Supplementary Information for

Visual object categorization in infancy

Céline Spriet¹, Etienne Abassi¹, Jean-Rémy Hochmann¹, Liuba Papeo*¹

Corresponding Author: Liuba Papeo. Email: liuba.papeo@isc.cnrs.fr

This PDF file includes:
Supplementary Results
Supplementary Methods
Supplementary References
Figs. S1 to S2
Tables S1 to S5
Supplementary Results

1. Stepwise linear regression analysis (Experiments 1-2)

We run a stepwise linear regression with the six synthetic models as regressors for each group in Experiments 1-2. As reported in the main text, this analysis showed an effect of the eight-category model, animacy model, and humanness model for 19-month-old infants (Experiment 1), and of the animacy model only for 10-month-old infants (Experiment 1) and 4-month-old infants in Experiment 2. See Supplementary Table 1 for statistical values of this analysis.

Supplementary Table 1. Results of representational similarity analysis (Stepwise linear regression) reflecting relationships between the infants’ DLT-RDMs and the synthetic models of categorization

| Exp. | Age | Regressor          | Mean β (SD) | CI (min – max) | t (df) | P      | Cohen’s d |
|------|-----|--------------------|-------------|----------------|--------|--------|-----------|
| 1    | 19 m| Eight-categories   | .090 (.112) | .026 – .155    | 4.023 (24) | <.001  | .804     |
|      |     | Animacy            | .077 (.109) | .014 – .139    | 3.514 (24) | .002  | .702     |
|      |     | Humanness          | .133 (.215) | .009 – .256    | 3.091 (24) | .005  | .618     |
|      |     | Faces/Bodies       | .022 (.135) | -.056 – .099   | 0.811 (24) | n.s.  | .163     |
|      |     | Natural/Artificial | .017 (.219) | -.109 – .143   | 0.384 (24) | n.s.  | .078     |
|      |     | Big/Small          | .066 (.176) | -.036 – .167   | 1.866 (24) | .074  | .375     |
| 1    | 10 m| Eight-categories   | .051 (.113) | -.015 – .118   | 2.222 (23) | .036  | .451     |
|      |     | Animacy            | .059 (.098) | .002 – .117    | 2.981 (23) | .007  | .608     |
|      |     | Humanness          | .002 (.200) | -.116 – .119   | 0.039 (23) | n.s.  | .010     |
|      |     | Faces/Bodies       | .045 (.181) | -.062 – .152   | 1.207 (23) | .240  | .249     |
|      |     | Natural/Artificial | -.026 (.122) | -.098 – .046   | -.105 (23) | n.s.  | .213     |
|      |     | Big/Small          | .038 (.128) | -.038 – .113   | 1.448 (23) | .161  | .297     |
| 1    | 4 m | Eight-categories   | .013 (.101) | -.047 – .073   | 0.634 (23) | n.s.  | .129     |
|      |     | Animacy            | .014 (.118) | -.055 – .084   | 0.598 (23) | n.s.  | .119     |
|      |     | Humanness          | .034 (.206) | -.088 – .156   | 0.809 (23) | n.s.  | .165     |
|      |     | Faces/Bodies       | .012 (.200) | -.106 – .130   | 0.304 (23) | n.s.  | .060     |
|      |     | Natural/Artificial | .031 (.250) | -.117 – .178   | 0.606 (23) | n.s.  | .124     |
|      |     | Big/Small          | .021 (.246) | -.124 – .166   | 0.421 (23) | n.s.  | .085     |
| 2    | 4 m | Eight-categories   | .026 (.142) | -.057 – .110   | 0.906 (23) | n.s.  | .183     |
|      |     | Animacy            | .074 (.098) | .016 – .132    | 3.697 (23) | .001  | .755     |
|      |     | Humanness          | -.061 (.223) | -.192 – .070   | -1.341 (23) | .193  | .274     |
|      |     | Faces/Bodies       | .019 (.211) | -.106 – .143   | 0.430 (23) | n.s.  | .090     |
|      |     | Natural/Artificial | .096 (.296) | -.079 – .271   | 1.591 (23) | .125  | .324     |
|      |     | Big/Small          | .008 (.232) | -.129 – .145   | 0.165 (23) | n.s.  | .034     |

Note: Exp., experiment; m, months; CI, 99.17% confidence interval; Highlighted in bold are the significant results; α = .0083, two-tailed; n.s. = non-significant results with P > .250.
2. Difference between within-category and between-category DLTs (Experiments 1-2).

For each Experiment, for each age group, for each of the six categorization models, we tested whether within-category differential looking times (DLTs) were lower than between-category DLTs. As mentioned in the main text, this analysis showed an effect of the eight-category model, animacy model, and humanness model in 19-month-old infants (Experiment 1), and of the animacy model only in 10-month-old infants (Experiment 1) and in 4-month-old infants of Experiment 2.

Supplementary Table 2 report all statistical values of this analysis.

Supplementary Table 2. Results of the DLTs analyses of within versus between category comparisons

| Exp. Age | Comparisons within/between | Mean of the difference (SD) | CI (min – max) | t (df) | P | Cohen’s d |
|----------|----------------------------|-----------------------------|----------------|--------|---|-----------|
| 1 19 m   | Eight-categories           | -.135 (.085)                | -inf – -.091   | -7.967 (24) | <.0001 | 1.588     |
|          | Animacy                   | -.095 (.106)                | -inf – -.041   | -4.481 (24) | <.0001 | .896      |
|          | Humanness                 | -.158 (.190)                | -inf – -.060   | -4.160 (24) | <.001  | .832      |
|          | Faces/Bodies              | -.057 (.141)                | -inf – .016    | -2.013 (24) | .028   | .404      |
|          | Natural/Artificial        | -.035 (.210)                | -inf – .074    | -2.825 (24) | .109   | .167      |
|          | Big/Small                 | -.074 (.178)                | -inf – .108    | -2.071 (24) | .025   | .016      |
| 1 10 m   | Eight-categories           | -.057 (.137)                | -inf – .016    | -2.018 (23) | .028   | .416      |
|          | Animacy                   | -.061 (.103)                | -inf – .007    | -2.919 (23) | .004   | .592      |
|          | Humanness                 | -.003 (.202)                | -inf – .104    | -.066 (23)  | n.s.   | .015      |
|          | Faces/Bodies              | -.045 (.202)                | -inf – .061    | -1.996 (23) | .142   | .223      |
|          | Natural/Artificial        | .029 (.139)                 | -inf – .102    | 1.036 (23)  | n.s.   | .209      |
|          | Big/Small                 | -.041 (.145)                | -inf – .036    | -1.370 (23) | .092   | .283      |
| 1 4 m    | Eight-categories           | -.048 (.120)                | -inf – .016    | -1.940 (23) | .032   | .400      |
|          | Animacy                   | -.022 (.121)                | -inf – .042    | -.873 (23)  | .196   | .182      |
|          | Humanness                 | -.036 (.216)                | -inf – .078    | -.811 (23)  | .213   | .167      |
|          | Faces/Bodies              | -.013 (.205)                | -inf – .095    | -.316 (23)  | n.s.   | .063      |
|          | Natural/Artificial        | -.043 (.252)                | -inf – .090    | -.840 (23)  | .205   | .171      |
|          | Big/Small                 | -.027 (.262)                | -inf – .111    | -.513 (23)  | n.s.   | .103      |
| 2 4 m    | Eight-categories           | -.024 (.133)                | -inf – .046    | -.891 (23)  | .191   | .180      |
|          | Animacy                   | -.076 (.104)                | -inf – .021    | -3.583 (23) | <.001  | .731      |
|          | Humanness                 | .087 (.232)                 | -inf – .209    | 1.836 (23)  | n.s.   | .375      |
|          | Faces/Bodies              | .018 (.225)                 | -inf – .136    | .383 (23)   | n.s.   | .080      |
|          | Natural/Artificial        | -.086 (.308)                | -inf – .077    | -1.360 (23) | .094   | .279      |
|          | Big/Small                 | -.014 (.259)                | -inf – .123    | -2.267 (23) | n.s.   | .054      |

Note: Exp., experiment; m, months; CI, 99.17% confidence interval; Highlighted in bold are the significant results; α = .0083, one-tailed; n.s. = non-significant results with P > .250.
3. Analysis of mean looking times (Experiments 1-2)

Separate one-way ANOVAs for 4-, 10- and 19-month-olds in Experiment 1 and for 4-month-olds in Experiment 2, revealed the effect of categories on mean looking times (MLTs). The following post-hoc analyses were carried out to follow up on the effect of category. For 19-month-old infants in Experiment 1, pairwise comparisons (one sample t-tests, $\alpha_{\text{corrected}}: 0.0018$) revealed longer looking times towards animate than inanimate objects (9 out of 16 comparisons were significant; 4 of the 7 remaining comparisons showed non-significant trends, that is, effects that did not reach the significance level, corrected for multiple comparisons). A t-test comparing all animate vs. all inanimate categories confirmed this effect (animate: $M \pm SD = 2.103 s \pm 0.344$; inanimate: $M \pm SD = 1.598 s \pm 0.247$; $M_{\text{difference}} \pm SD = 0.505 \pm 0.335$; 95% CI = 0.367 – 0.643; $t(24) = 7.551, P < 0.0001; d = 1.509$). Comparisons between human and nonhuman categories showed significant or non-significant trends ($Ps < .03$), which can be summarized as a preference for nonhuman categories ($M \pm SD = 2.317 s \pm 0.463$) over human ($M \pm SD = 1.889 s \pm 0.452$; $M_{\text{difference}} \pm SD = -0.428 \pm 0.602$; 95% CI = -0.676 – -0.179; $t(24) = -3.552, P = 0.002; d = 0.711$). Among the inanimate categories, only the comparison between natural-small vs. artificial-big showed a non-significant trend ($P = 0.019$; for other comparisons all $Ps > 0.11$).

The same analyses on 10-month-old infants (Experiment 1) showed that infants looked longer at the animate than inanimate categories (10 out of 16 comparisons were significant with $\alpha_{\text{corrected}} = 0.0018$; 4 of the 6 remaining comparisons showed non-significant trends with $Ps < 0.05$; for the remaining two comparisons, $Ps > 0.062$). A t-test comparing MLTs for all animate vs. all inanimate objects showed significantly longer MLTs for animate ($M \pm SD = 2.044 s \pm 0.316$) than for inanimate categories ($M \pm SD = 1.601 s \pm 0.263$; $M_{\text{difference}} \pm SD = 0.443 \pm 0.314$; 95% CI = 0.310 – 0.576; $t(23) = 6.901, P < 0.0001; d = 1.409$). When two animate or two inanimate categories were compared, only 1 out of 12 comparisons showed a non-significant trend ($P = 0.007$; for all other comparisons $Ps > 0.056$).
For 4-month-old infants in Experiment 1, pairwise comparisons (one sample $t$-tests, $\alpha_{\text{corrected}} = 0.0018$, two-tailed) revealed that infants looked longer at human faces relative to all other categories ($Ps < 0.001$ for 6 of the 7 comparisons between the human face category and other categories respectively), except the natural big objects, for which the MLTs only marginally differed ($P = 0.002$). No other difference was observed within the animate categories. Within the inanimate categories, infants looked longer at big than small objects, whether artificial or natural ($Ps < 0.0001$).

Four-month-old infants in Experiment 2 showed to prefer (i.e., looked longer at) human faces vs. all other categories (all $Ps < 0.0001$) and big vs. small (natural or artificial) inanimate objects (all $Ps < 0.001$). Unlike in Experiment 1, the preference for human faces and big objects in Experiment 2 was not conflated with the preference for large images, as all images featured the same number of pixels. However, the mean looking time for individual images, averaged across participants, negatively correlated with the image shape elongation ($\rho = -0.337$, $P = 0.004$) and positively with image compactness ($\rho = 0.590$, $P < 0.001$): the less elongated the shape and the more compact, the longer the looking time (see also Supplementary Results 4). See Supplementary Table 3 for $t$-values and $P$-values of all the pairwise comparisons reported here.

### 4. Relationship between MLTs and low-level features of images.

Because 4-month-olds’ looking behavior appeared to be guided by low-level features such as the size of images, the elongation and compactness (Supplementary Fig. 1), we tested whether the categories of images that infants looked at for a longer time, corresponded to the categories including larger images, less elongated and/or more compact objects. To this end, we computed the MLTs, the mean number of pixels, the mean elongation ratio and the mean compactness for each category (Supplementary Fig. 2). The three preferred categories, human faces and natural/artificial big objects, constituted the largest (Experiment 1), least elongated and among the most compact (Experiments 1-2) images. Furthermore, mean looking times (MLTs) correlated significantly with
size, elongation and compactness in Experiment 1, and with elongation and compactness in Experiment 2 (see Main text).

**Supplementary Figure 1.** Signed DLT-RDM (left), size RDM (middle left), elongation RDM (middle), compactness RDM (middle right) and color RDM (right) for 4-, 10- and 19-month-olds in Experiment 1 and 4-month-olds in Experiment 2. Example of a signed DLT-RDM, a size RDM, an elongation RDM, a compactness RDM and a color RDM for one subject of the 19-month-old group (a), the 10-month-old group (b) and the 4-month-old group (c) in Experiment 1, and one subject of the 4-month-old group in Experiment 2 (d).
Supplementary Figure 2. Looking times as a function of image size (left), elongation (middle) and compactness (right) for 4-month-olds in Experiments 1-2. Relationship between looking times (s) and average size of the images (number of pixels) for each category, average elongation (ratio between width and height) for each category and average compactness for each category for 4-month-olds in Experiment 1 (a) and relationship between looking times (s) and average elongation for each category and average compactness for each category for 4-month-olds in Experiment 2 (b). Small differences in the values of elongation between Experiment 1 and 2 reflect handmade measurement.
**Supplementary Table 3. Absolute t-values and P-values for pairwise comparisons (t tests) of MLTs**

| Age | Categories | HB | NHF | NHB | NB | NS | AB | AS |
|-----|------------|----|-----|-----|----|----|----|----|
|     |            | t  | P   | t   | P  | t  | P  | t  | P  |
| 19 m | HF         | 1.083 | n.s. | 4.083 | <0.001 | 3.079 | 0.005 | 2.645 | 0.022 | 3.003 | 0.006 | 1.187 | 0.247 | 1.921 | 0.067 |
|      | HB         | 3.138 | 0.005 | 2.317 | 0.029 | 3.706 | 0.001 | 3.461 | 0.002 | 1.903 | 0.069 | 2.567 | 0.017 | 5.914 | <0.0001 |
|      | NHF        | 0.472 | n.s. | 6.875 | <0.0001 | 8.575 | <0.0001 | 6.005 | <0.0001 | 4.702 | <0.0001 | 5.553 | <0.0001 |
|      | NHB        | 5.624 | <0.0001 | 1.264 | 0.219 | 1.076 | n.s. | 0.271 | n.s. | 2.505 | 0.019 | 1.650 | 0.112 | 0.800 | n.s. |
|      | NS         | 0.800 | n.s. | 2.800 | 0.007 | 1.927 | n.s. | 1.765 | n.s. | 1.235 | 0.219 | 1.076 | n.s. | 0.271 | n.s. |
|      | AB         | 2.972 | 0.007 | 5.050 | <0.0001 | 6.702 | <0.0001 | 3.138 | 0.005 | 3.404 | 0.002 | 1.836 | 0.079 | 2.321 | 0.030 |
| 10 m | HF         | 1.152 | n.s. | 0.564 | n.s. | 1.830 | 0.080 | 5.528 | <0.0001 | 4.148 | <0.0001 | 3.779 | <0.0001 | 6.163 | <0.0001 |
|      | HB         | 0.452 | n.s. | 1.050 | n.s. | 3.717 | 0.001 | 2.974 | 0.007 | 2.517 | 0.019 | 5.050 | <0.0001 | 4.287 | <0.0001 |
|      | NHF        | 1.566 | 0.131 | 3.642 | 0.001 | 3.701 | 0.001 | 2.878 | 0.009 | 4.918 | <0.0001 | 4.212 | 0.056 | 0.786 | 0.375 |
|      | NHB        | 2.737 | 0.012 | 1.959 | 0.062 | 1.404 | 0.174 | 2.012 | 0.056 | 2.945 | 0.007 | 2.855 | 0.009 | 1.848 | 0.078 |
|      | NS         | 0.800 | n.s. | 1.927 | n.s. | 1.765 | n.s. | 1.765 | n.s. | 1.235 | 0.219 | 1.076 | n.s. | 0.271 | n.s. |
|      | AB         | 2.972 | 0.007 | 5.050 | <0.0001 | 6.702 | <0.0001 | 3.138 | 0.005 | 3.404 | 0.002 | 1.836 | 0.079 | 2.321 | 0.030 |
| 4 m  | HF         | 5.878 | <0.0001 | 6.702 | <0.0001 | 6.877 | <0.0001 | 3.517 | 0.002 | 8.056 | <0.0001 | 4.115 | <0.0001 | 7.954 | <0.0001 |
|      | HB         | 0.457 | n.s. | 0.497 | n.s. | 5.563 | <0.0001 | 1.872 | 0.074 | 3.033 | 0.006 | 2.945 | 0.007 | 2.855 | 0.009 |
|      | NHF        | 0.716 | n.s. | 3.404 | 0.002 | 1.836 | 0.079 | 2.321 | 0.030 | 2.769 | 0.031 | 6.944 | <0.0001 | 4.552 | <0.0001 |
|      | NHB        | 5.470 | <0.0001 | 1.931 | 0.066 | 3.409 | 0.002 | 2.769 | 0.031 | 6.944 | <0.0001 | 4.552 | <0.0001 | 4.552 | <0.0001 |
|      | NS         | 0.786 | n.s. | 1.765 | n.s. | 1.765 | n.s. | 1.765 | n.s. | 1.235 | 0.219 | 1.076 | n.s. | 0.271 | n.s. |
|      | AB         | 2.972 | 0.007 | 5.050 | <0.0001 | 6.702 | <0.0001 | 3.138 | 0.005 | 3.404 | 0.002 | 1.836 | 0.079 | 2.321 | 0.030 |

Note: Exp., experiment; m, months; HF: Human Face; HB: Human Body; NHF: NonHuman Face; NHB: NonHuman Body; NB: Natural Big; NS: Natural Small; AB: Artificial Big; AS: Artificial Small; Highlighted in bold are the significant results; α = .0018; n.s.= non-significant with P > .250
Given these results, for each group of 4-month-olds, we performed a novel stepwise linear regression analysis and DLTs analysis, after removing the variance explained by size, elongation and compactness in Experiment 1, and by elongation and compactness in Experiment 2 (Supplementary Tables 4-5).

**Supplementary Table 4.** Results of the stepwise linear regression with artificial models, removing variance explained by size, elongation and compactness of 4-month-old infants' DLT-RDM in Experiments 1 and 2

| Exp. | Regressor       | Mean β (SD) | CI (min – max) | t (df) | P     | Cohen’s d |
|------|----------------|-------------|----------------|--------|-------|-----------|
| 1    | Eight-categories | .001 (.115) | -.067 – .068    | .024 (23) | n.s.  | .009      |
|      | Animacy         | .005 (.144) | -.079 – .090    | .177 (23) | n.s.  | .035      |
|      | Humanness       | .053 (.249) | -.094 – .199    | 1.043 (23) | n.s.  | .213      |
|      | Faces/Bodies    | -.021 (.239) | -.162 – .120    | -.434 (23) | n.s.  | .088      |
|      | Natural/Artificial | -.016 (.188) | -.126 – .095    | -.407 (23) | n.s.  | .085      |
|      | Big/Small       | -.039 (.195) | -.154 – .077    | -.969 (23) | n.s.  | .198      |
| 2    | Eight-categories | -.009 (.120) | -.080 – .062    | -.367 (23) | n.s.  | .075      |
|      | Animacy         | .055 (.085) | .005 – .106     | 3.176 (23) | .004  | .647      |
|      | Humanness       | -.083 (.207) | -.205 – .039    | -.1967 (23) | .061  | .401      |
|      | Faces/Bodies    | -.022 (.216) | -.149 – .106    | -.496 (23) | n.s.  | .102      |
|      | Natural/Artificial | .059 (.307) | -.122 – .240    | .948 (23) | n.s.  | .192      |
|      | Big/Small       | -.038 (.218) | -.167 – .090    | -.859 (23) | n.s.  | .174      |

Note: Exp., experiment; CI, confidence interval; Highlighted in bold are the significant results; α = .0083, two-tailed; n.s. = non-significant results with P > .250.

**Supplementary Table 5.** Results of the DLTs analysis of within versus between category comparisons, removing variance explained by size, elongation and compactness of 4-month-old infants in Experiments 1 and 2

| Exp. | Comparisons within/between | Mean of the difference (SD) | CI (min – max) | t (df) | P     | Cohen’s d |
|------|-----------------------------|-----------------------------|----------------|--------|-------|-----------|
| 1    | Eight-categories            | -.019 (.113)                | - inf – .040   | -.837 (23) | .206  | .171      |
|      | Animacy                     | -.008 (.146)                | - inf – .069   | -.251 (23) | n.s.  | .051      |
|      | Humanness                   | -.058 (.250)                | - inf – .074   | -1.128 (23) | .136  | .230      |
|      | Faces/Bodies                | .012 (.241)                 | - inf – .139   | .243 (23) | n.s.  | .049      |
|      | Natural/Artificial          | .026 (.210)                 | - inf – .137   | .604 (23) | n.s.  | .124      |
|      | Big/Small                   | .038 (.194)                 | - inf – .140   | .960 (23) | n.s.  | .196      |
| 2    | Eight-categories            | .009 (.130)                 | - inf – .077   | .323 (23) | n.s.  | .066      |
|      | Animacy                     | -.057 (.090)                | - inf – -.009  | -3.081 (23) | .003  | .629      |
|      | Humanness                   | .101 (.227)                 | - inf – .220   | 2.176 (23) | n.s.  | .444      |
|      | Faces/Bodies                | .050 (.232)                 | - inf – .172   | 1.055 (23) | n.s.  | .215      |
|      | Natural/Artificial          | -.059 (.327)                | - inf – .114   | -.879 (23) | .194  | .179      |
|      | Big/Small                   | .041 (.260)                 | - inf – .178   | .781 (23) | n.s.  | .159      |

Note: Exp., experiment; CI, confidence interval; Highlighted in bold are the significant results; α = .0083, two-tailed; n.s. = non-significant results with P > .250.
5. Experiment 2: Data analysis with more lenient inclusion criteria

In Experiment 2, the attrition rate for 4-month-olds was larger compared to 4-month-olds in Experiment 1 (37% in Exp. 1 and 50% in Exp. 2). Data from 24 out of 48 tested infants were discarded in Experiment 2, while data from 14 out of 38 tested infants were discarded in Experiment 1. A larger attrition rate may lead to the inclusion of more attentive infants, which could explain the different results in Experiment 2 compared to Experiment 1.

To take into account this possibility, we ran a novel analysis on the data from Experiment 2, using inclusion criteria that yielded an attrition rate comparable to that of Experiment 1. In particular, we modified the criterion for trial inclusion, selecting all trials in which infants looked at images for at least 800 ms (instead of 1000 ms in the original analysis reported in the main text). This change yielded the exclusion of 16 out of 48 infants (attrition rate of 33%). The analysis of this dataset confirmed the results reported in the main text. The stepwise linear regression analyzing the structure of the absolute RDMs identified only one significant regressor ($\alpha_{\text{corrected}} = 0.0083$), corresponding to the animacy model ($\text{mean } \beta \pm SD = 0.066 \pm 0.094; 99.17\% \text{ CI } = 0.020 - 0.113$; $t(31) = 3.991; P < 0.001; d = 0.705$) (for other regressors, $t$s < 2.313; $Ps > 0.028$). The DLTs analysis also showed that the difference between within-category and between-category DLTs was only significant for the animacy model ($M_{\text{difference}} \pm SD = -0.065 \pm 0.094; 99.17\% \text{ CI } = \text{-inf} - 0.023$; $t(31) = -3.927; P < 0.001; d = 0.691$) (for other models, $t$s > -2.208; $Ps > 0.017$).
Supplementary Methods

fMRI study on adults

Participants. Fifteen participants took part in the fMRI study (eight females; mean age: 24.9 years ± 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 approved the study.

Stimuli and experimental design. The fMRI study involved a main experiment on the same 72 object-stimuli (and eight categories) and a functional localizer session (see below). In the main experiment, all images were presented over two runs of 7.63 min each. Each run began with a warm-up block (10 s of fixation) and ended with a cool-down block (16 s of fixation), and included six sequences of eight blocks (one per category), for a total of 48 blocks of 6 s each (12 blocks per category in the whole experiment). Each block featured all nine exemplars of a category, in random order. Within a run, the inter-block interval duration was jittered (range: 0-6 s; total inter-block time per run: 144 s) to remove the overlap between estimates of the hemodynamic response. Jittering was optimized using the optseq tool of Freesurfer (1). During a block, a blue cross was always present in the center of the screen, while stimuli appeared for 667 ms without interval. Participants were instructed to fixate the cross throughout the experiment, detect and report a change in the color of the cross (from blue to red in 40% of the blocks) by pressing a button with the right-index finger. This task was used to minimize eye movements and maintain vigilance. The main experiment lasted 15.26 min. Participants, lying down in the scanner, viewed the stimuli binocularly (~7° of visual angle) through a mirror above their head. Stimuli were back-projected onto a screen by a liquid crystal projector (frame rate: 60 Hz; screen resolution: 1024×768 pixels, screen size: 40x30 cm). For all the stimuli, the center of the image overlapped with the center of the screen. Stimulus presentation, response collection and synchronization with the scanner were controlled with the Psychtoolbox (2) through MATLAB (MathWorks).
**Functional localizer session.** Stimuli and task were adapted from the fLoc package (3). In this task, participants saw 180 grayscale photographs of five classes: body-stimuli (headless bodies and body parts), faces, places (houses and corridors), inanimate objects (cars and musical instruments) and scrambled objects. Stimuli were presented over two runs (5.27 min each). Runs began with a warm-up block (12 s) and ended with a cool-down block (16s), and included 72 blocks of 4 s each: 12 blocks for each object class with eight images per block (500 ms per image without interruption), randomly interleaved with 12 baseline blocks featuring an empty screen. To minimize low-level differences across categories, the view, size, and retinal position of the images varied across trials, and each item was overlaid on a 10.5° phase-scrambled background generated from another image of the set. During blocks of images, some images were repeated twice, interleaved by a different image. Participants pressed a button when they detected the repetition. Stimulus presentation, response collection and synchronization with the scanner were as in the main experiment.

**Data acquisition.** Imaging was conducted on a MAGNETOM-Prisma3T scanner (Siemens Healthcare). T2*-weighted functional volumes were acquired using a gradient-echo echo-planar imaging sequence (GRE-EPI; TR/TE: 2000/30 ms, flip angle: 80°, acquisition matrix: 96x92, FOV: 210x201, 56 transverse slices, slice thickness: 2.2 mm, no gap, multiband acceleration factor: 2 and phase encoding set to anterior/posterior direction). For the main experiment and the functional localizer session, we acquired four runs for a total of 790 frames per participant. Acquisition of high-resolution T1-weighted anatomical images was performed before the functional runs and lasted eight min (MPRAGE; TR/TE/TI: 3000/3.7/1100 ms, flip angle: 8°, acquisition matrix: 320x280, FOV: 256x224 mm, slice thickness: 0.8 mm, 224 sagittal slices, GRAPPA accelerator factor: 2). Acquisition of two field maps was performed at the beginning of the fMRI session.

**Preprocessing.** Functional images were preprocessed and analyzed using SPM12 (4) and the CoSMoMVPA toolbox (www.cosmomvpa.org) in combination with MATLAB. The first four volumes of each run were discarded, taking into account initial scanner gradient stabilization. Preprocessing of the remaining volumes involved despiking, slice time correction, geometric
distortions correction using field maps, spatial realignment and motion correction using the first volume of each run as reference. Anatomical volumes were co-registered to the mean functional image, segmented into gray matter, white matter and cerebrospinal fluid in native space, and aligned to the probability maps in the Montreal Neurological Institute (MNI-152) as included in SPM12. The DARTEL method (5) was used to create a flow field for each subject and an inter-subject template, which was registered in the MNI-152 space and used for normalization of functional images. Final steps included spatial smoothing (Gaussian kernel of 2 mm FWHM for main experiment and 6 mm FWHM for functional localizer), and removing of low-frequency drifts with a temporal high-pass filter (cutoff 128 s).

Analyses – Voxel Selection. The blood-oxygen-level-dependent (BOLD) signal of each voxel in each participant in the functional localizer task was modeled using five regressors, one for each of the five conditions, one regressor for baseline blocks, and six regressors of no interest for movement correction parameters. Within a gray matter mask based on the statistical map of the second-level (group) analysis, voxels that responded to visual object stimulation across the whole brain were selected with the contrast ([bodies+faces+places+objects+scrambled objects] > [baseline]) (voxelwise $P = 0.05$).

ROIs. We created a model of visual object categorization based on the neural activity patterns evoked by the eight visual object categories, in three ROIs across the visual cortex. Voxels selected with the functional localizer task were divided in three sectors, using three bilateral masks of the early visual cortex (EVC), ventral occipitotemporal cortex (VOTC) and lateral occipitotemporal cortex (LOTC), extracted from the SPM Anatomy toolbox (6). The EVC included all voxels that responded to visual object stimulation during the functional localizer task and fell within early visual areas V1/V2 or ventral/dorsal extrastriate areas V3v/V3d. The VOTC-ROI was defined within a mask encompassing the extrastriate area V4v and fusiform gyrus (FG1/FG2/FG3/FG4). The LOTC-ROI included voxels within a mask encompassing lateral occipital areas V4la/V4lp and V5.
**Vector-of-ROIs.** To capture more local effects of visual categorization, we created RDMs from neural patterns evoked by the eight objects categories, at each consecutive partition along the antero-posterior axis that forms the visual ventral stream. First, a mask of the visual ventral stream was defined by selecting all the voxels that responded to visual object presentation in the functional localizer task and fell within V1/V2, V3v/V4v and FG1/FG2/FG3/FG4. This mask was then divided into 38 consecutive slices of 2.2 mm width, along the antero-posterior axis.

**fMRI-based RDMs.** Independently for each run of the main fMRI experiment, the BOLD signal of each voxel in each participant was modeled in a general linear model, using eight regressors for the eight object-categories, one regressor for baseline fixation blocks, and six regressors of no interest for movement correction parameters. For each voxel in the brain, groupwise beta values for a given category were extracted in a second-level group analysis using the contrast [category > baseline]. Finally, for each ROI and for each slice along the vector of ROIs, we computed an RDM representing dissimilarities between neural patterns (one minus Pearson coefficient $r$) for each between-category and within-category comparison (Fig. 3b).

**RSA.** For each RDM extracted from the three ROIs (EVC, VOTC, and LOTC), we computed the correlation with the DLT-RDM of each individual infant (Spearman correlation). Correlation coefficients $\rho$ of all infants of an age group were fisher-transformed and entered in a one-sample $t$ test (two-tailed), to assess the relationship between the infants’ looking behavior and object representations in different sectors of the adults’ visual cortex. The above analysis was repeated using RDMs derived from each slice of the vector of ROIs. Clusters of consecutive slices showing correlation with the infants’ DLT-RDMs were identified as follows: for each slice, the individual’s fisher-transformed coefficients $\rho$ were entered into a one-sample $t$ test; the $t$-values for each slice were transformed in $z$-values and analyzed in a cluster mass test to correct for multiple comparisons (7) with a threshold of $z = 1.64$ ($P = 0.05$, two-tailed; Fig. 3d) and 1000 permutations.
Eye-tracking and fMRI data and the codes for the main analyses are available in the Open Science Framework repository created for this project (https://osf.io/6rm7a/?view_only=dcd418e45e074e379edee09ba36840be).

**Supplementary References**

1. B. Fischl, FreeSurfer. *NeuroImage* **62**, 774–781 (2012).

2. D. H. Brainard, The psychophysics toolbox. *Spatial Vision* **10**, 433–436 (1997).

3. A. Stigliani, K. S. Weiner, K. Grill-Spector, Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience* **35**, 12412–12424 (2015).

4. K. Friston, J. Ashburner, T. Nichols, S. Kiebel, W. D. Penny, *Statistical Parametric Mapping: the Analysis of Functional Brain Images* (Elsevier, 2007) https://doi.org/10.1016/B978-0-12-372560-8.X5000-1 (January 29, 2021).

5. J. Ashburner, A fast diffeomorphic image registration algorithm. *NeuroImage* **38**, 95–113 (2007).

6. S. B. Eickhoff, *et al.*, A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* **25**, 1325–1335 (2005).

7. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**, 177–190 (2007).