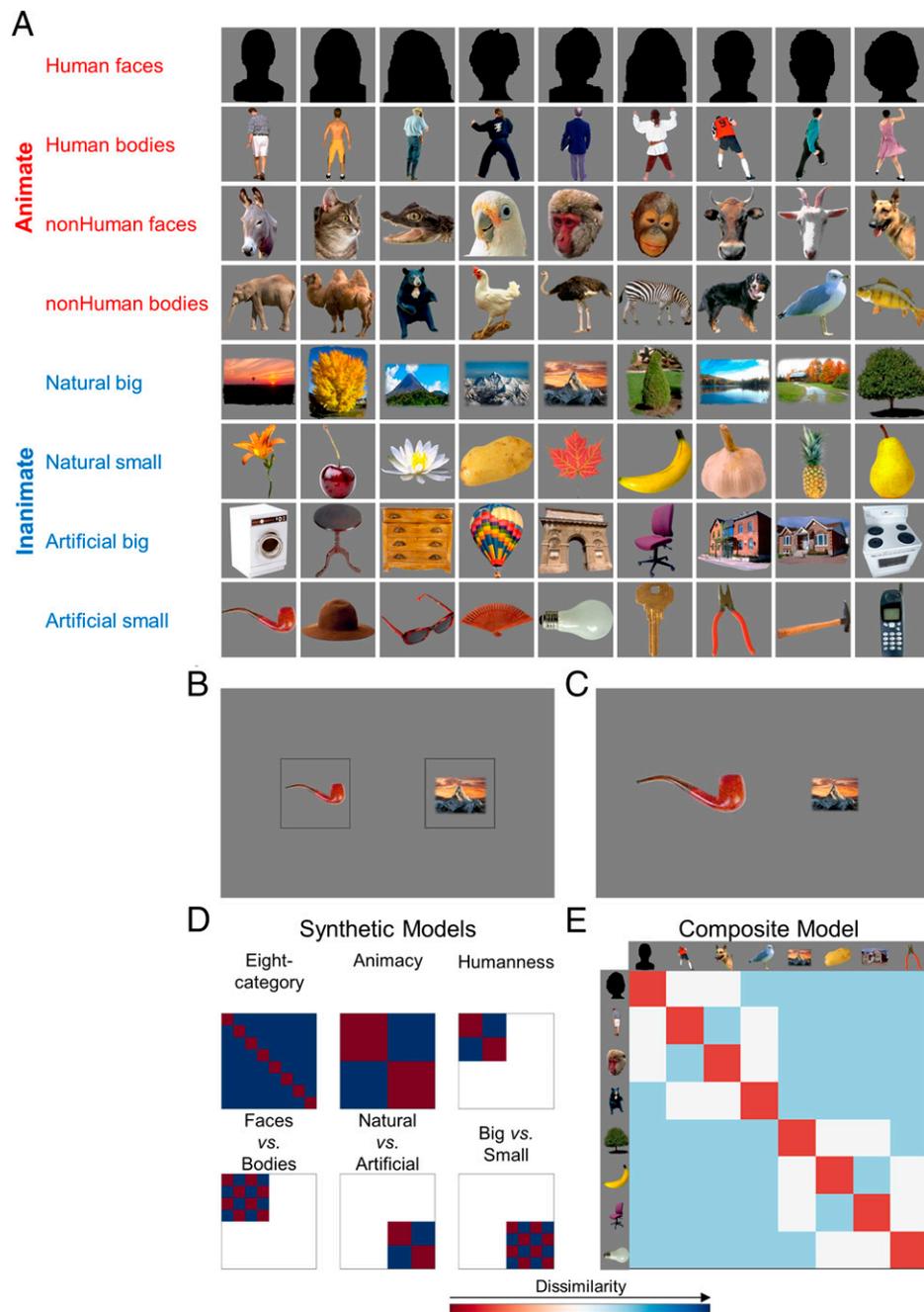
connecting data from brain-activity recording, behavioral measurements in adults, and computational modeling (49). Here, by studying the relationship between the infants' looking behavior and the organization of visual object information in the adults' brain, we connected another branch, which is another step toward a unified theory of the origin and development of functional organization in the human brain.

## Results

**Experiment 1.** Three groups of infants of 4 ( $n = 24$ ), 10 ( $n = 24$ ), and 19 mo ( $n = 25$ ) saw 36 pairs of images, each featuring an object from one of eight categories: human faces, human bodies, nonhuman faces, nonhuman bodies, natural-big and natural-small objects, artificial-big and artificial-small objects (hereafter, “big” and “small” refer to real-world size) (Fig. 1A). The set of images used here had previously been used to study object representation in the visual cortex (49, 50). They depicted naturalistic views of real-world objects without meeting any specific requirement or manipulation. Objects thus reproduced the natural combination of visual features typical of their category, so that we could expect the variations in the objects' appearance to reflect the natural variations within and between categories. All subjects saw all possible combinations (Fig. 1B) of between-category and within-category pairs. For each infant, for each pair, we measured the absolute difference in looking times between the left and right images (DLT) (*Materials and Methods*). DLTs were used to build a representational dissimilarity matrix (RDM), in which cells off the diagonal represented between-category comparisons and cells on the diagonal represented within-category comparisons (Fig. 2). Since different infants saw different exemplars for each category, group-averaged RDMs represented relationships (i.e., dissimilarities) between categories, rather than between individual objects.

**Reference models of visual object categorization.** Using representational similarity analysis (49), we computed the relationship between RDMs based on infants' DLT (DLT-RDMs) and models (i.e., RDMs) of visual object categorization in adults, defined with two independent approaches. The first approach defined a set of categorization models based on fMRI responses evoked in human adults, when viewing the same objects presented to infants. In the fMRI-based RDMs, pairwise between- and within-category dissimilarities reflected correlations between neural activity patterns (*Materials and Methods*). Three RDMs were computed from activations in three broad regions-of-interest (ROIs) (Fig. 3A and B) of the visual cortex (early visual cortex, EVC; ventral occipitotemporal cortex, VOTC; and lateral occipitotemporal cortex, LOTC), and at each location along the antero-posterior axis of the visual ventral stream (i.e., vector-of-ROIs analysis). The second approach defined six synthetic categorization models (RDMs) that may apply to the current stimulus set: animate–inanimate (animacy model), human–nonhuman animates (humanness model), faces–bodies, natural–artificial inanimates, big–small inanimates, and eight-category model, where each category was defined as a category of its own, distinct from the other seven (Fig. 1D). In each cell of an RDM, the values 0 or 1 indicated dissimilarities within-category (lowest dissimilarity) and between-category (highest dissimilarity), respectively. As a model of visual object categorization in adults, a composite-RDM was obtained by averaging the above six models (Fig. 1E).

In addition to, or instead of, categorical information, infants' look might be guided by physical properties of the stimuli, such as size of the image on the retina (51, 52), elongation (53), compactness (54), and color (53), among others. To assess systematic relations between looking times and visual features of the images, irrespective of the category, we computed RDMs

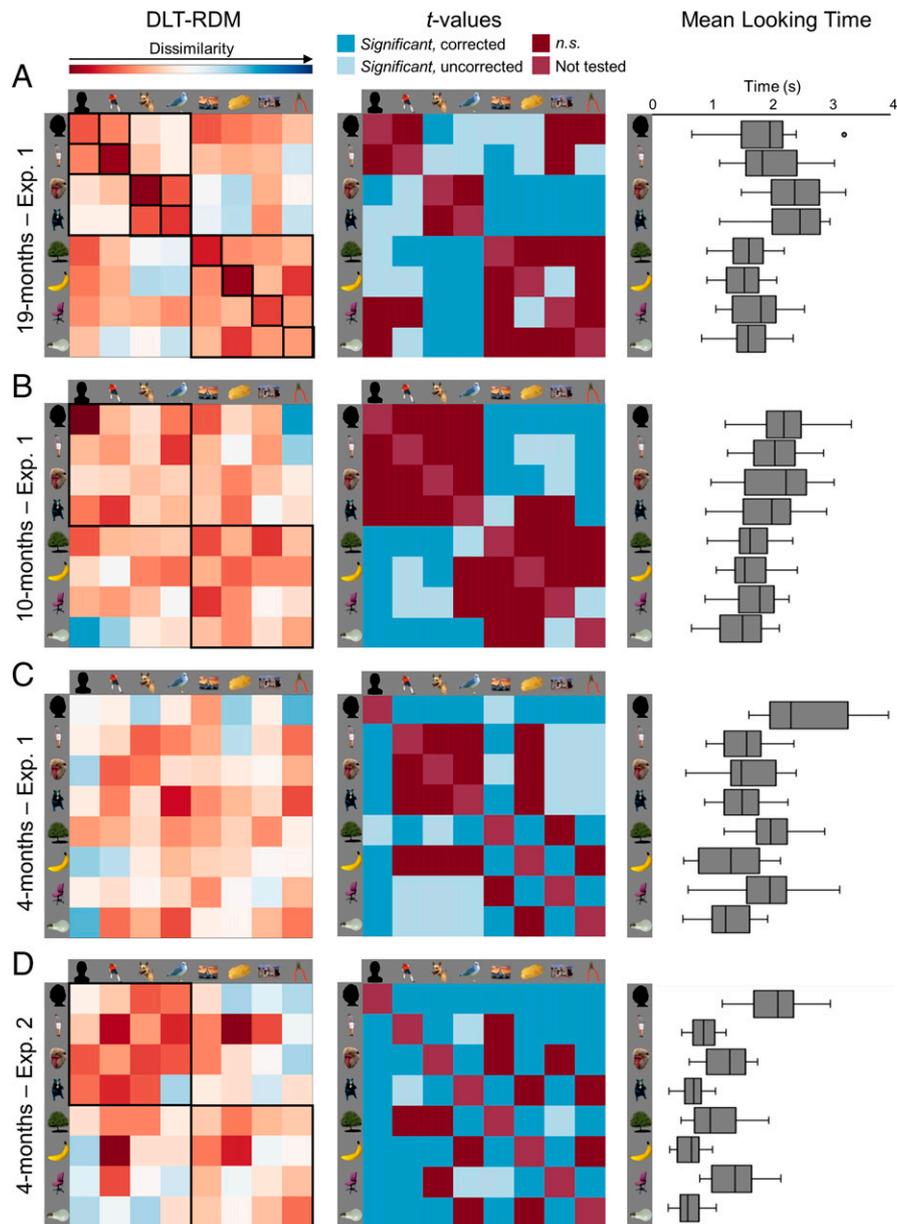


**Fig. 1.** Stimuli, trials, and hypothesis-based models of categorization considered in the design of Exps. 1 and 2. (A) Stimuli were 72 images depicting 9 objects from each of 8 different categories. Silhouettes instead of the actual colorful female human faces used in the experiments, are shown for illustration. (B) In each trial of Exp. 1, two images were presented within two gray frames of identical size, on the right and on the left, equally distant from the center of the screen. (C) In each trial of Exp. 2, the image frame was removed and the image size was modified so that each object had the same number of pixels. (D) Hypothesis-driven (synthetic) models reflecting the categorical object representations tested in the current design. (E) The composite model reflecting the mean of the six synthetic models.

representing differences in size, elongation, and compactness, relying on signed values to appreciate the looking-time difference between two objects but also which one was looked at the longer (the larger one; the more/less elongated one; the more/less compact one). A fourth RDM was computed to represent differences in the image color (*SI Appendix, Fig. S1*). Note that other image properties may affect infants' behavior; here, size, elongation, compactness, and color should be taken as proxies of unspecific physical properties of our stimuli that are not necessarily distinctive of a category, but could affect the looking behavior.

**Nineteen-month-olds.** The group-averaged DLT-RDM (Fig. 2A) showed an adult-like organization, as reflected by significant correlations with the composite-RDM, and the RDMs derived from the EVC, VOTC, and LOTC (Fig. 3C; see statistics in Table 1). The vector-of-ROIs analysis showed that the DLT-RDM was maximally correlated with object-related responses in early visual areas (V1 to V3) and fusiform gyrus ( $P_s < 0.001$ ) (Fig. 3D).

Next, we asked which of the six categorical models underlying the composite-RDM, best represented the infants' DLT-RDMs. A stepwise linear regression ( $\alpha_{\text{corrected}}: 0.0083$ , two-tailed) showed



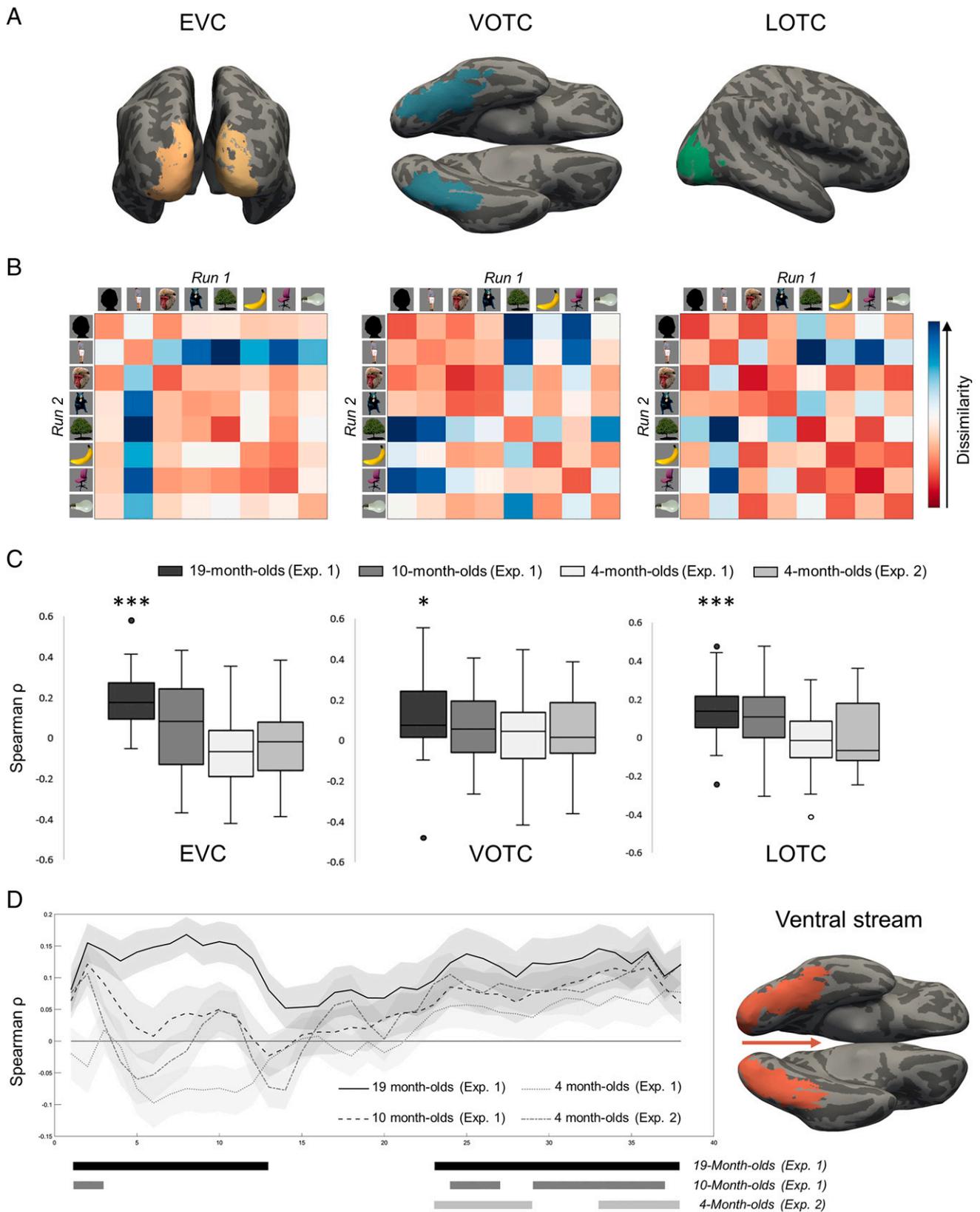
**Fig. 2.** Results of representational similarity analysis and of the pairwise comparisons of MLT between- and within-categories for each age group in Exps. 1 and 2. (*Left*) Mean RDM reflecting dissimilarities between- and within-categories in terms of DLTs. Black squares in the RDMs highlight categorization by animacy, humanness, and by the eight categories in 19-mo-olds (A), categorization by animacy in 10-mo-olds (B), and in 4-mo-olds of Exp. 2 (D). (*Center*) Matrix of  $t$ -values for each pairwise comparison between MLTs of the individual categories for 19- (A), 10- (B), and 4-mo-olds (C) in Exp. 1 and 4-mo-olds in Exp. 2 (D). Squares in dark blue denote significant effects; squares in lighter blue denote effects that did not survive the multiple comparison correction (trends); red squares denote nonsignificant (n.s.) or nontested comparisons. (*Right*) Distribution of MLTs in 19- (A), 10- (B), and 4-mo-olds (C) of Exp. 1 and of 4-mo-olds of Exp. 2 (D). Box-plots represent the minimum, the first quartile, the median, the third quartile and the maximum of the population distribution; outliers are denoted by dots (one in the 19-mo-old group).

an effect of the eight-category model [mean  $\beta = 0.090$ ; 99.17% confidence interval (CI) [100  $\times$  (1-0.0083)] = 0.026 to 0.155;  $t(24) = 4.023$ ,  $P < 0.001$ ;  $d = 0.804$ ], the animacy model [mean  $\beta = 0.077$ ; 99.17% CI = 0.014 to 0.139;  $t(24) = 3.514$ ,  $P = 0.002$ ;  $d = 0.702$ ], and the humanness model [mean  $\beta = 0.133$ ; 99.17% CI = 0.009 to 0.256;  $t(24) = 3.091$ ,  $P = 0.005$ ;  $d = 0.618$ ] (for all other regressors:  $P_s > 0.07$ ) (SI Appendix, Table S1).

In another analysis, we considered another measure of categorization: for the six categorization models, we tested whether average between-category DLTs were higher than average within-category DLTs ( $\alpha_{\text{corrected}} = 0.0083$ , one-tailed). Confirming the results above, we found that this was the case for the eight-

category model [mean difference =  $-0.135$ ; 99.17% CI =  $-\infty$  to  $-0.091$ ;  $t(24) = -7.967$ ,  $P < 0.0001$ ;  $d = 1.588$ ], the animacy model (mean difference =  $-0.095$ ; 99.17% CI =  $-\infty$  to  $-0.041$ ;  $t(24) = -4.481$ ,  $P < 0.0001$ ;  $d = 0.896$ ) and the humanness model [mean difference =  $-0.158$ ; 99.17% CI =  $-\infty$  to  $-0.060$ ;  $t(24) = -4.160$ ,  $P < 0.001$ ;  $d = 0.832$ ], but not for the other models (SI Appendix, Table S2).

Following evidence of categorization based on the eight-category model, we asked which of the eight categories the infants could indeed represent. Separately for each of the eight categories, we tested whether average within-category DLTs were lower than average between-category DLTs ( $t$  tests;



**Fig. 3.** Relationship between infants' looking behavior and visual object representation in the adults' visual cortex. (A) ROIs in the adults' brain: EVC, VOTC, and LOTC. (B) Mean RDMs reflecting relationships (i.e., dissimilarities) between object categories in terms of dissimilarities in the neural activity patterns evoked by viewing objects in the EVC, VOTC, and LOTC of adults (fMRI-based RDMs). (C) Results of the representational similarity analysis between the mean fMRI-based RDM in each ROI and the DLT-RDM of each infant in each age group of Exps. 1 and 2. Box-plots represent the minimum, the first quartile, the median, the third quartile, and the maximum of the population distribution as well as outliers (dots); \* $P < 0.017$ ; \*\*\* $P < 0.0003$ . (D) Results of the representational similarity analysis between the infants' DLT-RDMs and the fMRI-based RDM derived from each partition along the ventral visual stream. Solid bars represent clusters with significant correlation (above 0) for each age group of Exps. 1 and 2.

**Table 1. Results of the representational similarity analysis reflecting the relationship of the infants' DLT-RDMs with the fMRI-based RDMs, the RDM for the synthetic composite model of categorization, and the RDMs based on size, elongation, compactness, and color differences**

Exp.	Age (mo)	Model	Mean $\rho$ (SD)	CI (minimum to maximum)	$t$ (df)	$P$	Cohen's $d$		
1	19	<b>EVC</b>	<b>0.182 (0.123)</b>	<b>0.123 to 0.273</b>	<b>6.795 (24)</b>	<b>&lt;0.0001</b>	<b>1.474</b>		
		<b>VOTC</b>	<b>0.110 (0.200)</b>	<b>0.017 to 0.227</b>	<b>2.985 (24)</b>	<b>0.006</b>	<b>0.552</b>		
		<b>LOTC</b>	<b>0.133 (0.147)</b>	<b>0.065 to 0.226</b>	<b>4.615 (24)</b>	<b>&lt;0.001</b>	<b>0.902</b>		
		<b>CM</b>	<b>0.248 (0.195)</b>	<b>0.168 to 0.328</b>	<b>6.371 (24)</b>	<b>&lt;0.0001</b>	<b>1.272</b>		
		Size	-0.034 (0.194)	-0.139 to 0.071	-0.874 (24)	n.s.	0.175		
		Elongation	-0.010 (0.151)	-0.092 to 0.071	-0.333 (24)	n.s.	0.066		
		Compactness	-0.074 (0.166)	-0.163 to 0.016	-2.218 (24)	0.036	0.444		
		Color	0.009 (0.186)	-0.092 to 0.109	0.234 (24)	n.s.	0.048		
		1	10	EVC	0.053 (0.252)	-0.079 to 0.185	1.029 (23)	0.314	0.210
				VOTC	0.077 (0.179)	-0.017 to 0.171	2.112 (23)	0.046	0.431
LOTC	0.086 (0.200)			-0.019 to 0.191	2.102 (23)	0.047	0.429		
<b>CM</b>	<b>0.125 (0.205)</b>			<b>0.038 to 0.211</b>	<b>2.984 (23)</b>	<b>0.007</b>	<b>0.610</b>		
Size	0.068 (0.208)			-0.048 to 0.183	1.592 (23)	0.125	0.327		
Elongation	-0.041 (0.197)			-0.150 to 0.068	-1.013 (23)	n.s.	0.208		
Compactness	0.054 (0.190)			-0.052 to 0.159	1.379 (23)	0.181	0.281		
Color	0.044 (0.253)			-0.096 to 0.183	0.843 (23)	n.s.	0.174		
1	4			EVC	-0.067 (0.201)	-0.172 to 0.039	-1.629 (23)	0.117	0.333
				VOTC	0.033 (0.174)	-0.058 to 0.124	0.929 (23)	n.s.	0.190
		LOTC	-0.025 (0.184)	-0.122 to 0.071	-0.673 (23)	n.s.	0.137		
		CM	0.049 (0.182)	-0.028 to 0.126	1.320 (23)	0.200	0.269		
		<b>Size</b>	<b>0.315 (0.173)</b>	<b>0.220 to 0.411</b>	<b>8.950 (23)</b>	<b>&lt;0.0001</b>	<b>1.821</b>		
		<b>Elongation</b>	<b>-0.321 (0.187)</b>	<b>-0.425 to -0.218</b>	<b>-8.401 (23)</b>	<b>&lt;0.0001</b>	<b>1.717</b>		
		<b>Compactness</b>	<b>0.208 (0.155)</b>	<b>0.123 to 0.294</b>	<b>6.597 (23)</b>	<b>&lt;0.0001</b>	<b>1.347</b>		
		Color	0.035 (0.204)	-0.078 to 0.148	0.836 (23)	n.s.	0.172		
		2	4 m	EVC	-0.026 (0.199)	-0.131 to 0.079	-0.642 (23)	n.s.	0.131
				VOTC	0.051 (0.191)	-0.050 to 0.151	1.295 (23)	0.208	0.264
LOTC	0.017 (0.184)			-0.080 to 0.114	0.453 (23)	n.s.	0.092		
<b>CM</b>	<b>0.102 (0.205)</b>			<b>0.015 to 0.189</b>	<b>2.433 (23)</b>	<b>0.023</b>	<b>0.498</b>		
<b>Elongation</b>	<b>-0.150 (0.187)</b>			<b>-0.253 to -0.046</b>	<b>-3.924 (23)</b>	<b>&lt;0.001</b>	<b>0.802</b>		
<b>Compactness</b>	<b>0.269 (0.221)</b>			<b>0.146 to 0.391</b>	<b>5.949 (23)</b>	<b>&lt;0.001</b>	<b>1.214</b>		
Color	0.068 (0.182)			-0.033 to 0.169	1.837 (23)	0.079	0.374		

CM, composite model; mean  $\rho$  are the Fisher-transformed  $\rho$ ; CI, 98.3% confidence interval for EVC, VOTC and LOTC; 95% CI for CM; 98.8% CI for size, elongation, compactness, and color. Highlighted in bold are the significant results;  $\alpha = 0.017$  for EVC, VOTC, and LOTC;  $\alpha = 0.05$  for CM;  $\alpha = 0.0125$  for size, elongation, compactness, and color; n.s.= nonsignificant results with  $P > 0.250$ .

$\alpha_{\text{corrected}}$ : 0.0063, one-tailed). We found that, in addition to animates, inanimates, humans, and nonhumans, 19-mo-olds showed an ability to represent the subordinate categories of human bodies [mean<sub>difference</sub> = -0.195; 99.37% CI =  $-\infty$  to -0.060;  $t(24) = -3.902$ ,  $P < 0.001$ ,  $d = 0.780$ ], nonhuman bodies [mean<sub>difference</sub> = -0.160; 99.37% CI =  $-\infty$  to -0.014;  $t(24) = -2.956$ ,  $P = 0.004$ ,  $d = 0.603$ ], nonhuman faces [mean<sub>difference</sub> = -0.216; 99.37% CI =  $-\infty$  to -0.094;  $t(24) = -4.763$ ,  $P < 0.0001$ ,  $d = 0.953$ ], and natural-small objects [mean<sub>difference</sub> = -0.179; 99.37% CI =  $-\infty$  to -0.076;  $t(24) = -4.727$ ,  $P < 0.0001$ ,  $d = 0.986$ ] (all other  $P$ s  $> 0.024$ ).

Difference between images in color, size, elongation, or compactness did not account for the infants' looking behavior (Table 1), suggesting priority of categorical information, over more general physical differences, in the processing of visual objects at 19 mo.

Finally, we assessed possible preferences, considering the mean looking times (MLTs) toward each category, averaged across trials and subjects. A one-way repeated-measures ANOVA showed an effect of Category [ $F(7, 168) = 16.259$ ,  $P < 0.0001$ ;  $\eta^2 = 0.790$ ], which reflected a preference (i.e., longer looking times) for animate (mean = 2.103 s  $\pm$  0.344) over inanimate categories [mean = 1.598 s  $\pm$  0.247; mean<sub>difference</sub> = 0.505; 95% CI = 0.367 to 0.643;  $t(24) = 7.551$ ,  $P < 0.0001$ ;  $d = 1.509$ ], and for nonhuman animals (mean = 2.317 s  $\pm$  0.463) over humans [mean = 1.889 s  $\pm$  0.452; mean<sub>difference</sub> = -0.428; 95% CI = -0.676 to -0.179;  $t(24) = -3.552$ ,  $P = 0.002$ ;  $d = 0.711$ ] (see *SI Appendix, Supplementary Results* for details). The matrix representing  $t$  values for each

pairwise comparison of MLTs remarkably replicated the structure of the DLT-RDM (Fig. 24; see *SI Appendix, Table S3* for  $t$  and  $P$  values), showing categorization and discrimination based on animacy and humanness.

**Ten-month-olds.** The looking behavior of 10-mo-olds (Fig. 2B) was significantly correlated with the composite-RDM of adult categorization as well as with fMRI-based RDMs reflecting object-related responses in selective aspects of the adults' visual ventral stream. Further analyses showed that objects were principally categorized by animacy.

More precisely, although correlations of the infants' DLT-RDM with activations in the broad ROIs (EVC, VOTC, and LOTC) did not reach the significance level (Fig. 3C and Table 1), the vector-of-ROIs analysis showed correlation with RDMs derived from the early visual cortex (V1) and fusiform gyrus ( $P$ s  $< 0.001$ ) (Fig. 3D). The DLT-RDMs also correlated with the synthetic composite-RDM. Infant's behavior was not explained by differences in visual feature, such as color, size, elongation, or compactness (Table 1).

Which of the six models underlying the composite-RDM best represented the infants' behavior? A stepwise linear regression showed correlation of the infants' DLT-RDM with the animacy model only [ $\alpha_{\text{corrected}}$ : 0.0083, two-tailed; mean  $\beta = 0.059$ ; 99.17% CI = 0.002 to 0.117;  $t(23) = 2.981$ ,  $P = 0.007$ ;  $d = 0.608$ ; for all other regressors:  $P$ s  $> 0.036$ ] (*SI Appendix, Table S1*). Consistent with this finding, average within-category DLTs were significantly lower than average between-category DLTs for the animacy model [mean<sub>difference</sub> = -0.061; 99.17% CI =  $-\infty$  to -0.007;

$t(23) = -2.919, P = 0.004; d = 0.592; \alpha_{\text{corrected}}: 0.0083$ , one-tailed], but not for the other models (*SI Appendix, Table S2*).

The analysis of the MLTs revealed an effect of Category [one-way repeated-measures ANOVA:  $F(7, 161) = 9.422, P < 0.0001; \eta^2 = 0.808$ ], reflecting preference (i.e., longer looking times) for animate (mean =  $2.044 \text{ s} \pm 0.316$ ) over inanimate categories [mean =  $1.601 \text{ s} \pm 0.263; m_{\text{difference}} = 0.443; 95\% \text{ CI} = 0.310 \text{ to } 0.576; t(23) = 6.901, P < 0.0001; d = 1.409$ ] (*SI Appendix, Supplementary Results 3*). Relationships between categories computed on the MLTs replicated the structure of the DLT-RDM, showing that 10-mo-olds categorized objects based on animacy, with a preference for animate objects (Fig. 2B and *SI Appendix, Table S3*).

**Four-month-olds.** Unlike older infants, 4-mo-olds showed no evidence of categorization; they looked longer at human faces and big-inanimate objects or, otherwise, at the larger, less elongated and more compact of two images on the screen.

More precisely, the infants' DLT-RDM did not match the organization of object-related information in any broadly defined ROIs of the adults' visual cortex (Fig. 3C and Table 1), or in any smaller partition of the visual ventral stream (vector-of-ROIs analysis) (Fig. 3D). No correlation was found with the composite model of adult categorization (Table 1) or with any of the six underlying models (stepwise linear regression: all  $t_s < 1$ , n.s.) (*SI Appendix, Table S1*). For none of the categorization models were DLTs larger for between-category than within-category comparisons ( $P_s > 0.032$ ) (*SI Appendix, Table S2*). Instead, infants' DLT-RDM correlated positively with the RDMs based on image size and compactness, and negatively with the RDM based on elongation (Table 1). No correlation was found with the color model.

A one-way repeated-measures ANOVA on the MLTs showed an effect of Category [ $F(7, 161) = 22.970; P < 0.0001; \eta^2 = 0.802$ ], which was driven by a preference for human faces over all other categories ( $\alpha_{\text{corrected}}: 0.0018$ ; all  $P_s < 0.001$ ) (Fig. 2C and *SI Appendix, Table S3*). Within the inanimate categories, infants looked longer at big over small objects, whether artificial or natural ( $P_s < 0.0001$ ). We note that the two most preferred categories (human faces and big-inanimate) were those with the largest image size ( $>15,000$  pixels), the least elongated, and among the most compact shape (*SI Appendix, Supplementary Results 4 and Fig. S2*). Thus, size, elongation, and compactness, rather than object identity, could explain object preferences in 4-mo-olds. In line with this, the MLTs computed for each image across subjects correlated positively with image size ( $\rho = 0.515, P < 0.0001$ ) and compactness ( $\rho = 0.397, P < 0.001$ ), and negatively with elongation ( $\rho = -0.531, P < 0.0001$ ). That is, the larger the image, the less elongated, or the more compact the shape, the longer the looking time. Given this result, with a new stepwise linear regression, we reassessed the relationship of the DLT-RDM with the six categorical models, after removing the variance explained by size, elongation, and compactness. Yet, no model accounted for the remaining variance (all  $P_s > 0.308$ ) (*SI Appendix, Supplementary Results 4 and Table S4*). Likewise, we found no evidence of categorization comparing average within- and between-category DLTs (all  $P_s > 0.136$ ) (*SI Appendix, Supplementary Results 4 and Table S5*).

**Comparison between groups.** *Categorical distinctions.* The above analyses showed that categorization by animacy emerged by 10 mo, while categorization by humanness, and additional categories of the eight-category model emerged by 19 mo. Additional between-subjects analyses confirmed the differences between age groups, with respect to object categorization by animacy, humanness, and by the eight-category model. In effect, for each of the three categorization models that appeared to change over time (i.e., across groups), we analyzed the variation of the mean difference between averaged between-categories vs.

averaged within-category DLTs with a one-way ANOVA, including Age as between-subject factor (4 mo, 10 mo, 19 mo).

As for the animacy model, we found a trend for the effect of Age [ $F(2, 70) = 2.752; P = 0.071; \eta^2 = 0.073$ ]. Pairwise comparisons showed that 4-mo-olds differed from 19-mo-olds [mean<sub>4-mo-olds</sub> =  $-0.022 \pm 0.121 \text{ SD}$ ; mean<sub>19-mo-olds</sub> =  $-0.095 \pm 0.106$ ; 95% CI = 0.009 to 0.139;  $t(47) = 2.273; P = 0.028; d = 0.649$ ] but not from 10-mo-olds [mean<sub>10-month-olds</sub> =  $-0.061 \pm 0.103$ ; 95% CI =  $-0.026$  to 0.105;  $t(46) = 1.225; P = 0.227; d = 0.354$ ]. Ten- and 19-mo-olds did not differ [95% CI =  $-0.026$  to 0.094;  $t(47) = 1.144; P = 0.258; d = 0.327$ ]. As for the humanness model, the effect of Age was significant [ $F(2, 70) = 4.027; P = 0.022; \eta^2 = 0.103$ ]. Nineteen-month-olds differed from both 4-mo-olds [mean<sub>4-mo-olds</sub> =  $-0.036 \pm 0.216$ ; mean<sub>19-mo-olds</sub> =  $-0.158 \pm 0.190$ ; 95% CI = 0.006 to 0.239;  $t(47) = 2.109; P = 0.040; d = 0.602$ ] and 10-mo-olds [mean<sub>10-mo-olds</sub> =  $-0.003 \pm 0.202$ ; 95% CI = 0.043 to 0.268;  $t(47) = 2.780; P = 0.008; d = 0.794$ ]. Four- and 10-mo-olds did not differ [95% CI =  $-0.155$  to 0.088;  $t(46) = -0.548; P = 0.586; d = 0.158$ ]. As for the eight-categories model, the effect of Age was significant [ $F(2, 70) = 4.292; P = 0.018; \eta^2 = 0.109$ ]. Nineteen-month-olds differed from both 4- [mean<sub>4-mo-olds</sub> =  $-0.048 \pm 0.120$ ; mean<sub>19-mo-olds</sub> =  $-0.135 \pm 0.085$ ; 95% CI = 0.028 to 0.147;  $t(47) = 2.966; P = 0.005; d = 0.845$ ] and 10-mo-olds [mean<sub>10-mo-olds</sub> =  $-0.057 \pm 0.137$ ; 95% CI = 0.013 to 0.144;  $t(47) = 2.426; P = 0.019; d = 0.690$ ]. Four- and 10-mo-olds did not differ [95% CI =  $-0.066$  to 0.084;  $t(46) = 0.243; P = 0.809; d = 0.070$ ].

In sum, categorization by humanness and by the so-called eight-categories model changed between 10 and 19 mo. Categorization by animacy differed between 4 and 19 mo. However, it did not differ significantly between 4 and 10 mo, although separate analyses for each group suggest a change in the representation of animate and inanimate categories between the two age groups. Given the robust evidence for categorization by animacy in 10-mo-olds, the absence of a difference between 4- and 10-mo-olds might suggest a latent categorization by animacy in the younger age group. Exp. 2 speaks to this question.

*Lower-level dimensions.* The above analyses showed that features such as size, elongation, and compactness contributed to driving the behavior of 4-mo-olds but not of 10- and 19-mo-olds. We assessed these differences across groups with three one-way ANOVAs, testing the effect of Age (4 mo, 10 mo, 19 mo) on the variation of correlation coefficients for the correlation between the DLT-RDMs and the RDMs for each of three visual dimensions (image size, elongation, and compactness). The correlation between looking behavior and image size changed over time [Effect of Age:  $F(2, 70) = 21.259; P < 0.001$ ]. In particular, correlation was higher in 4-mo-olds than in 10-mo-olds [mean<sub>4-mo-olds</sub> =  $0.315 \pm 0.173$ ; mean<sub>10-mo-olds</sub> =  $-0.068 \pm 0.208$ ; 95% CI = 0.136 to 0.359;  $t(46) = 4.483; P < 0.001; d = 1.294$ ] and 19-mo-olds [mean<sub>19-mo-olds</sub> =  $-0.034 \pm 0.194$ ; 95% CI = 0.244 to 0.455;  $t(47) = 6.649; P < 0.001; d = 1.903$ ]. Ten- and 19-mo-olds did not differ [95% CI =  $-0.014$  to 0.217;  $t(47) = 1.768; P = 0.084; d = 0.505$ ]. There was also a significant effect of Age for the correlation between elongation and looking behavior [ $F(2, 70) = 22.180; P < 0.001$ ]. Correlation was higher in 4-mo-olds than in 10-mo-olds [mean<sub>4-mo-olds</sub> =  $-0.321 \pm 0.187$ ; mean<sub>10-mo-olds</sub> =  $-0.041 \pm 0.197$ ; 95% CI =  $-0.392$  to  $-0.169$ ;  $t(46) = -5.051; P < 0.001; d = -1.458$ ] and 19-mo-olds [mean<sub>19-mo-olds</sub> =  $-0.010 \pm 0.151$ ; 95% CI =  $-0.409$  to  $-0.214$ ;  $t(47) = -6.418; P < 0.001; d = -1.830$ ]. Ten- and 19-mo-olds did not differ [95% CI =  $-0.131$  to 0.070;  $t(47) = -0.614; P = 0.542; d = -0.175$ ]. Analogous results were found for the correlation between looking behavior and compactness [Effect of Age:  $F(2, 70) = 16.694; P < 0.001$ ]; correlation was higher in 4-mo-olds than in 10-mo-olds [mean<sub>4-mo-olds</sub> =  $0.208 \pm 0.155$ ; mean<sub>10-mo-olds</sub> =  $0.054 \pm 0.190$ ; 95% CI = 0.054 to 0.255;  $t(46) = 3.091; P = 0.003; d = 0.892$ ] and 19-mo-olds [mean<sub>19-mo-olds</sub> =  $-0.074 \pm 0.166$ ; 95% CI = 0.189 to 0.374;

$t(47) = 6.144$ ;  $P < 0.001$ ;  $d = 1.757$ ], and higher in 19-mo-olds than in 10-mo-olds [95% CI = 0.025 to 0.230;  $t(47) = 2.496$ ;  $P = 0.016$ ;  $d = 0.712$ ]. These results, together with the above correlation analysis, demonstrated that visual features of the stimuli such as image size, elongation, and compactness predicted the behavior of 4-mo-olds but not of older infants.

**Experiment 2.** In Exp. 1, 4-mo-olds showed no evidence of categorization, but a preference for human faces and big-inanimate objects, which might be explained by physical properties, such as image size, elongation, and compactness (i.e., a tendency to look at the larger/less elongated/more compact image on the screen). We asked whether a preference for certain physical properties might have overshadowed categorical effects. To this end, we tested a new group of 4-mo-olds ( $n = 24$ ) with the same images of Exp. 1, but all matched for size (i.e., number of pixels) (Fig. 1C). Size, but not elongation and compactness, was modified because only the former can change without affecting object identity or recognizability. Results confirmed the preference for human faces and big-inanimate object, but also showed that, when size was no longer available to discriminate between two stimuli, 4-mo-olds showed categorization by animacy.

Specifically, the infants' DLT-RDMs (Fig. 2D) correlated with the composite-RDM, and the fMRI-based RDMs extracted from the anterior fusiform gyrus ( $P < 0.001$ ) (Fig. 3D) in the vector-of-ROIs analysis (Table 1). There remained a significant negative correlation with the elongation model, a significant positive correlation with the compactness model, and no correlation with the color model (Table 1).

Of the six synthetic models that contributed to the composite-RDMs, infants' behavior was best represented by the animacy model [stepwise linear regression,  $\alpha_{\text{corrected}}$ : 0.0083, two-tailed; mean  $\beta = 0.074$ ; 99.17% CI = 0.016 to 0.132;  $t(23) = 3.697$ ;  $P = 0.001$ ;  $d = 0.755$ ; for all other regressors,  $P_s > 0.12$ ] (SI Appendix, Table S1). The comparison between within-category and between-category DLTs confirmed this result, showing lower within-category than between-category DLTs for the animacy model [ $\alpha_{\text{corrected}}$ : 0.0083, one-tailed; mean<sub>difference</sub> =  $-0.076$ ; 99.17% CI =  $-\infty$  to  $-0.021$ ;  $t(23) = -3.583$ ,  $P < 0.001$ ;  $d = 0.731$ ], but not for the other models (SI Appendix, Table S2).

A one-way repeated-measures ANOVA on the MLTs showed a significant effect of Category [ $F(7, 161) = 59.466$ ;  $P < 0.0001$ ;  $\eta^2 = 0.869$ ], which reflected a preference for human faces over all other categories ( $\alpha_{\text{corrected}}$ : 0.0018; all  $P_s < 0.0001$ ), for non-human faces over human or nonhuman bodies and for bigger small-inanimate objects (all  $P_s < 0.001$ ) (SI Appendix, Supplementary Results 3 and Table S3). Thus, the preference for human faces and big objects, which were the largest objects in Exp. 1, remained despite matching images for size. Moreover, the average MLTs for individual images were negatively correlated with elongation ( $\rho = -0.337$ ,  $P = 0.004$ ) and positively correlated with compactness ( $\rho = 0.590$ ,  $P < 0.001$ ), confirming the bias for the less elongated and more compact shapes. Given the last result, we reassessed the correlation of the DLT-RDM with the six categorical models, after removing the variance explained by elongation and compactness. Again, the animacy model was the only significant regressor [ $\alpha_{\text{corrected}}$ : 0.0083; mean  $\beta = 0.055$ ; 99.17% CI = 0.005 to 0.106;  $t(23) = 3.176$ ;  $P = 0.004$ ;  $d = 0.647$ ; for all other regressors,  $P_s > 0.061$ ] (SI Appendix, Table S4). Categorization by animacy was confirmed by higher between- than within-category DLTs for the animacy model only [ $\alpha_{\text{corrected}}$ : 0.0083, one-tailed; mean<sub>difference</sub> =  $-0.057$ ; 99.17% CI =  $-\infty$  to  $-0.009$ ;  $t(23) = -3.081$ ;  $P = 0.003$ ;  $d = 0.629$ ; for all other comparisons  $P > 0.194$ ] (SI Appendix, Supplementary Results 4 and Table S5).

Can face preference explain the categorization by animacy at 4 mo? The animate-inanimate distinction in a group that

showed strong face preference could reflect the distinction between objects-with-face (preferred) vs. objects-without-face. In effect, infants looked longer at human faces but also at non-human faces, compared to body stimuli (SI Appendix, Table S3). However, they also looked longer at human bodies, whose faces were not visible than nonhuman bodies, whose faces were visible. An additional analysis addressing the correlation between the DLT-RDM of 4-mo-olds and a synthetic model considering all the faces as one category, and all other stimuli as another category, only yielded a nonsignificant trend [mean  $\beta = 0.055$ ; 99.17% CI =  $-0.027$  to 0.137;  $t(23) = 1.932$ ;  $P = 0.066$ ;  $d = 0.394$ ]. Thus, although it could contribute to it, the face preference cannot fully account for the animate-inanimate categorization in 4-mo-olds.

## Discussion

Categorization is the mechanism through which the human mind makes sense of the environment by organizing the things of the world in categories. Categorization begins at a young age, with the ability to appreciate perceptual similarities between objects, and acquires refinement with knowledge and language acquisition. What type of real-world object categories infants can represent before developing a sizable lexicon and a rich system of knowledge about the world? We considered the hypothesis that the early stages of visual object categorization are guided by the same dimensions that structure object representations in the visual cortex of the primate brain.

Our findings demonstrate that early visual object categorization along the fundamental dimensions represented in the human visual cortex, is an incremental process with two milestones. The first, between 4 and 10 mo, establishes the transition from an exploration of the environment guided by general visual saliency to an organization that corresponds to the animate-inanimate categorical distinction; the second, between 10 and 19 mo, presents a spurt of visual object categories toward mature organization.

All the categorization effects that we observed analyzing looking-time differences between objects were paired with preference effects, as indexed by MLTs. That is, when infants showed categorizing objects by animacy, they also looked longer at animate than inanimate objects; when they showed categorizing objects by humanness, they also looked longer at nonhuman than human animals. The cooccurrence of preference and categorization suggests that the earliest visual categories to emerge in infancy are those that are important enough to give rise to a hierarchy of preferences.

However, we emphasize that categorization is not equal to preference. We operationalized categorization in terms of larger difference for between-category comparisons than within-category comparisons. The systematic cooccurrence of effects in DLTs and in MLTs suggests that categorization effects were mainly carried by high between-categories DLTs. However, our approach can in principle detect categorization effects that are not paired with a systematic preference (i.e., categorical distinctions that do not yield consistent differences in the MLTs); for example, categorization by animacy could have been observed as long as within-category differences were lower than between-categories differences, and even if some infants preferred one category, and others, the other category yielding null group preference effect. Clearly, categorization effects are easier to detect, the lower the within-category differences (i.e., for rather homogenous object categories).

**Within 4 mo.** Four-month-olds showed no evidence of categorization based on any of the categorical dimensions considered

here, when image size allowed discriminating between two images on the screen (Exp. 1). Accordingly, their looking behavior did not match visual object representation in any sector of the adults' visual cortex. Infants' look, however, was not random. They looked longer at the larger image and the less elongated and more compact object on the screen, with size, elongation, and compactness differences predicting looking-time differences. Moreover, MLTs revealed preference for human faces and real-world size big (vs. small) inanimate objects.

The preference for human faces, extensively documented in very young infants (33–35), has been explained by the detection of the characteristic eyes–mouth configuration, and iris–pupil–sclera contrast of human eyes (33). However, if performance here reflected detection of those features, infants would have shown categorization of human faces, as all our human faces carried those features. Instead, preference for faces occurred without categorization (i.e., within-category DLTs were as high as between-categories DLTs). This suggests that, in processing two faces, infants focused on individual- rather than category-level features, possibly reflecting a propensity or need for individuation (i.e., for processing faces as individuals rather than category members) (48), which makes the differences between two faces as salient to the visual system, as the difference between a face and another object (55).

The preference for real-world big (vs. small) objects also emerged without evidence of categorization. In this case, however, within-category DLTs comparable to between-category DLTs could reflect the visual heterogeneity of the category of big inanimate objects, which included landscapes (e.g., view of a lake), landmarks (e.g., building), and various large objects (e.g., washing-machine and stool). This observation leaves open the possibility that categorization of big objects could be found for more homogeneous sets than the current one.

The effect of real-world object size at such a young age is unprecedented and open to multiple interpretations. Scholars are debating a proper characterization of the big-small object distinction, which might relate to differences in perceptual properties (e.g., texture, spatial frequency) (8) and behavior-relevant properties of the objects (56). The current results add a new piece to the puzzle, showing early asymmetry in the attention toward big vs. small objects. In the adult brain, big objects, which typically function as landmarks, are represented in ventral aspects of the visual cortex, adjacent to place- and scene-specific areas. Small objects, which are by definition graspable, are represented in lateral aspects of the occipitotemporal cortex, also hosting areas for tool and action representation (4). Areas of the scene- and place-specific network are functionally interconnected already in the first weeks of life (32), and respond strongly to scenes in 4- to 6-mo-olds (31). In contrast, at 4 mo, infants are unable to grasp objects, showing immaturity of the networks that control hand movements and hand–object interaction. Interest in graspable objects increases during the first year of life (52, 53), as infants develop grasping skills (57). Consistent with this trajectory, we found that by 10 mo, the preference for big over small objects had disappeared. Thus, different developmental trajectories of different networks in the visual cortex might contribute to object distinctions captured by the effect of real-world size.

**From 4 to 10 mo.** When size was no longer available to discriminate between two images on the screen (Exp. 2), 4-mo-olds continued to show a preference for human (as well as nonhuman) faces and big (vs. small) objects, but they also showed categorization by animacy. Thus, categorization by animacy was functional at 4 mo but was overshadowed by physical features, such as size, making an object more visible, independently from the category. By 10 mo, infants showed an ability to overcome the importance of low-level visual features in favor of

categorical information: categorization by animacy emerged despite differences in image size. Moreover, by 10 mo, the preference for human faces and big (vs. small) objects had given way to interest in the broader category of animate entities.

Thus, the looking behavior of both 10-mo-olds in Exp. 1 and 4-mo-olds in Exp. 2 revealed categorization by animacy and matched the cortical organization of object-related information recorded from anterior (temporal) aspects of the visual ventral stream in adults. Yet, a change between 4 and 10 mo happens as infants move from prioritizing image size to prioritizing categorical information.

Animacy is the earliest categorical distinction of visual objects in infancy. This implies that representation of animate entities is not an extension of the representation of conspecifics (48, 58). Infants would rather start with a broad, underspecified representation of what animates look like, which might function as a coarse “life detector” to identify conspecifics as well as predators and preys (59). The animate–inanimate distinction would thus lay the foundation for subtler categorical distinctions, and possibly sets conditions for domain-specific processes of naive psychology (60, 61) vs. naive physics (62, 63).

**From 10 to 19 mo.** With the second developmental change between 10 and 19 mo, infants showed an ability to represent the categories animate and inanimate, but also human, nonhuman, human bodies, nonhuman bodies, and nonhuman faces. The spurt of categories by 19 mo represents another step toward the model of mature visual object representation addressed here.

While in 4- and 10-mo-olds, categorization limited to two categories was associated with object-related responses in the most anterior aspects of the visual cortex, 19-mo-olds' behavior correlated with object-related responses across the broad visual cortex of adults (from early visual cortex to ventral and lateral higher-level areas and from posterior to anterior regions along the ventral stream). This suggests that the ability to form new visual categories, from very general (e.g., animate–inanimate) to finer-grained (e.g., human vs. nonhuman bodies), involves the progressive recruitment of more and more feature spaces distributed over the visual cortex, and representing features with different complexity: as integration across regions (and feature spaces) increases, more and more categories can be represented. Promoter of this development, among other structural and functional maturation phenomena, could be the myelination of fiber tracts connecting distant areas (64), which begins around 4 mo in the occipital lobe and continues later through the temporal lobe (65). In the second year of life, categorization may further thrive with language development. Verbal labeling and communication of information about objects promote and shape the formation of new categories and, in some models, govern the transition from perceptual to conceptual categories (46, 58, 66–69).

The developmental course of visual object categorization described here confirms and extends current knowledge of object categorization in infancy. Previous studies have shown that, by 5 mo, infants can represent the abstract categories of animate and inanimate, which constrain their ability to individuate objects (58, 70, 71) and make inference about them. Thus, infants expect animate objects to be nonhollow (72), intentional (60, 73), to have beliefs (74), social affiliation (75, 76), and morality (77, 78). Inanimate objects are rather expected to be solid and obey the continuity principle (63), to be moved by contact (79), and lack intentionality (60) and strong causal power (80). In that body of research, infants identify animate entities primarily based on cues such as self-propelled motion, eyes, furry texture (62, 72), and agentive/contingent behavior (62, 72, 81–84). Three- to 4-mo-olds can also learn to represent narrower, basic-level animate categories (e.g., dogs) after repeated exposure to various exemplars of a category to

emphasize their visual similarity (41–43, 85). By contrasting two species, or basic-level categories [e.g., human vs. ape faces (29); human body vs. horse, human body vs. cat (48)]; but see refs. 47 and 86], the human-nonhuman distinction has been observed even before 19 mo. Categorization of exemplars from homogeneous, basic-level categories (e.g., a few exemplars of canonical dogs) can rely on a few, very specific physical properties of the stimuli. In contrast, here, infants were faced with the harder task of extracting category-relevant information from a heterogeneous set of static visual features. That is, categorization required infants to recognize as members of the same (animate) category, human faces, zebras, fish, and parrots, on the one hand, and hammers, washing-machines, apples, and trees (inanimate objects), on the other hand, or to recognize that a human body and a human face fall in the same category, and a horse is closer to a fish than to a human body.

In studying when and how this task is achieved, the present study introduces two important advances. First, it shows that the infants' DLTs are a reliable measure of categorical similarity, with variations in DLTs reflecting variations in the similarity of image-computable features and categorical information (see ref. 46 for converging evidence). Second, by using larger category boundaries than in previous studies, and without systematic manipulation of typical cues (e.g., self-propelled motion or agentive/contingent behavior for animacy), we have exposed in the infants' looking behavior, the mechanism through which vision extracts category-relevant information from a large, heterogeneous set of features. The correlation between infants' looking behavior and fMRI data in adults suggests that infants can form categories using the same static visual information that yields categorical object representation in the visual ventral stream of primates. The so-formed visual categories could constitute another cue for early conceptual distinctions such as animate/inanimate and human/nonhuman. To test so, future research should study how infants label those categories, and what inferences they make about them: for example, do they infer from its static appearance that a crocodile is self-propelled, has intentions and beliefs?

Future studies should also address whether other categories can be captured in the infants' looking behavior by changing the category boundaries (e.g., more/less homogeneous categories), adding other real-world features (e.g., motion), or giving infants more time to explore the images. More, or finer-grained, categories could also be uncovered by going beyond the unidimensional characterization of the infants' looking behavior afforded by looking times (87), or replicating the current methodology using neural correlates of infants' categorization. Finally, the exact nature of the category-relevant visual features that drove DLTs in infants remains to be studied. While we focused on visual features, as we targeted the model of object representation in the visual cortex, a role of other representational levels, mediated, for example, by language or semantic knowledge, remains to be tested.

**Conclusions.** We have shown that infants initiate their exploration of the visual world by giving priority to images that are more visible (i.e., the larger ones) and displaying preferences for faces and big objects. By 10 mo, they show the ability to learn that categorical information relevant to the animate-inanimate distinction is more important than general physical properties. Thus, the first act of visual object categorization divides the world into animate and inanimate entities. Other categories represented in the visual cortex emerge by 19 mo. As visual categories multiply, infants' behavior correlates with neural activity in ever-larger aspects of the adult visual cortex. Integration through growing connections within category-specific networks and between distant visual areas could be the driving force of this process. Increasing representation, and reliance

on, category-relevant information in the first years of life may signal the coupling between seeing and thinking.

## Materials and Methods

### Eye-Tracking Study.

**Participants.** The study involved 97 infants in total. Exp. 1 involved 24 infants of 4 mo (11 females; mean age: 4 mo, 15 d; range: 4:0 to 4:24), 24 infants of 10 mo (8 females; mean age: 10:26; range: 10:1 to 11:30), and 25 infants of 19 mo (11 females; mean age: 19:5; range: 18:1 to 20:1). Exp. 2 involved 24 infants of 4 mo (14 females; mean age: 4:17; range: 4:3 to 5:0). The sample size of 24 was arbitrarily chosen for the initial group of 19-mo-olds. Next, we verified that it was superior to the minimal sample size ( $n = 18$ ) required to obtain the smallest categorical effect found in 19-mo-olds (human vs. nonhuman:  $d_{\text{Cohen}} = 0.6182$ , power = 0.80,  $\alpha = 0.05$ ; GPower 3.1), and kept it constant across groups. We continued testing until we reached 24 participants per group. The last 19-mo-old infant had a twin; parents asked to test him too and we kept him in the sample. Fifty additional infants were tested in Exps. 1 and 2, but discarded (see *Analyses*). Infants were tested in the Babylab of the Institute of Cognitive Sciences Marc Jeannerod (Bron, France). Parents received travel reimbursement and gave informed consent before participation. The study was approved by the local ethics committee (CPP sud-est II).

**Stimuli.** We selected 72 total color photographs of isolated real-world objects from publicly available sets (50) or from the internet. For Exp. 1, objects were superimposed on a gray background and scaled to fit a  $350 \times 350$  pixels black frame. The final set of images consisted of nine exemplars for each of eight categories. Human faces were all female faces. For Exp. 2, all objects were resized to have the same number of pixels (54,135) without gray background (Fig. 1C).

**Procedure.** Infants sat on their parent's lap, ~60 cm away from a Tobii Eye-tracker T60XL screen, in a dark room. Parents were instructed to keep their eyes closed throughout the experiment. The experiment began after the calibration for eye-tracking and consisted of 36 trials. In a trial, two images were presented for 5 s on the left and right side of the screen, equally distant from central fixation (Fig. 1B and C). Each infant saw a unique set of pairs including all 28 possible between-category combinations and eight within-category combinations. The experiment ended after 36 trials (~3 min), or because the infant expressed discomfort, or stopped looking at the screen. Stimulus presentation and data recording were controlled through PsyScopeX ([psy.cns.sissa.it](http://psy.cns.sissa.it)).

**Analyses.** On the eye-tracking screen, we defined two areas-of-interest overlapping with the locations of the two images. Areas-of-interest were two  $350 \times 350$ -pixel squares in Exp. 1, and two masks encompassing all nonbackground pixels in Exp. 2. For every trial, we computed the cumulative looking times in each area-of-interest.

Only trials with  $\geq 1$  s of look within the areas-of-interest were considered valid. For the analyses, we discarded infants with fewer than 27 valid trials (3 of 4 of total trials) or with a strong side bias (i.e., fixation on the same side for >80% of the experiment duration). In the final analysis of Exp. 1, 4-mo-olds contributed, on average,  $35 \pm 1$  trials, 10-mo-olds,  $34 \pm 2$  trials, and 19-mo-olds,  $34 \pm 2$  trials. In Exp. 2, 4-mo-olds included in the final analysis contributed on average  $29 \pm 2$  trials. Of all the infants tested in Exp. 1, exclusion criteria led us to discard 24 because of insufficient data (13 4-mo-olds, 7 10-mo-olds, and 4 19-mo-old) and 1 4-mo-old because of a side bias. Twenty-five tested in Exp. 2 were discarded because of a side bias ( $n = 2$ ) or insufficient data ( $n = 23$ ). However, stimulus presentation differed between Exps. 1 and 2 (Fig. 1B and C). In Exp. 1 areas-of-interest were fixed, squared areas delimited (and highlighted) by a black frame, which contained object and background. In Exp. 2, areas-of-interest overlapped with the object's contours without background and frame. In sum, in Exp. 2, stimuli might have been less visually salient and, therefore, the exclusion criteria, more stringent, than in Exp. 1, yielding higher attrition rate. To address this, we carried out additional analyses adopting more lenient criteria in Exp. 2, to reach an attrition rate closer to that of Exp. 1. This analysis confirmed all of our results (*SI Appendix, Supplementary Results 5*).

For each infant, for each trial, one DLT was computed as the difference in the cumulative looking time (LT) between the right and the left area-of-interest divided by the sum of the two (i.e., the total time the infant attended to the areas of interest):  $(LT_{\text{right}} - LT_{\text{left}})/(LT_{\text{right}} + LT_{\text{left}})$ . Absolute and signed DLT values were entered in absolute and signed DLT-RDMs, respectively. Values on the diagonal (within-category DLTs) and off-diagonal values in one half of the DLT-RDM (between-category DLTs) were used for analysis.

**Category effects.** Separately for each experiment, for each group, we performed RSA to correlate the absolute DLT-RDMs with each of six

categorization models, and the composite model of adult categorization, reflecting the average of the above six models. Each model defined an RDM, where cells had value of 0, 1, or 0.5 corresponding to within-category comparisons (i.e., lowest dissimilarity), between-category comparisons (i.e., highest dissimilarity), and comparisons irrelevant for a given categorization, respectively. For example, the humanness model had 0 for human–human (e.g., human face–human body) and nonhuman–nonhuman comparisons (e.g., cow face–elephant body), 1 for human–nonhuman comparisons (e.g., human body–camel body), and 0.5 for irrelevant comparisons (e.g., artificial small object–natural large object).

First, we computed the correlation between the composite-RDM and the DLT-RDM of each infant. Individual Spearman correlation coefficients  $\rho$  for a group of infants were Fisher-transformed and tested against chance-level 0 ( $t$  test). Then, we performed a stepwise linear regression analysis for each infant, with the above six categorical models as regressors. For each regressor, the distribution of coefficients  $\beta$  in a group was compared against chance ( $t$  test). Categorization was further addressed by assessing whether, for each model, average within-category DLTs were higher than average between-category DLTs ( $t$  tests, one-tailed). All above analyses were computed considering the DLTs over the total 5-s trial duration.

**Effects of general properties of the images.** In Exp. 1, for each image, we computed a score for: 1) size (i.e., total number of pixels [350 × 350] minus number of background pixels); 2) shape-elongation (i.e., height-to-width ratio with ratio tending to 1 indicating lowest elongation); 3) shape-compactness, computed as the ratio between the area of the shape and the area of the disk with the same perimeter (values between 0 and 1, with 1 indicating maximal compactness); and 4) color (i.e., for the RGB format, the average of the mean values for red, green, and blue). Since each infant of a group saw different exemplars of a category, for each infant we created: an RDM representing signed size-differences for each pairwise comparison  $[(\text{Size}_{\text{right-image}} - \text{Size}_{\text{left-image}})/(\text{Size}_{\text{right-image}} + \text{Size}_{\text{left-image}})]$ ; an RDM representing signed elongation-differences  $[(\text{Elongation}_{\text{right-image}} - \text{Elongation}_{\text{left-image}})/(\text{Elongation}_{\text{right-image}} + \text{Elongation}_{\text{left-image}})]$ ; an RDM representing signed compactness-differences  $[(\text{Compactness}_{\text{right-image}} - \text{Compactness}_{\text{left-image}})/(\text{Compactness}_{\text{right-image}} + \text{Compactness}_{\text{left-image}})]$ ; and an RDM representing color-differences in the form of Euclidean distance between the average color-value vector of two images. For each infant, we computed Spearman correlations between the DLT-RDM and each of the three RDMs. For each age group, individual correlation coefficients  $\rho$  were Fisher-transformed and entered in a one-sample  $t$  test (chance-level 0). For the signed RDMs (size, elongation, and compactness), positive correlation values indicated longer looking times toward larger/more elongated/compact objects.

**Effects of size, elongation, and compactness on categorization.** As size, elongation, and compactness correlated with the 4-mo-olds' DLTs in Exp. 1, and elongation and compactness correlated with 4-mo-olds' DLT in Exp. 2, we reassessed the effects of categorization after removing the variance explained by those physical properties of the images (SI Appendix, Supplementary Results 4). We performed a stepwise linear regression on the signed DLT-RDM of 4-mo-olds, with size RDM, elongation RDM, and compactness RDM as regressors. Next, we performed the stepwise linear regression analysis on the absolute values of the residual matrices  $R [R = |\text{signed DLT-RDM} - \beta_{\text{size}} \text{size-RDM} - \beta_{\text{elongation}} \text{elongation-RDM} - \beta_{\text{compactness}} \text{compactness-RDM}|]$ .

**Analysis of MLT.** For each group, we computed the MLT toward each category. Differences across categories were analyzed with a one-way repeated-measures ANOVA and followed up with pairwise  $t$  tests.

#### fMRI Study on Adults.

**Participants.** Fifteen participants took part in the fMRI study (eight females; mean age: 24.9  $\pm$  3.6 SD). All had normal or corrected-to-normal vision, were screened for contraindications to fMRI, gave informed consent before participation, and were paid for their time. The local ethics committee (Comité de Protection des Personnes Sud Est V) approved the study.

**Stimuli, procedures, and analyses.** The fMRI study involved: 1) a main experiment to record neural responses to the same 72 object-stimuli shown to infants, and 2) a functional localizer session, used to define the whole visual ventral stream (from V1 to the fusiform gyrus) and three broad ROIs within the visual cortex (bilateral EVC, VOTC, and LOTC). fMRI data were used to create models of visual object categorization (i.e., RDM) based on the neural activity patterns evoked by the eight categories within each ROI, and, to capture more local effects of visual categorization, within each of the 38 consecutive slices along the antero-posterior axis that forms the visual ventral stream. RSA was used to test the correlation between the RDM extracted from each ROI and each slice of the ventral stream, and the DLT-RDM of each individual infant. See SI Appendix, Supplementary Methods for a detailed description of procedures and analysis.

**Data Availability.** Stimuli, group map fMRI data, eye-tracking data, and code for the main analyses have been deposited in the Open Science Framework repository created for this project ([https://osf.io/6rm7a/?view\\_only=dcd418e45e074e379edee09ba36840be](https://osf.io/6rm7a/?view_only=dcd418e45e074e379edee09ba36840be)).

**ACKNOWLEDGMENTS.** We thank Marion Gasselien for help with infant recruitment. Funding for this project was provided by a European Research Council Starting Grant Project: THEMPO, Grant Agreement 758473 (to L.P.) and by Fyssen Foundation Research Grant Contract 233173 (to L.P.).

1. J. R. Anderson, The adaptive nature of human categorization. *Psychol. Rev.* **98**, 409–429 (1991).
2. E. Rosch, "Principles of categorization" in *Foundations of Cognitive Psychology: Core Readings*, D. J. Levitin, Ed. (MIT Press, 1999) pp. 251–270.
3. T. Konkle, A. Caramazza, Tripartite organization of the ventral stream by animacy and object size. *J. Neurosci.* **33**, 10235–10242 (2013).
4. T. Konkle, A. Oliva, A real-world size organization of object responses in occipitotemporal cortex. *Neuron* **74**, 1114–1124 (2012).
5. M. A. Cohen, G. A. Alvarez, K. Nakayama, T. Konkle, Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *J. Neurophysiol.* **117**, 388–402 (2017).
6. J. V. Haxby et al., A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
7. N. Kriegeskorte et al., Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).
8. B. Long, C.-P. Yu, T. Konkle, Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9015–E9024 (2018).
9. H. P. Op de Beeck, J. Haushofer, N. G. Kanwisher, Interpreting fMRI data: Maps, modules and dimensions. *Nat. Rev. Neurosci.* **9**, 123–135 (2008).
10. H. P. Op de Beeck, I. Pillot, J. B. Ritchie, Factors determining where category-selective areas emerge in visual cortex. *Trends Cogn. Sci.* **23**, 784–797 (2019).
11. L. Papeo, M. F. Wurm, N. N. Oosterhof, A. Caramazza, The neural representation of human versus nonhuman bipeds and quadrupeds. *Sci. Rep.* **7**, 1–8 (2017).
12. P. E. Downing, Y. Jiang, M. Shuman, N. Kanwisher, A cortical area selective for visual processing of the human body. *Science* **293**, 2470–2473 (2001).
13. R. Epstein, N. Kanwisher, A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
14. N. Kanwisher, J. McDermott, M. M. Chun, The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
15. G. McCarthy, A. Puce, J. C. Gore, T. Allison, Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* **9**, 605–610 (1997).
16. B. Z. Mahon, A. Caramazza, Concepts and categories: A cognitive neuropsychological perspective. *Annu. Rev. Psychol.* **60**, 27–51 (2009).
17. S. Bobadilla-Suarez, C. Ahlheim, A. Mehrotra, A. Panos, B. C. Love, Measures of neural similarity. *Comput. Brain Behav.* **3**, 369–383 (2020).
18. T. Davis, R. A. Poldrack, Measuring neural representations with fMRI: Practices and pitfalls. *Ann. N. Y. Acad. Sci.* **1296**, 108–134 (2013).
19. N. Sigala, N. K. Logothetis, Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415**, 318–320 (2002).
20. B. Z. Mahon, S. Anzellotti, J. Schwarzbach, M. Zampini, A. Caramazza, Category-specific organization in the human brain does not require visual experience. *Neuron* **63**, 397–405 (2009).
21. S. Mattioni et al., Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *eLife* **9**, e50732 (2020).
22. E. Striem-Amit, A. Amedi, Visual cortex extra-striate body-selective area activation in congenitally blind people "seeing" by using sounds. *Curr. Biol.* **24**, 687–692 (2014).
23. J. van den Hurk, M. Van Baelen, H. P. Op de Beeck, Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4501–E4510 (2017).
24. A. Caramazza, J. R. Shelton, Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J. Cogn. Neurosci.* **10**, 1–34 (1998).
25. P. Adibpour, J. Dubois, G. Dehaene-Lambertz, Right but not left hemispheric discrimination of faces in infancy. *Nat. Hum. Behav.* **2**, 67–79 (2018).
26. A. de Heering, B. Rossion, Rapid categorization of natural face images in the infant right hemisphere. *eLife* **4**, e06564 (2015).
27. M. Buiatti et al., Cortical route for facelike pattern processing in human newborns. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4625–4630 (2019).
28. T. Gliga, G. Dehaene-Lambertz, Development of a view-invariant representation of the human head. *Cognition* **102**, 261–288 (2007).
29. S. Peykarjou, S. Hoehl, S. Pauen, B. Rossion, Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Sci. Rep.* **7**, 12526 (2017).
30. S. Peykarjou, S. Pauen, S. Hoehl, How do 9-month-old infants categorize human and ape faces? A rapid repetition ERP study: Infants' categorization of human and ape faces. *Psychophysiol* **51**, 866–878 (2014).

31. B. Deen *et al.*, Organization of high-level visual cortex in human infants. *Nat. Commun.* **8**, 13995 (2017).
32. F. S. Kamps, C. L. Hendrix, P. A. Brennan, D. D. Dilks, Connectivity at the origins of domain specificity in the cortical face and place networks. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6163–6169 (2020).
33. T. Farroni *et al.*, Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17245–17250 (2005).
34. E. Valenza, F. Simion, V. M. Cassia, C. Umiltà, Face preference at birth. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 892–903 (1996).
35. M. H. Johnson, S. Dziurawiec, H. Ellis, J. Morton, Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19 (1991).
36. L. Bardi, L. Regolin, F. Simion, Biological motion preference in humans at birth: Role of dynamic and configural properties. *Dev. Sci.* **14**, 353–359 (2011).
37. F. Simion, L. Regolin, H. Bulf, A predisposition for biological motion in the newborn baby. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 809–813 (2008).
38. R. S. Bhatt, A. Hock, H. White, R. Jubran, A. Galati, The development of body structure knowledge in infancy. *Child Dev. Perspect.* **10**, 45–52 (2016).
39. N. Zieber *et al.*, Body representation in the first year of life. *Infancy* **15**, 534–544 (2010).
40. V. Slaughter, M. Heron, S. Sim, Development of preferences for the human body shape in infancy. *Cognition* **85**, B71–B81 (2002).
41. P. C. Quinn, P. D. Eimas, Perceptual cues that permit categorical differentiation of animal species by infants. *J. Exp. Child Psychol.* **63**, 189–211 (1996).
42. P. C. Quinn, P. D. Eimas, M. J. Tarr, Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *J. Exp. Child Psychol.* **79**, 78–94 (2001).
43. S. Pauen, S. Peykarjou, "Neural correlates of human categorization in early infancy" in *The Oxford Handbook of Developmental Cognitive Neuroscience*, K. Cohen Kadosh, Ed. (Oxford Handbooks Online, 2021), <https://10.1093/oxfordhob/9780198827474.013.9>. Accessed 9 June 2021.
44. R. N. Aslin, What's in a look? *Dev. Sci.* **10**, 48–53 (2007).
45. J.-R. Hochmann, Word frequency, function words and the second gavagai problem. *Cognition* **128**, 13–25 (2013).
46. A. S. LaTourrette, S. R. Waxman, Naming guides how 12-month-old infants encode and remember objects. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21230–21234 (2020).
47. S. Pauen, Early differentiation within the animate domain: Are humans something special? *J. Exp. Child Psychol.* **75**, 134–151 (2000).
48. P. C. Quinn, P. D. Eimas, Evidence for a global categorical representation of humans by young infants. *J. Exp. Child Psychol.* **69**, 151–174 (1998).
49. N. Kriegeskorte, M. Mur, P. Bandettini, Representational similarity analysis—Connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
50. R. Kiani, H. Esteky, K. Mirpour, K. Tanaka, Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J. Neurophysiol.* **97**, 4296–4309 (2007).
51. E. M. Brannon, D. Lutz, S. Cordes, The development of area discrimination and its implications for number representation in infancy. *Dev. Sci.* **9**, F59–F64 (2006).
52. K. Libertus *et al.*, Size matters: How age and reaching experiences shape infants' preferences for different sized objects. *Infant Behav. Dev.* **36**, 189–198 (2013).
53. C. Newman, J. Atkinson, O. Braddick, The development of reaching and looking preferences in infants to objects of different sizes. *Dev. Psychol.* **37**, 561–572 (2001).
54. L. Huang, Space of preattentive shape features. *J. Vis.* **20**, 10 (2020).
55. D. Maurer, R. L. Grand, C. J. Mondloch, The many faces of configural processing. *Trends Cogn. Sci.* **6**, 255–260 (2002).
56. C. Magri, T. Konkle, A. Caramazza, The contribution of object size, manipulability, and stability on neural responses to inanimate objects. *NeuroImage*, **237**, 118098 (2021).
57. M. E. McCarty, R. K. Clifton, D. H. Ashmead, P. Lee, N. Goubet, How infants use vision for grasping objects. *Child Dev.* **72**, 973–987 (2001).
58. L. Bonatti, E. Frot, R. Zangl, J. Mehler, The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognit. Psychol.* **44**, 388–426 (2002).
59. G. Vallortigara, L. Regolin, F. Marconato, Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biol.* **3**, e208 (2005).
60. A. L. Woodward, Infants selectively encode the goal object of an actor's reach. *Cognition* **69**, 1–34 (1998).
61. E. S. Spelke, E. P. Bernier, A. E. Skerry "Core social cognition" in *Navigating the Social World*, M. R. Banaji, S. A. Gelman, Eds. (Oxford University Press, 2013), pp. 11–16.
62. R. Saxe, J. B. Tenenbaum, S. Carey, Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychol. Sci.* **16**, 995–1001 (2005).
63. E. Spelke, Initial knowledge: Six suggestions. *Cognition* **50**, 431–445 (1994).
64. J. Dubois *et al.*, The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. *Neuroscience* **276**, 48–71 (2014).
65. S. C. L. Deoni *et al.*, Mapping infant brain myelination with magnetic resonance imaging. *J. Neurosci.* **31**, 784–791 (2011).
66. K. Dewar, F. Xu, Do 9-month-old infants expect distinct words to refer to kinds? *Dev. Psychol.* **43**, 1227–1238 (2007).
67. F. Xu, S. Carey, Infants' metaphysics: The case of numerical identity. *Cognit. Psychol.* **30**, 111–153 (1996).
68. S. R. Waxman, D. B. Markow, Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognit. Psychol.* **29**, 257–302 (1995).
69. G. Westermann, D. Mareschal, From perceptual to language-mediated categorization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20120391 (2013).
70. F. Xu, S. Carey, N. Quint, The emergence of kind-based object individuation in infancy. *Cognit. Psychol.* **49**, 155–190 (2004).
71. L. Surian, S. Caldi, Infants' individuation of agents and inert objects. *Dev. Sci.* **13**, 143–150 (2010).
72. P. Setoh, D. Wu, R. Baillargeon, R. Gelman, Young infants have biological expectations about animals. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15937–15942 (2013).
73. Y. Luo, R. Baillargeon, Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychol. Sci.* **16**, 601–608 (2005).
74. R. Baillargeon, R. M. Scott, Z. He, False-belief understanding in infants. *Trends Cogn. Sci.* **14**, 110–118 (2010).
75. L. J. Powell, E. S. Spelke, Preverbal infants expect members of social groups to act alike. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3965–E3972 (2013).
76. L. J. Powell, E. S. Spelke, Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition* **170**, 31–48 (2018).
77. J. K. Hamlin, Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Curr. Dir. Psychol. Sci.* **22**, 186–193 (2013).
78. M. Buyukozer Dawkins, F. Ting, M. Stavans, R. Baillargeon, "Early moral cognition: A principle-based approach" in *The Cognitive Neurosciences VI*, D. Poeppel, G. R. Mangun, M. S. Gazzaniga, Eds. (MIT Press, Cambridge, MA, 2020), pp. 7–16.
79. E. S. Spelke, A. Phillips, A. L. Woodward, "Infants' knowledge of object motion and human action" in *Causal Cognition: A Multidisciplinary Debate*, D. Sperber, D. Premack, A. J. Premack, Eds. (Clarendon Press/Oxford University Press, 1995), pp. 44–78.
80. P. Muentener, S. Carey, Infants' causal representations of state change events. *Cognit. Psychol.* **61**, 63–86 (2010).
81. R. Baillargeon, D. Wu, S. Yuan, J. Li, Y. Luo, "Young infants' expectations about self-propelled objects" in *The Origins of Object Knowledge*, B. M. Hood, L. R. Santos, Eds. (Oxford University Press, 2009), pp. 285–352.
82. L. Markson, E. S. Spelke, Infants' rapid learning about self-propelled objects. *Infancy* **9**, 45–71 (2006).
83. G. E. Newman, F. C. Keil, V. A. Kuhlmeier, K. Wynn, Early understandings of the link between agents and order. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17140–17145 (2010).
84. F. Deligianni, A. Senju, G. Gergely, G. Csibra, Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Dev. Psychol.* **47**, 1499–1503 (2011).
85. P. C. Quinn, P. D. Eimas, S. L. Rosenkrantz, Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception* **22**, 463–475 (1993).
86. L. M. Oakes, J. M. Plumert, J. M. Lansink, J. D. Merryman, Evidence for task-dependent categorization in infancy. *Infant Behav. Dev.* **19**, 425–440 (1996).
87. J. E. Kiat *et al.*, Linking patterns of infant eye movements to a neural network model of the ventral stream using representational similarity analysis. *Dev. Sci.* **25**, e13155 (2022).