When Do Visual Category Representations Emerge in Infants' Brains?

Xiaoqian Yan[1,2,3], Sarah Tung[1+], Bella Fascendini[1+], Yulan Diana Chen[1,2], Anthony M. Norcia[1,2*], Kalanit Grill-Spector[1,2,4*]

1. Department of Psychology, Stanford University, Stanford, CA 94305, USA
2. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94305, USA
3. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China
4. Neurosciences Program, Stanford University, Stanford, CA 94305, USA

* co-senior-author

+ equal contribution

Abstract (141/150 words)

Visual categorization structures infants' interactions with their environment. However, it is unknown when visual category representations emerge during the first year of life. Here we measured cortical representations of faces, limbs, corridors, characters, and cars using an electroencephalography (EEG) frequency-tagging paradigm in infants aged 3-4 months (N = 17), 4-6 months (N = 14), 6-8 months (N= 15), and 12-15 months (N = 15). We discovered a staggered development of category representations. Representations of faces emerge first, between 4-6 months and strengthen thereafter. Whereby representations for limbs, corridors, and characters emerge later, between 6-15 months. Not only do these data suggest that category representations are learned, but also that representations of everyday categories take different durations to develop. Thus, these findings necessitate a rethinking of the development of category representations in infancy and, have important implications for assessing atypical cortical development.

Introduction

Visual categorization is a key human cognitive skill that is supported by both selective responses to items of ecologically-relevant categories over others[1–7] as well as distributed representations[8,9] in higher-level visual cortex of children and adults. The ability to categorize the visual scene starts in infancy, supporting the development of attention, learning, and language. Behavioral studies have shown that within the first year of life, infants can perform some level of visual categorization such as distinguishing between faces and objects[10,11] and between animate categories such as cats and dogs[12,13]. Recent EEG and fMRI studies have also found that there is some level of representation of face[14–16] and other category (e.g., body, scene) information[17–20] in infants' cortex within the first year of life. However, very little is known about the temporal sequence over which visual category representations emerge in the developing infants' brain. Thus, a key open question is: When do visual category representations emerge in infants' brains?

One possibility is that infants need to organize the barrage of visual input they see, and thus representations of multiple categories are present and/or develop simultaneously in their brains. Supporting this hypothesis, selective responses to faces vs. scenes were observed in ventral

temporal cortex of 4-6 months old infants[18], and selective responses to faces, scenes, and body parts vs. objects were observed in high-level visual cortex of 2-9 month olds[19]. Moreover, recent studies have decoded category information for toys, bodies, faces, houses and animals slightly but significantly above chance from distributed electroencephalographic (EEG) evoked responses in 6-8[20] and 12-15[17] month-olds, respectively.

Alternatively, representations for different categories may develop at different times during infancy. One possibility is that specialization for some categories, like faces, is innate because of their evolutionary importance[21], whereas representations for other categories, like words, are learnt[22]. Supporting this possibility is evidence that newborns and young infants tend to orient to faces[23] and face-like stimuli[24] as well as have cortical responses to face-like stimuli[25], but word representations only emerge in childhood with the onset of reading instruction[6]. Another possibility is that visual experience is necessary for the development of all category representations, but representations for different categories may develop at different times during infancy because infants' behavior and visual diet (daily visual experiences) vary across categories. Supporting this view, research in nonhuman primates revealed that monkeys reared without seeing faces do not develop face representations[26], whereas nondeprived monkeys develop face representations gradually during the first year of life[27]. Further, behavioral measurements in humans shows that infants' visual experience varies substantially across categories. While faces constitute about 25% of the visual diet[28] of 0-3-month-olds, hands constitute less than 10% of their visual diet. However, as infants become more mobile and dexterous, they look less at faces and more at hands[28,29], with hands consisting 20% of the visual diet of 12-15-month-olds (on par with faces) and 25% of the visual diet of 24-month-olds. Thus, even if all categories are learnt, the dominance of faces in infants' early visual experience suggests the possibility that representation of faces may emerge before that of other categories that are less frequently seen.

To determine when do representations to different visual categories emerge during infancy, we used EEG and a Steady-State Visual Evoked Potential (SSVEP) paradigm (**Fig 1**) to measure brain responses in infants of four age groups (3-4-, 4-6-, 6-8-, 12-15-month-olds) when they viewed images of faces, limbs, corridors, characters, and cars. We tested at each age if there are

category-selective responses and/or distinct distributed representations to each of these categories. The first hypothesis predicts that representations to all categories will emerge at the same age, and the second hypothesis predicts that the representations of different categories will emerge at different ages, with faces, perhaps the earliest.

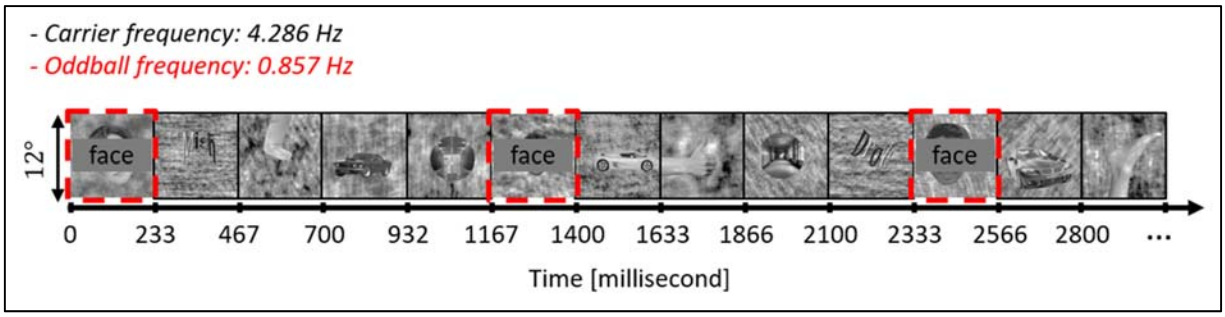

Fig. 1: Experimental Design. Example portion of a presentation sequence in a condition with faces as the oddball category. Images appeared for 233ms, such that the update (carrier) frequency is 4.286 Hz. A different exemplar from a single category appeared every 5th image at a rate of 0.857 Hz; this is referred to as the oddball frequency. In each condition, blocked by category, images were presented for 14 seconds and contained 12 such oddball cycles. In each presentation sequence (70 seconds), participants viewed in random order each of 5 conditions where each condition had a different oddball category drawn from: faces, limbs, corridors, characters, and cars. We collected on average 6 sequences per category per participant. Each 70-second sequence used different images. Images spanned 12°. Here face images are covered with the text "face" to meet with bioRxiv's publication policy.

Results

We used EEG to measure brain responses in 45 infants (for analysis, including 12 longitudinal infants) from four age groups: 3-4 months (n = 17, 7 female), 4-6 months (n = 14, 7 female), 6-8 months (n = 15, 6 female), and 12-15 months (n = 15, 4 female) using a frequency-tagging SSVEP paradigm[30,31,16,32,33] (**Fig 1**), which affords high signal-to-noise ratio with short acquisitions making it effective for infant brain measurements[16,31]. During the EEG recording, infants viewed 144 grayscale images from each of five categories: adult faces, limbs, corridors, characters, and cars presented at a rate of 4.286 Hz that were controlled for low-level visual properties[34]. We used these categories because (i) these are everyday categories present in infants environments, (ii) child and adult ventral temporal cortex (VTC) and lateral occipitotemporal cortex (LOTC) contain both clustered category-selective regions[1,2,7,35,36], and distinct distributed representations[8,9,37,38] to these categories, and (iii) these categories have regular visual features and spatial configuration[34,39]. In each condition, blocked by category, images were drawn randomly from four categories and every fifth image was a different item from a single category (**Fig. 1**). Frequency tagging the EEG signal at 4.286 Hz (carrier frequency) and its harmonics measures the

visual response to changing images. Frequency tagging the EEG signal at 0.857 Hz (oddball frequency) and its harmonics measures category-selective responses as it reflects the relative response to the target category above the general visual response.

Validation of experimental paradigm

We first conducted several analyses to determine whether our experimental paradigm enables measurement of visual category responses in infants. As visual acuity develops during the first year of life[40,41] (Supplementary **Fig S1**), one concern is that our controlled natural, gray-level stimuli may not be distinguishable to infants. Measurements of visual evoked potentials[40,41] suggest that visual acuity in 3 month olds is 5–8 cycles per degree (cpd) and in 6 month olds is 10–16 cpd (Supplementary **Fig S1**). Thus, to simulate how our images may appear to infants, we filtered all images at 5 cpd. Despite being blurry, images of different categories are distinguishable and individual items retain their identity (Supplementary **Movies 1-4**, note that the blurred face movie is not provided to meet with bioRxiv' publication policy but is available upon request). Another concern is that the amount of empirical data typically collected in infants is less than in adults, which may compromise the experimental power to detect responses in infants as it may affect the noise-bandwidth (sensitivity) of the frequency tagging analysis. Thus, we tested whether it is possible to measure category-selective responses in adults using the same experiment and same amount of data we have collected in infants. Despite lower response amplitudes in adults than infants, using the same amount of data, we found significant category-selective responses to each of these five categories (Supplementary **Fig S2**) and we were able to successfully decode category information from spatiotemporal patterns of brain responses (Supplementary **Fig S3**). This suggests that the experimental paradigm has sufficient power to identify category representations in infants. A third concern is that the infants' immature visual system may not produce robust visual responses to our rapidly presented stimuli. Thus, we next tested whether there are time and phase-locked responses in infants' brains to the images they viewed in the experiment.

Robust visual responses in occipital regions to visual stimuli in all infant age groups

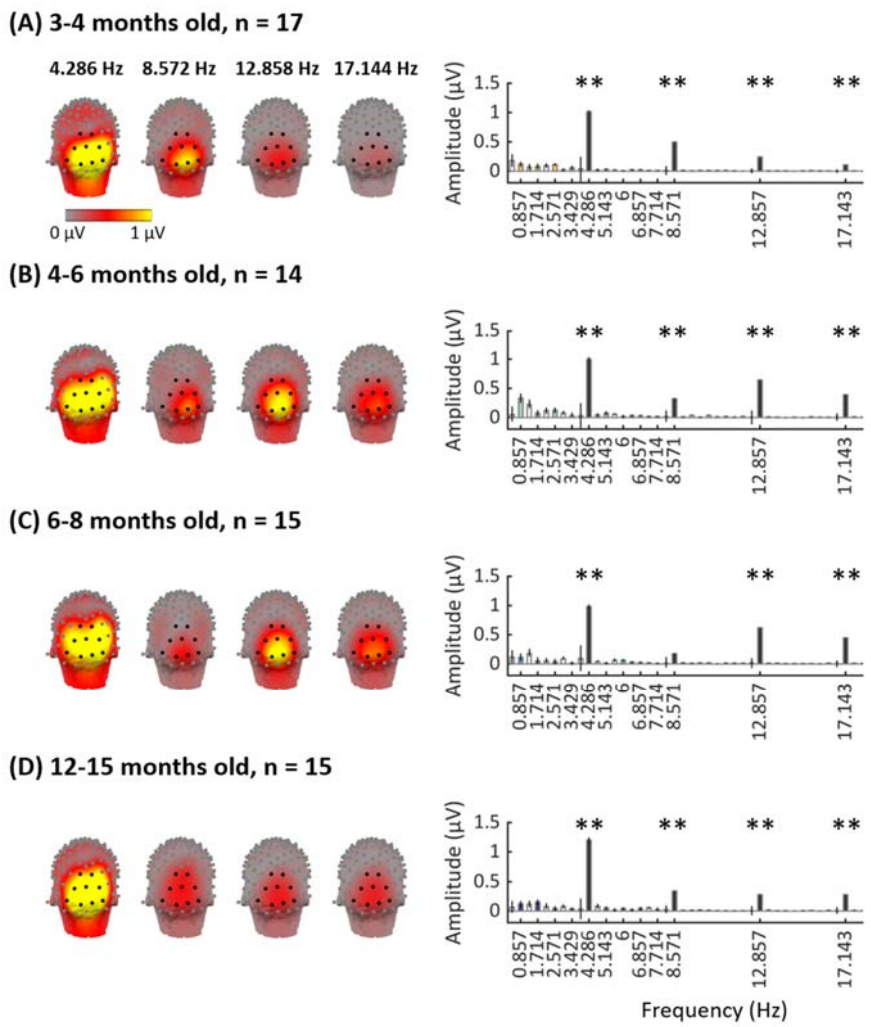

Fig. 2. Strong visual responses over occipital cortex at the carrier frequency and its harmonics in all age groups. Each panel shows mean responses across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds, n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of visual response at the carrier frequency and its first, second, and third harmonics. *Right panels in each row:* mean Fourier Amplitude Spectrum across 9 occipital electrodes of the occipital ROI (shown in black in the right panel). Data are first averaged in each participant and then across participants. *Error bars:* standard error of the mean across participants. *Asterisks:* significant response amplitude vs. zero, $p < .01$, FDR corrected. *Black bars:* amplitude of response at carrier frequency and its harmonics; *Colored bars:* amplitude of response at oddball frequency and its harmonics. *White bars:* amplitude of response at noise frequencies.

To quantify visual responses, we analyzed the mean response to changing images in a region of interest (ROI) spanning 9 occipital electrodes located over early visual cortex. As images were presented at a rate of 4.286 Hz, data were transformed to the frequency domain, and we examined if there are significant responses at 4.286 Hz (carrier frequency) and its harmonics.

Examination of the spatial distribution of responses at 4.286 Hz and its harmonics revealed strong responses concentrated spatially over occipital electrodes in all age groups (**Fig 2-left**). Quantification of the mean response amplitude over the occipital ROI revealed significant visual responses in all age groups at the carrier frequency (4.286 Hz) and its first 3 harmonics (except

for the second harmonic 8.571 Hz in 6-8-month-olds, **Fig 2-right**). This indicates that there are significant visual responses even in the youngest 3-4 months old infants.

We also evaluated whether there are developmental changes in the dynamics of these visual responses in the time domain. Time-locked visual responses in the youngest infants (3-4-month-olds) had two deflections (one negativity and one positivity) during each 233-ms image duration, whereas for infants that were 4 months old and older, we observed four deflections within this duration (two minima and two maxima, **Fig 3A,B**).

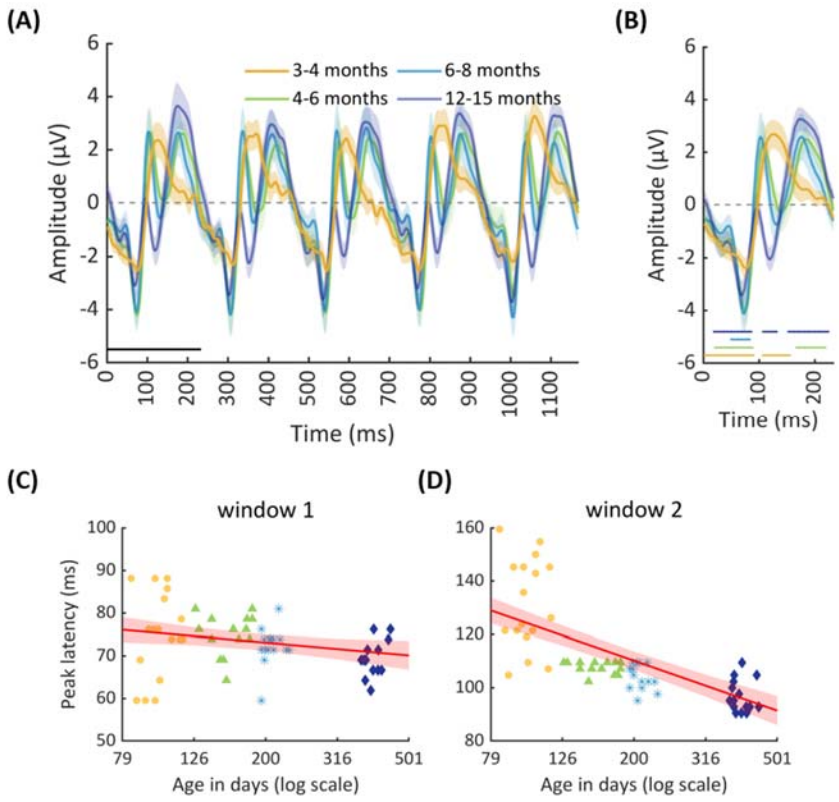

Fig. 3. Latency of the first and second peaks of the visual response decreases from 3 to 12 months of age. (A) Mean visual response over occipital electrodes for each age group. Responses are filtered for the carrier frequency and its harmonics up to 30 Hz (see Methods). Waveforms are shown for a time-window of 1.17 seconds during which 5 images are shown. *Lines:* mean response. *Shaded areas*: standard error of the mean across participants of each group. (B) Same as (A) but for a 233-ms segment. *Horizontal lines colored by age group:* significant responses vs. zero ($p < .05$ with a cluster-based analysis, see Methods). (C) Peak latency for the first peak in the 60-90 ms interval after stimulus onset. Each dot is a participant; Dots are colored by age group. *Line*: linear mixed model (LMM) estimate of peak latency as a function of log10(age). *Shaded area*: 95% confidence interval (CI). (D) Same as (C) but for the second peak in the 90-160 ms interval for 3-4-month-olds, and 90-110 ms for older infants.

Additionally, we examined if the latency of the peak visual response develops. Thus, we analyzed the timing of the peak latency during two time-windows corresponding to the first deflection (60-90 ms), and the second deflection (90-160 ms for 3-4-month-olds, and 90-110 ms for other age groups, see expanded plot in **Fig 3B**). Mean peak latency shows little change in the first window (~14 ms; **Fig 3C**), but decreases by ~32 ms in the second window from the age of 3 months to 15 months (**Fig 3D**). We quantified the developmental effects on these peak latencies

using a linear mixed model (LMM) with age (log transformed) and time window as two fixed effects and participant as a random effect. The slopes (β_{age}) of the LMM indicate the rate of change in the latency of the peak responses as a function of age. Results reveal a significant interaction between age and time window ($\beta_{age\ x\ time\ window}$ = -45.8, 95% CI: -58.4 – -33.1, $t_{(118)}$ = -7.2, p < .001) and significant main effects of age (β_{age} = 39.8, 95% CI: 19.5 – 60.0, $t_{(118)}$ = 3.9, p < .001) and time window ($\beta_{time\ window}$ = 141.7, 95% CI: 112.9 – 170.5, $t_{(118)}$ = 9.8, p < .001). Post-hoc LMM analyses for each time window with age as a fixed effect and participant as a random effect revealed a significant effect of age on the latency of the peak at both time windows (window 1: β_{age} = -7.4, 95% C:I -13.8 – -1.1, $t_{(118)}$ = -2.3, p_{FDR} < .05; window 2: β_{age} = -46.9, 95% CI: -56.6 – -37.3, $t_{(59)}$ = -9.7, p_{FDR} < .001, FDR corrected), with larger age-related decreases in latency in the second than first time window (**Fig 3C,D**). Similar analyses of the amplitude of the peak response as a function of age (log transformed) and time window (participant, random effect) revealed a significant interaction between age and time window ($\beta_{age\ x\ time\ window}$ = -4.9, 95% CI: -8.7 – -1.1, $t_{(118)}$ = -2.6, p = .01). A follow-up LMM revealed that the peak amplitude significantly decreased with age only in the second time window (β_{age} = -0.008, 95% CI: -0.01 – -0.002, $t_{(59)}$ = -2.8, p < .01; Supplementary **Tables S1, S2**). These data suggest that there are significant and reliable visual responses in all age groups, and there is development of the temporal dynamics of these brain responses to images from 3 to 15 months of age.

When do category-selective responses emerge during the first year of life?

We next asked if in addition to general visual responses there are also category-selective responses in infants. Therefore, we examined in each age group if there are significant responses at the oddball frequency (0.857Hz) and its harmonics separately for each of the five category conditions (faces, limbs, corridors, cars, and characters). The oddball response is a selective response as it reflects the relative response to the target category above the general visual response to images of other categories. **Fig 4** shows the spatial distribution and amplitude of the response to faces at the oddball frequency and its harmonics; Supplementary Figures **S4-S7** show the same analysis for the other four categories.

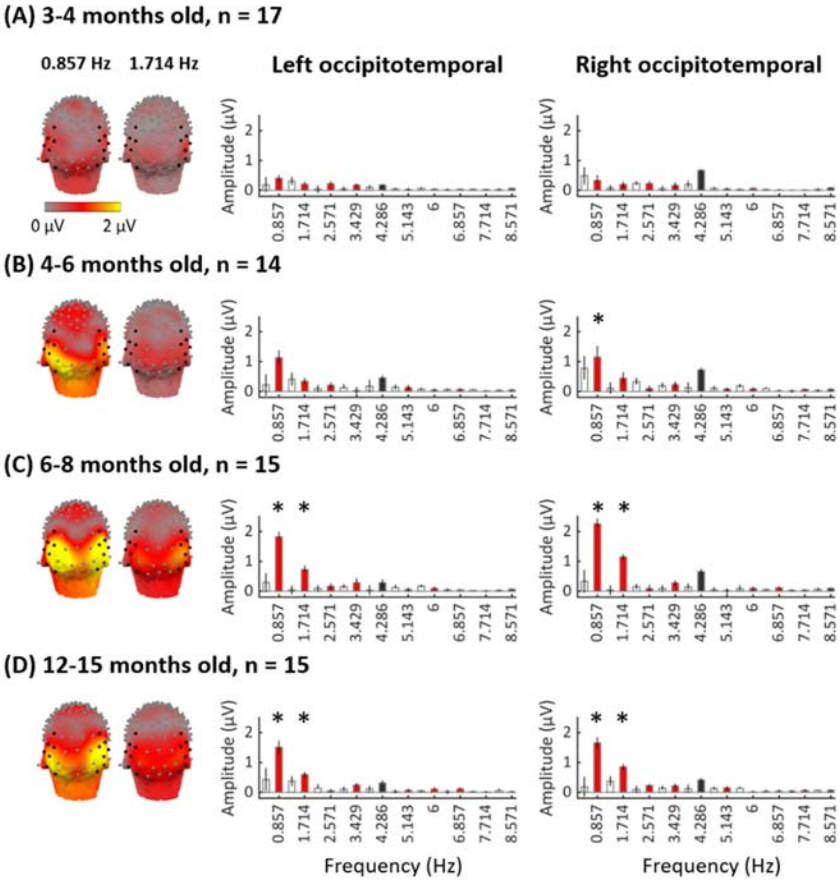

Fig. 4. Face responses emerge over occipitotemporal electrodes after 4 months of age. Each panel shows mean responses at the oddball frequency and its harmonics across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds; n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of response to oddball frequency at the 0.857 Hz and its first harmonic. Harmonic frequencies are indicted on the top. *Right two panels in each row*: mean Fourier amplitude spectrum across two ROIS: 7 left and 7 right occipitotemporal electrodes (shown in black on the left panel). Data are first averaged across electrodes in each participant and then across participants. *Error bars:* standard error of the mean across participants of an age group. *Asterisks:* significant amplitude vs. zero (p < .05, FDR corrected). *Black bars:* carrier frequency and harmonics; *Colored bars:* oddball frequency and harmonics. *White bars:* noise frequencies. Responses for the other categories (limbs, corridors, characters and cars) in Supplementary **Figs S4-S7.**

We found selective responses to some but not all categories and a differential development of category-selective responses. The largest and earliest developing category-selective responses were to faces. In contrast to the early visual responses, which were centered over occipital electrodes (**Fig 2**-left), significant category-selective responses to faces at the oddball frequency and its second harmonic were observed over lateral occipitotemporal electrodes (**Fig 4**-left). Analysis of the mean response separately for the right and left occipitotemporal ROIs in each age group indicated that there were no significant responses to faces in 3-4-month-olds at either the oddball frequency or its harmonics (**Fig 4A**). Importantly, there were significant responses to faces at the oddball frequency over the right lateral occipitotemporal electrodes in 4-6-month-olds (**Fig 4B,** p_{FDR} < .05, FDR corrected over 4 levels: the oddball frequency and its first 3 harmonics), and significant bilateral responses to faces in 6-8-month-olds and 12-15-month-olds

at both the oddball frequency and its second harmonic (**Fig 4C,D,** p_{FDR} < .05). Our paradigm reveals that face-selective responses start to reliably emerge between 4-6 months of age.

We did not find significant responses to any of the other categories over lateral occipitotemporal electrodes before 6 months of age (except for a weak but statistically significant response for cars at the oddball frequency, only in the right occipitotemporal ROI of 3-4-month-olds, Supplementary **Fig S7**). Instead, we found significant responses for (i) limbs in 6-8-month-olds in the right occipitotemporal ROI at the oddball frequency (Supplementary **Fig S4**), (ii) corridors in 6-8-month-olds in the left occipitotemporal ROI at the oddball frequency and in 12-15-month-olds in the left occipitotemporal ROI at the oddball frequency and its first harmonic (Supplementary **Fig S5**), (iii) characters in 12-15-month-olds in the left occipitotemporal ROI at the oddball frequency and its first harmonic (Supplementary **Fig S6**);

To determine how the temporal dynamics of category-selective waveforms develop from 3 to 15 months, we examined the mean time averaged waveforms (data filtered to the oddball frequency and its harmonics up to until 30 Hz) of the left and right occipitotemporal ROIs for each category and age group.

Consistent with frequency domain analyses and our results from adult participants (Supplementary **Fig S2B**), categorical responses exhibit different spatiotemporal characteristics across categories (**Fig 5**). Analyses in the time domain showed the dynamics of the significant responses to faces which emerged at 4 months of age (**Fig 5A**, horizontal lines). Responses were characterized with an initial positive deflection after stimulus onset peaking at ~500 ms followed by a negative deflection peaking at ~900 ms. For limbs, corridors, and characters, we found significant responses that emerged at 6 months of age (**Fig 5B-D**). The spatiotemporal dynamics associated with these three categories were different from faces: there was only a single negative defection after stimulus onset which peaked at ~500 ms and was evident only in 6-8 and 12-15-month-olds. We found no significant categorical responses to cars in any of the infant age groups (**Fig 5E**).

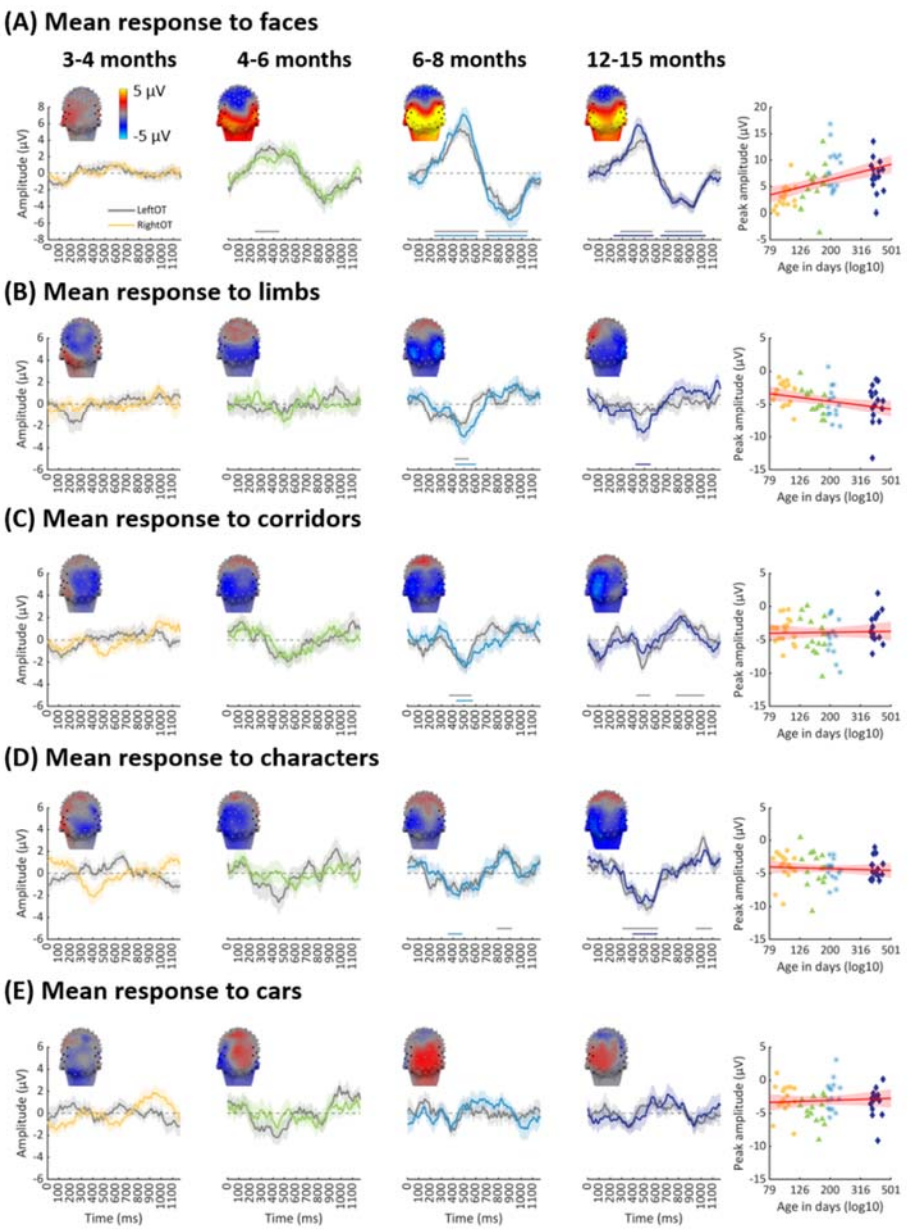

Fig. 5. Temporal dynamics of category-selective responses as a function of age. Category-selective responses to (A) faces, (B) limbs, (C) corridors, (D) characters, and (E) cars over left and right occipitotemporal ROI. Data are filtered to the oddball frequency and its harmonics in each individual, then averaged across electrodes of an ROI and across individuals. Left four panels in each row show the responses in the time domain, for the four age groups. *Colored lines:* mean responses in the right occipitotemporal ROI. *Gray lines:* mean responses in the left occipitotemporal ROI. *Colored horizontal lines above x-axis:* significant responses relative to zero for the right OT ROI. *Gray horizontal lines above x-axis:* significant responses relative to zero for the left OT ROI. *Top*: 3-D topographies of the spatial distribution of the response to oddball stimuli at a 483-500 ms time window after stimulus onset. *Right panel in each row:* amplitude of the peak deflection defined in a 400-700 ms time interval after stimulus onset. Each dot is a participant; Dots are colored by age group. Red *line:* linear mixed model (LMM) estimate of peak amplitude as a function of log10(age). *Shaded area:* 95% CI.

Further analyses of the two lateral occipitotemporal ROIs examined if there are developments of peak response amplitude and latency. LMM of peak response amplitude as function of age (log transformed) and category; participant, random effect reveal a significant category × age interaction in the right occipitotemporal ROI ($\beta_{category\ x\ age}$ = -1.1, 95% CI – -2.0 to -0.1, $t_{(301)}$ = -2.3, p_{FDR} < .05) suggesting differential development across categories. Post-hoc

analyses performed separately for each category (LMM of peak response as function of age (log transformed); participant, random effect) revealed significantly age-related increases response amplitude for faces in the right occipitotemporal ROI (β_{age} = 7.3, 95% CI: 4 to 10.5, $t_{(59)}$ = 4.5, p_{FDR} < .05) and a weaker age-related decreases of peak response amplitude for limbs in the right occipitotemporal ROI (significance did not survive FDR correction, β_{age} = -2.9, 95% CI: -5.4 to -0.4, $t_{(59)}$ = -2.3, p < .05; **Fig 5**-right). Analysis of the latency of the peak response (LMM as a function of age (log transformed) and category; participant, a random effect) found a significant decrease of peak latency in the right occipitotemporal ROI (β_{age} = -173.2, 95% CI: -284.7 to -61.6, $t_{(301)}$ = -3.1, p < .01), suggesting that the latency of the peak response shortens for all categories from 3 to 15 months of age. Other developmental trends were not significant (Supplementary **Table S3 and Table S4**).

Are categorical responses consistent across infants?

To determine whether the spatiotemporal pattern of brain responses evoked by different visual categories are distinct from one another and are reliable across infants, we used a leave-one-out infant-cross-validation (LOOCV) classifier approach (**Fig 6A**). At each iteration, the classifier was trained with the spatiotemporal patterns of response of each category using data from N-1 infants and then tested on how well it could predict the category the left-out infant was viewing from their brain responses. The spatiotemporal response was the mean response over the occipital and lateral occipitotemporal ROIs at the oddball frequency and its harmonics. The classifier was trained and tested separately for each age group.

Results reveal two main findings. First, we find significant development of reliable category spatiotemporal responses from 3 to 15 months of age (**Fig 6B**). That is, the LOOCV classifier decoded category information from brain responses significantly above the 20% chance level in infants aged 6 months and older but not in the younger 3-6-month-olds (6-8-month-olds, $t_{(14)}$ = 4.1, p < .01, Bonferroni corrected over 4 age groups, one-tailed; 12-15-month-olds, $t_{(14)}$ = 3.4, p < .01, Bonferroni corrected, one-tailed). Second, examination of classification by category, suggests that the development of a reliable spatiotemporal response to faces is dominant in driving the LOOCV classifier's performance (**Fig 6C**). While the LOOCV classifier could not decode

any category above chance in 3-4-month-olds, the classifier successfully determined from spatiotemporal responses when infants were viewing faces in 64% of 4-6-month-olds, in 93% of 6-8-month-olds, and 87% of 12-15-month-olds (**Fig 6C**). In contrast, classification was lower for other categories, but nonetheless above chance, especially for the 12-15-month-olds.

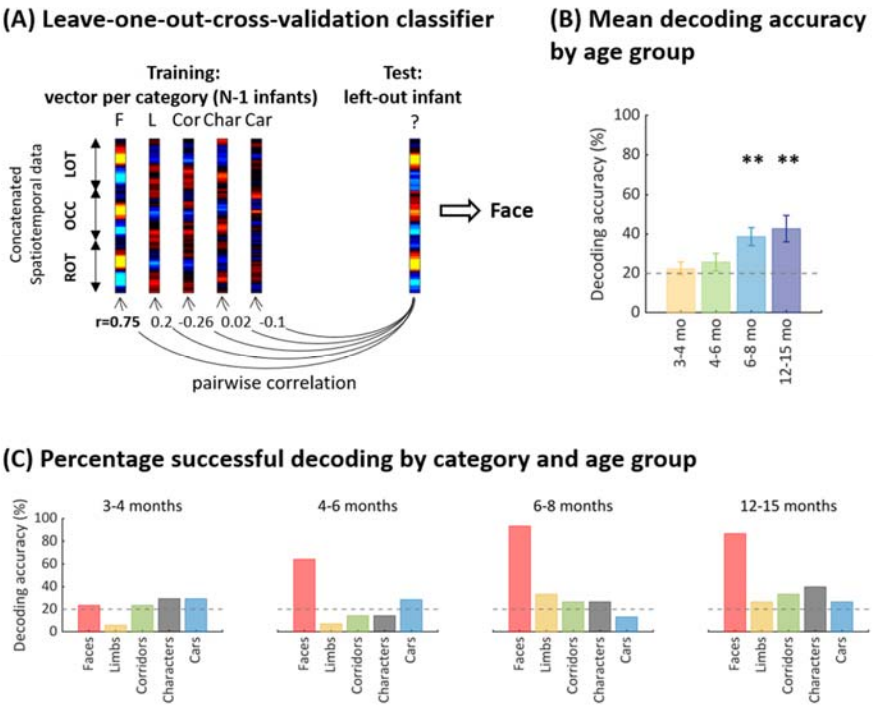

Fig. 6. Leave-one-out infant-cross-validation (LOOCV) reveals successful decoding of faces staring from 4 months of age. (A) An illustration of winner-takes-all LOOCV using the spatiotemporal response patterns of each category. Spatiotemporal patterns of response for each category are generated by concatenating the mean time courses from three ROIs: left occipitotemporal (LOT), occipital (OCC), and right occipitotemporal (ROT). At each iteration, we train the classifier with the mean spatiotemporal patterns of each category from N-1 infants, and test how well it predicts the category the left-out infant is viewing from their spatiotemporal brain response. The classifier is based on the highest pairwise correlation between the training and testing vectors. (B) Mean decoding accuracies across all five categories in each age group. *Asterisks:* significant decoding above chance level ($p < .01$, Bonferroni corrected, one-tailed). (C) Proportion of successful decoding of each category across infants of each age group. *Dashed lines*: chance level.

What is the nature of categorical spatiotemporal patterns in individual infants?

While the prior analyses leverage the power of averaging across electrodes and infants, averaging across electrodes of an ROI masks the intricate spatiotemporal patterns of response across electrodes and does not provide insight to representations within individual infants. To examine the finer-grain representation of category information within each infant's brain we examined spatiotemporal responses patterns across all 23 electrodes in left occipitotemporal (LOT), occipital (OCC), and right occipitotemporal (ROT) cortex in each individual. We tested: (i) if categorical representations in an infant's brain are reliable across different images of a category using split-half analyses of the data, and (ii) if category representations become more distinct

during the first year of life. We predicted that if representations become more similar across items of a category and more dissimilar between items of different categories then category distinctiveness would increase from 3 to 15 months of age.

Representation similarity matrices (RSM, **Fig 7A**) were generated from odd/even split-halves of the data in each infant. RSM cells on the diagonal represent within-category similarity (correlation, *r*) of spatiotemporal responses, while cells off the diagonal represent between-category similarity. We reasoned that reliable responses to different items of a category will be coupled with positive within-category similarity. Consistent with the LOOCV analyses, within individual infants there is no reliable category information in 3-4-month-olds as there is close to zero similarity between responses to different images of the same category (on-diagonal, **Fig 7A**, r: -0.03 ± 0.06, $p = .09$). At 4-6 months of age there is weak emergence of reliable patterns of responses to faces (r: 0.07 ± 0.17, $p = .14$, **Fig 7A**). Responses to faces becomes more reliable and significant above zero in 6-8-month-olds (r: 0.31 ± 0.23, $p < .001$), and stay reliable in 12-15-month-olds- (r: 0.26 ± 0.24, $p < .001$).

As the RSMs suggest that additional category information may be emerging (e.g., for limbs in 12-15-month-olds, **Fig 7A**), we evaluated category distinctiveness (within-category minus between-category similarity) for each infant and category and tested if it changed with age. Positive distinctiveness indicates that spatiotemporal patterns of response are more similar for different items within a category than between categories. Individual infants' category distinctiveness is shown in **Fig 7B-top** (infants ordered by age). Category distinctiveness increases with age and becomes more positive from 84 to 445 days of age (**Fig 7B**). The biggest increase, as expected, is for faces, where after ~6 months of age (194 days) face distinctiveness is consistently positive in individual infants (13/15 infants aged 194-364 days and 12/15 infants aged 365-445 days). The increase in distinctiveness is more modest for other categories and appears later in development. For example, positive distinctives for limbs in individual infants is consistently observed after 12 months of age (5/15 infants aged 194-364 days and 9/14 infants aged 365-445 days, **Fig 7B**).

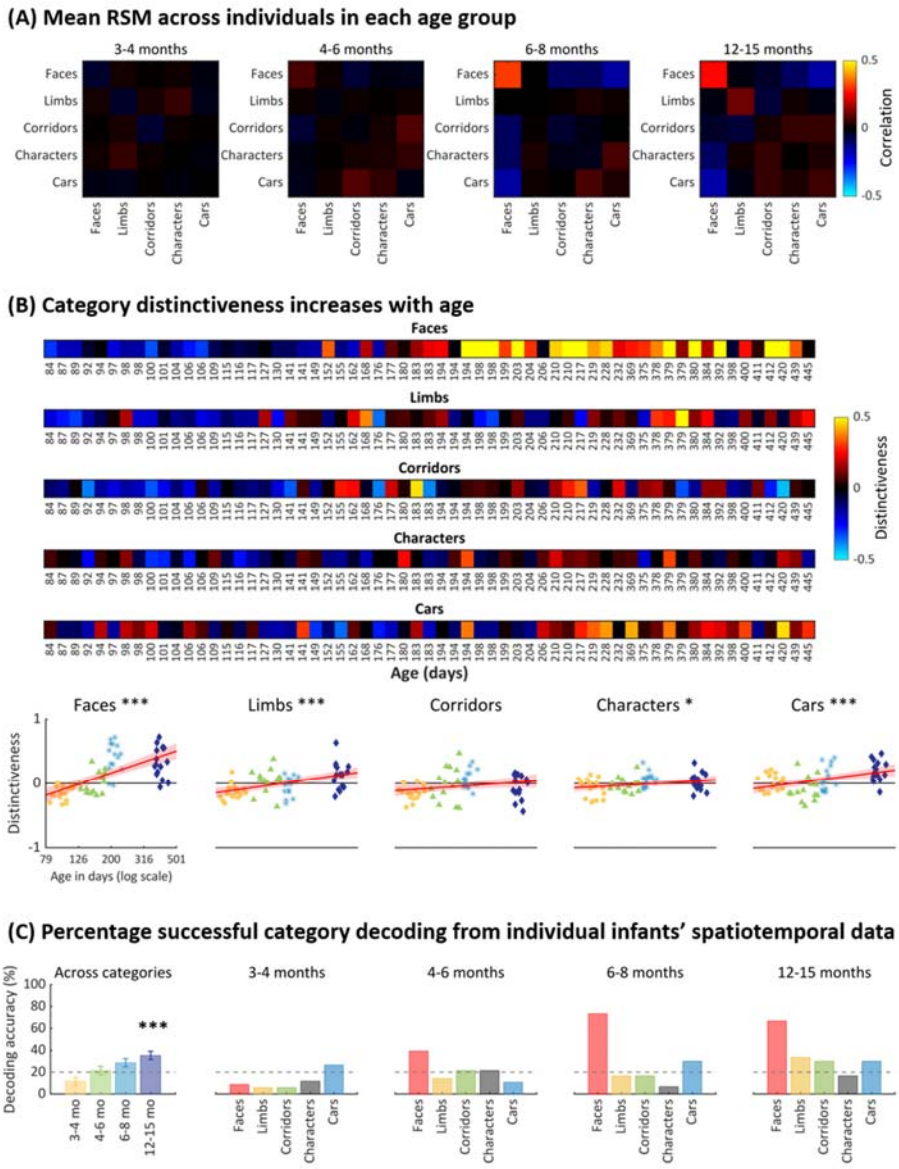

Fig.7. Individual split-half spatiotemporal pattern analyses reveal category information slowly emerges in the visual cortex after 6 months of age. (A) Representation similarity matrices (RSM) generated from odd/even split-halves of the spatiotemporal patterns of responses in individual infants. Spatiotemporal patterns for each category are generated by concatenating the mean time courses of each of 23 electrodes across left occipitotemporal (LOT), occipital (OCC), and right occipitotemporal (ROT). (B) *Top panel:* Category distinctiveness calculated for each infant and category, by subtracting the mean between-category correlation values from the within-category correlation value. *Bottom panel:* Distinctiveness as a function of age; panels by category; Each dot is a participant. Dots are colored by age group. *Red line:* linear mixed model (LMM) estimates of distinctiveness as a function of log10(age). *Shaded area:* 95% CI. (C). Decoding accuracy calculated with a winner-takes-all classifier of split-halves of the data. *Left panel:* mean decoding accuracy across five categories per group. *Right four panels:* proportion of successful decoding of each category across individuals in each age group. *Dashed lines:* chance level.

An LMM examining distinctiveness as a function of age (log transformed) and category (participant as random factor) revealed a significant interaction between age and category ($\beta_{age \times category}$ = -0.13, 95% CI: -0.2 to -0.06, $t_{(301)}$ = -3.61, p < .001), as well as significant main effects of age (β_{age} = 0.77, 95% CI: 0.54 to 1.00, $t_{(301)}$ = 6.62, p = 1.67×10^{-10}) and category ($\beta_{category}$ = 0.27, 95% CI: 0.11 to 0.43, $t_{(301)}$ = 3.38, p < .001). Further analyses for each category (**Fig 7B-bottom**) reveal a significant main effect of age for face distinctiveness (β_{age} = 0.9, 95% CI: 0.6 to 1.1, $t_{(59)}$ =

6.8, p_{FDR} < .001), limb distinctiveness (β_{age} = 0.4, 95% CI: 0.2 to 0.6, $t_{(59)}$ = 5.0, p_{FDR} < .001), character distinctiveness (β_{age} = 0.2, 95% CI: 0.02 to 0.3, $t_{(59)}$ = 2.2, p_{FDR} < .05), and cars distinctiveness (β_{age} = 0.4, 95% CI: 0.2 to 0.5, $t_{(59)}$ = 3.7, p_{FDR} < .001). This suggests that category distinctiveness slowly emerges in the visual cortex of infants from 3 to 15 months of age, with the largest and fastest development for faces.

Finally, we used a winner-takes-all classifier on split-halves of the data to test if the category an infant is viewing can be decoded from their spatiotemporal response pattern across 23 electrodes over bilateral occipitotemporal and occipital regions. The classifier was able to decode the category significantly above chance only in 12-15-month-olds (**Fig 7C-left**, $t_{(14)}$ = 2.98, p < .05, Bonferroni corrected, one-tailed). Examination of classification performance by category revealed overall better performance in 12-15-month-olds compared to younger infants, with numerically above chance classification of faces, limbs, corridors, and cars (**Fig 7C**). Again, the best classification was of faces. A classifier trained on infant's brain data can classify from an infant's brain responses if they are looking at faces in 76.7% of 6-8-month-olds and 73.3% of 12-15-month-olds.

Discussion

We find that both category-selectivity and the distinctiveness of distributed visual category representations progressively develop from 3 to 15 months of age. Notably, we find a differential development of category representations, whereby reliable face representations appear to emerge earliest at around 4-6 months of age. Category-representations to limbs, cars, corridors, and characters follow, emerging after 6 months of age. These findings suggest that representations for different everyday categories are learned and develop at different times during the first year of life. Together these data suggest a rethinking of the development of category representations during infancy.

Reliable category representations start to emerge at 4 months of age

While 3-4 months old infants have significant and reliable visual responses over early visual cortex, we find no reliable category representations of faces, limbs, cars, corridors, or characters

in these young infants. Our data suggest that category representations only start to reliably emerge at 4-6 months of age and progressively strengthen during the first 15 months of age.

Is it possible that there are some category representations in 3-4-month-olds, but we lack the sensitivity to measure them? We believe this is unlikely, because (i) we can measure significant visual responses from the same 3-4-month-olds, (ii) with the same amount of data, we can measure category selective responses and significant decoding of category information in infants older than 4 months and in adults (which have lower response amplitudes), and (iii) our results are largely consistent with other findings of immature but above chance decoding of category information from distributed EEG responses in 6-8-month-olds[20] and 12-15-month-olds[17].

We found some similarities and some differences between our finding of category-selective responses and a prior fMRI study in 2-10-month-olds[19]. We found the earliest reliable category-selective responses for faces only in 4-6-month-olds and category-selective responses to limbs and corridors only after 6 months of age, while [19] reported category-selective responses to faces, bodies, and scenes in 2-10-month-olds. Nonetheless, it is interesting that [19] found significant face-selective and place-selective responses in infants' ventral temporal cortex (VTC) but not in lateral occipital cortex (LOTC), and they found significant body-selective responses in LOTC, but not VTC[19]. Additionally, we note that there are also differences in stimuli across studies: in other studies infants typically observed colored, isolated objects for 1-2 seconds. However, in our study infants observed grayscale images of objects on a phase-scrambled textured background, which were controlled for low level features such as contrast and shown for 233ms. Thus, differences across studies may be due to differences in the visual presentation formats. For example, these differences raise the intriguing possibility that category-selective responses in VTC of infants younger than 6 months of age may be contingent on visual inputs consisting of isolated, colored, and moving items, which is different than category-selective responses in adults that generalize across visual formats[42–44]. Thus, a question for future research is whether category representations in young infants generalize across formats or are format specific.

Face representation emerge around 4-6 months of age

Recognizing faces (e.g., a caregiver's face) is crucial for infant's daily lives. Converging evidence from many studies suggest that infants have significant and reliable face-selective responses at 4-6 months of age[18,15,45,46,16]. While some studies report responses to face-like (high contrast paddle-like) stimuli[25,47,24] in newborns and significant visual evoked responses to faces in 3 month olds[48–51], these studies have largely compared responses to an isolated face vs. a single other isolated object. In contrast, we do not find reliable face-selective responses or reliable distributed representations to faces in 3-4-month-olds when responses to faces are contrasted to many other items and when stimuli are shown on a background rather than in isolation. Therefore, our data are consistent with longitudinal research in macaques showing that robust cortical selectivity to faces takes several months to emerge[27] and support the argument that experience with faces is necessary for the development of cortical face selectivity[26,27].

Our data also reveal that face-selective responses and distributed representations to faces become more robust in 6-8-month-olds and remain robust in 12-15-month-olds. For example, after 6 months of age it is possible to decode if the infant is looking at a face or something else from their distributed responses over the scalp. Successful decoding of faces was observed in 80% of individual infants based on several minutes of EEG data. This robust decoding has important clinical ramifications as it may serve as an early biomarker for cortical face processing, which is important for early detection of social and cognitive developmental disorders such as Autism[52,53] and Williams Syndrome[54]. Future research is necessary for elucidating the relationship between the development of brain responses to faces to infant behavior. For example, it is interesting that at 6 months of age, when we find robust face representations, infants also start to exhibit recognition of familiar faces (like parents) and stranger anxiety[55].

One fascinating aspect of the development of cortical face selectivity is that among the categories we tested, selectivity to faces seems to emerge the earliest at around 4 months of age, yet the development of selectivity and distributed representations to faces is protracted compared to objects and places [4,56,5]. Indeed, in both our data and prior work face-selective responses and distributed representations to faces in infants are immature compared to adults[14,20], and a large body of work has shown that face selectivity[57,58,4,7,56,5] and distributed

representations to faces[9] continue to develop during childhood and adolescence. This suggests that not only experience during infancy but also life-long experience with faces, sculpts cortical face selectivity. We speculate that the extended cortical plasticity for faces may be due to both the expansion of social circles (family, friends and acquaintances) across the lifespans and also the changing statistics of faces we socialize with (e.g., child and adult faces have different appearance).

New insight about development: different category representations emerge at different times

A key finding of our study is that the development of category selectivity during infancy is non-uniform: face representations develop before representations to limbs, places, and characters. We hypothesize that this differential development of visual category representations may be due to differential visual experiences with these categories during infancy. This hypothesis is consistent with behavioral research using head-mounted cameras that revealed that the visual input during early infancy is dense with faces, while hands become more prevalent in the visual input later in development and especially when in contact with objects [28,59]. Future studies that examine the visual diet and brain development in the same children can test this hypothesis, and test how environmental and experiential differences may influence infants' developmental trajectories. Additionally, future studies will be important to determine the emergence of other behaviorally relevant object categories present in infants' environments, such as food[60–62].

Perhaps surprisingly, we observed selectivity and distributed responses to characters in 12-15-month-olds, despite our participants being unable to read. Letter representations have been reported in 4-6 years olds[57] and word representations reliably emerge, when children learn to read[6] and continue to develop into adolescence[63,7,9]. Ours is the first study to examine (and find) character selectivity in infancy. What might drive character selectivity in infants? Prior work on writing systems has suggested that written scripts may leverage topological features and orientation biases that are common in natural images[64,65] and domain-general visual features learned from natural images can be used to scaffold domain-specific learning of letters with only limited exposure to print[66]. This suggests the possibility that exposure to natural scenes, perhaps

in combination with some exposure to print may be sufficient to generate a character-selective response in the infant brain by 12-15 months of age.

In sum, our study shows that category information starts develops to develop in infants' brains after 4 months of age, with face representations developing before representations to other everyday categories such as limbs or places. These data suggest not only that experience is necessary for the development of category representations, including faces, but also generates a rethinking of how category representations develop in infant. Specifically, our results raise the intriguing hypothesis that differential development of category information in infants' brains may be influenced by differential visual experiences with items from different categories. Future longitudinal studies could combine EEG or fMRI techniques to track developmental trajectories of object categorization within individual infants and explore the relationship between brain development and recognition behavior in infants. Moreover, our findings have important implications for atypical brain development, where differences in the development of categorization for certain objects like faces may serve as an early diagnostic tool for conditions such as autism and Williams Syndrome. Third, our findings have important implications for any computational model of human vision, because future models would need to incorporate non-uniform development computational constraints during learning. Finally, we believe that our innovative methods and our findings of sequential development of category representations in infancy are relevant for large-scale studies of brain development, which impact policy making related to children's health and well-being. Overall, our study underscores the need to consider both visual experience and behavioral outcomes when studying object categorization in the infant brain.

Methods

Participants

Ethical permission for the study was obtained from the Institutional Review Board of Stanford University. Parents of the infant participants provided written informed consent before their first visit to the lab, and also provided written informed consent prior to each session if they came for

multiple sessions. Participants were paid 20$/hour for participation. Participants were recruited via ads on the social media platforms Facebook and Instagram.

Sixty-two full-term and typically developing infants were recruited to participate in this study. Twelve participants were part of an ongoing longitudinal study and came for several sessions spanning at least 3 months apart. Data from nineteen infants (nine 3-4-month-olds, six 4-6-month-olds, and eight 6-8-month-olds; among whom seven were longitudinal infants) were acquired in two visits within a two-week span to obtain a sufficient number of valid data epoch per participant. Participants' demographic information is reported in **Table 1.** The youngest infants were 3 months old, as the EEG setup requires the infants to be able to hold their head and look at the screen in front of them.

We also recruited 23 adults (14 females) to participate in the same study. All participants reported to have normal/corrected-to-normal vision. Written informed consent was acquired prior to test.

Table 1. Demographic Information

	3-4 MONTHS (N = 17)	**4-6 MONTHS (N = 14)**	**6-8 MONTHS (N = 15)**	**12-15 MONTHS (N = 15)**
Age at test (days)	84 - 117	127 - 183	194 - 232	369 - 445
Gender				
Female	7	7	6	4
Male	10	7	9	11
Race				
White	5	4	6	5
Black				
Asian		1	1	2
Mixed races	11	8	6	7
Unknown	1	1	2	1

Inclusion: We included in the analyses data from participants who had a least 20 valid epochs (1.17-s/epoch) per category, had few noise/muscular artifacts during the EEG recordings, and

had visual responses over the occipital electrodes (see below). *Exclusion:* We excluded data from (i) 5 infants due to an insufficient number of epochs, (ii) 2 infants who had no visual responses, (iii) 10 infants for which we had technical issues during data collection, and (iv) 3 adults due to excessive noise/muscular artifacts during EEG. In total, we report data from 45 infants (Table 1) and 20 adults (13 females, 19-38 years) that met our inclusion criteria.

Visual stimuli

Natural grayscale images of adult faces, limbs, corridors, characters, and cars are used as stimuli, with 144 images per category from the fLOC image database, https://github.com/VPNL/fLoc[34]. The size, view, and retinal position of the items varied, and the items were overlaid on phase-scrambled backgrounds that were generated from a randomly drawn image in the stimulus set to minimize low-level visual differences across categories (Fig. 1; see [34] for more details). The stimuli were presented on a gamma-corrected OLED monitor screen (SONY PVM-2451; SONY Corporation, Tokyo Japan) at a screen resolution of 1920 × 1080 pixels and a monitor refresh rate of 60 Hz. When viewed from 70 cm away, each image extended a visual angle of approximately 12°.

EEG Protocol

The experiments were conducted in a calm, dimly illuminated lab room. Stimuli were presented using custom stimulus presentation software with millisecond timing precision and minimal frame-dropping. During testing, infant participants were seated on their parent's laps in front of the screen at a distance of 70 cm. One experimenter stood behind the presentation screen to monitor the centration of visual fixation of the infant participant. The visual presentation was paused if the infant looked away from the screen and was continued when their attention to the center of the screen was regained.

A frequency-tagging paradigm[31,67] was used to measure brain responses. In the experiment, visual stimuli (referred to here as carrier images) were presented sequentially at a rate of 4.286 Hz (~233 ms per image) with no inter-stimulus interval during each 70-s sequence. Unique images from a given stimulus category (referred to here as oddball images) were inserted at every 5th image at a rate of 0.857 Hz (4.286/5 Hz). The presentation pattern can be illustrated as: OCCCCOCCCCOCCCC …. (O: oddball image; C: carrier image). Each experimental sequence was

composed of five 14-s long conditions which included 1.17-s stimulus fade-in and 1.17-s stimulus fade-out. In each condition, participants saw stimuli from 5 categories: faces, limbs, corridors, characters, and cars. The conditions differed as to which category was assigned to the oddball and which categories were assigned to the carrier. In each condition, different images from one category were randomly drawn and presented as oddball images and carrier images were randomly drawn from the other four categories (**Fig 1**). Images did not repeat within each sequence. The order of the category conditions was randomized within each 70-s sequence. To motivate the infants to fixate and look at the screen, we presented at the center of the screen small (~1°) colored cartoon images such as butterflies, flowers, ladybugs. For adults, we used a fixation cross of the same size instead of the cartoons, and asked the participants to fixate and indicate when the fixation's color changed by pressing a space bar key on a keyboard. We obtained between 2-12 different 70-s sequences per individual (e.g. 22.34 to 134 s per category of analyzed data). EEG measurements for infant participants continued until the infant no longer attended the screen. For adult participants, we acquired 12 sequences per individual (134 s per category).

EEG Acquisition

EEG data were recorded at 500 Hz from 128-channel EGI High-Density Geodesic Sensor Nets connected to a NetAmps 300 amplifier (Electrical Geodesics, Inc., Eugene, OR, USA) for the infant participants. For the adult participants, the EEG net was connected to a NetAmps400 amplifier. The EEG recording was referenced online to a single vertex electrode Cz and the channel impedance was kept below 50 KΩ.

Pre-processing

EEG recordings were offline down-sampled to 420 Hz and were filtered using a 0.03-50 Hz bandpass filter with custom signal processing software. For infant participants, channels with more than 20% of samples exceeding a 100-150 µV amplitude threshold were replaced with the average amplitude of its six nearest-neighbor channels. The continuous EEG signals were then re-referenced to the common average of all channels and segmented into 1.17-s epochs (i.e., one stimulation cycle of OCCCC). Epochs with more than 15% of time samples exceeding threshold (150 - 200 µV) were excluded further on a channel-by-channel basis[68]. For adult participants, the

two-step artifact rejection was performed with different criteria as EEG response amplitudes are lower in adults than infants[68]. EEG channels with more than 15% of samples exceeding a 30 µV amplitude threshold were replaced by the average value of their neighboring channels. Then the EEG signals were re-referenced to the common average of all channels and segmented into 1.17-s epochs. Epochs with more than 10% of time samples exceeding threshold (30 - 80 µV) were excluded further on a channel-by-channel basis[69].

Table 2 shows the number of epochs (i.e. 1.17-s) we acquired before and after data pre-processing summing across all five categories. We used data after pre-processing for further data analysis. For infants' pre-processed data, there were no significant epoch-level differences between age groups ($F_{(3,57)} = 1.5$, $p = .2$).

Table 2. Average number of valid epochs summed across all five categories for each age group before and after data pre-processing.

	3-4 MONTHS (N = 17)	4-6 MONTHS (N = 14)	6-8 MONTHS (N = 15)	12-15 MONTHS (N = 15)	ADULTS (N = 20)
Before pre-processing	281 (± 103)	270 (± 86)	346 (±111)	324 (±78)	600 (± 0)
After pre-processing	223 (± 89)	219 (± 73)	266 (± 91)	269 (± 77)	560 (± 37)
Ratio (after/before)	79%	81%	77%	83%	93%

Univariate EEG analyses

Both the carrier and oddball EEG responses are reported at the frequency and time domain over three regions-of-interests (ROIs): the occipito-temporal ROI (left hemisphere: channels 57, 58, 59, 63, 64, 65 and 68; right hemisphere channels: 90, 91, 94, 95, 96, 99, and 100) and the occipital ROI (channels 69, 70, 71, 74, 75, 76, 82, 83 and 89). These ROIs were selected a priori based on previously published studies[70]. Since some outer rim channels from the infant participants were consistently noisy, three outer rim channels (i.e., 73, 81, and 88) were not included in the occipital ROI for further data analysis for both infant and adult participants.

Frequency domain analysis

Individual participant's preprocessed EEG signals for each stimulus condition were averaged over two epochs that contained two cycles of 'OCCCC'. The time averaged were then converted

to the frequency domain at a frequency resolution of 0.427 Hz via a Discrete Fourier Transform (DFT). The frequency bins of interest are therefore at every other bin in the frequency spectrum.

Category responses (oddball frequency): The amplitude and phase of the response at the oddball frequency and its first harmonic (0.857 Hz, 1.714 Hz) were extracted from the frequency spectra within each condition.

Visual responses (carrier frequency): The amplitude and phase of the evoked response at the carrier frequency and its first 3 harmonics (4.286 Hz, 8.571 Hz, 12.857 Hz, and 17.143 Hz) were extracted from the frequency spectra for each condition.

The real and imaginary Fourier coefficients of both the oddball and carrier responses for each condition were averaged across participants (vector averaging), to obtain the group data estimate of the oddball and carrier responses, respectively. Thus, the real and imaginary coefficients for a given harmonic were separately averaged across participants. The amplitudes of response were then computed from the coherently averaged vector. The Hotelling's T2 statistic[71] was used to test whether the carrier and oddball responses (and their harmonics, separately) were significantly different from zero. We used the Benjamini & Hochberg false discovery rate control procedure to correct for the multi-level comparisons.

Time domain analyses

The preprocessed time domain EEG signals of each individual were further low-passed filtered with a 30Hz cut-off. Then we filtered the time series keeping only the frequency bins of interest. To do that, we first transformed the time series into frequency domain with a DFT on the averaged epochs (same as above).

Category (oddball) responses: We kept responses only at frequencies of interest (such as 0.857 Hz and its harmonics up to 30 Hz) and set responses in the other frequency bins to 0 and then we applied an inverse Fourier transformation to transform the data back to the time domain. We then segmented the time series into 1.17-s epochs and averaged across these epochs for each condition and each individual. The mean and standard error across participants were computed for each condition at each time-point.

Visual (carrier responses): A similar procedure was performed to extract the visual responses; the only difference is that frequency bins of interest are at 4.286 Hz and its harmonics up to 30

Hz. As the carrier response represent the general visual responses, the EEG temporal waveforms were similar across different category conditions (see **Fig 3**). We, therefore, averaged the carrier waveforms across all five category conditions and report this mean response.

To define time windows in which amplitudes were significantly different from baseline (zero) for each individual condition, we used a cluster-based nonparametric permutation t-test (10000 permutations, with a threshold of $p < 0.05$, two-tailed) on the post-stimulus onset time-points (0 – 1167 ms)[72,73]. The null hypothesis was that the evoked waveforms were not different from zero at any time point. For each permutation, we assigned random signs to the data of individual participants and computed the group-level difference against zero using a t-test. We then calculated the cluster-level statistics as the sum of t-values in the consecutive time points with p-values less than 0.05[74]. We took the maximum cluster-level statistic for each permutation to make a non-parametric reference distribution of cluster-level statistics. We rejected the null hypothesis if the cluster-level statistics for any consecutive sequence in the original data was larger than 97.5% or smaller than 2.5% of the values in the null distribution.

Decoding analyses

We performed decoding analyses both at the group and individual level.

(1) At the group level, we used a leave-one-out participant-cross-validation (LOOCV) approach to test if spatiotemporal responses patterns to each of the five categories were reliable across participants. The classifier was trained on averaged data from N-1 participants and tested on how well it could predict the category the left-out participant was viewing from their brain activations. This procedure was repeated for each left-out participant. As the visual responses were mainly distributed over the posterior regions of the scalp, we used averaged time courses across electrodes across three ROIs as the exact location of the electrode array varies across individuals. The three ROIs comprised 7 left occipitotemporal, 9 occipital, and 7 right occipitotemporal electrodes (the same as those used in the univariate analysis). We calculated the average response in the time domain filtered to the oddball frequency and its harmonics up to 30 Hz in each ROI and concatenated these to form a spatiotemporal response vector. This spatiotemporal vector was calculated separately for each category. At each iteration, the LOOCV classifier computed the correlation between each of these vectors from the left-out participant

(test data, for an unknown stimulus) and each of the mean spatiotemporal vectors across the N-1 participants (training data, labeled data). The winner-take-all (WTA) classifier classifies the test vector to the category that yields the highest correlation with the training vector. For a given test pattern, correct classification yielded a score of 1 and an incorrect classification yielded a score of 0. We computed group mean decoding performance across all N iterations for each category and the group mean decoding accuracies across five categories.

(2) For the individual participant level decoding analysis, we also used a WTA classifier to predict the category an individual was viewing from the spatiotemporal patterns over their scalp. Here, we used individual level data and a finer spatial granularity. The classifier was trained on one half of the data (i.e., odd or even trials) and tested on how well it could predict the other half of the data (i.e., even or odd trials) for each individual participant. As we are interested in spatiotemporal responses over visual cortex, we concatenated the time courses from the 23 electrodes located over occipital and lateral occipital cortex in each individual. The WTA classifier computed the correlation between the spatiotemporal pattern of a given test vector (for an unknown stimulus) and each of the spatiotemporal patterns of the labelled training vector (see cross-correlation matrix in an example 6 months old infant, Supplementary **Fig S8**). The WTA classifier classified the test data based on the training data that yielded the highest correlation with the test vector. For a given test pattern, correct classification yields a score of 1 and an incorrect classification yields a score of 0. Performance was computed for each category from all combinations of odd and even splits, yielding possible scores of 1 (both classifications were correct), 0.5 (only 1 was correct), or 0 (both incorrect). In each age group, we evaluated the mean classification performance across individuals for each category as well as mean classification performance across categories.

Category distinctiveness

To quantify how consistent spatiotemporal responses within a category (across different images) are and how different they are from spatiotemporal responses to different categories, we measured category distinctiveness for each category and individual. Distinctiveness is defined as the difference between the similarity (correlation) of spatiotemporal responses within a category across odd and even splits and the average between-category similarity (correlation) of

spatiotemporal responses across odd and even splits. Distinctiveness is higher when the within category similarity is positive and the between category similarity is negative and varies from -2 (no category information) to 2 (maximal category information). We computed category distinctiveness for each of the 5 categories as a function of age to quantify how the representation of categories develops in infants' brain from 3 to 15 months of age.

Statistical analyses of developmental effects

In the study, to quantify development in responses as a function of age, we used linear mixed models (LMMs) with the 'fitlme' function in MATLAB version 2021b (MathWorks, Inc.). We fit our data with random-intercept models in which we allow intercepts to vary across participants. In all the LMMs, we transformed the age variable into log scale as the development is faster early on than later on.

(1) For visual responses (filtered responses at 4.286 Hz and harmonics up to until 30 Hz), to quantify development of peak latency with age and time window as fixed factors, we used the following formula:

peak latency ~ 1 + log10(age in days) × time window + (1|participant),

in which peak waveform latency is the latency of the peak response in each individual at a certain time window (window 1: 60-90 ms or window 2: 90-160 ms for 3-4-month-olds, and 90-110 ms for other age groups), age in days is a continuous fixed factor, time window is a categorical fixed factor (window 1 or window 2), and the term (1|participant) represents the individual level random intercept. The formula examines fixed effects including the main effects of age and time window as well as the interaction of age × time window, and the random effect.

As we found a significant interaction between age × time window for the peak latency of visual responses, to assess the effect of age in each time window, we further ran separate LMMs on the peak latency in each time window with age in days (log transformed) as a fixed factor (Supplementary **Table 1**):

Peak latency ~ 1+ log10(age in days) + (1|participant).

We used a similar LMM to assess the development of amplitude of the peak response and the results are reported in the Supplementary **Table 2**.

(2) To assess the developmental of category responses we examined the effect of age on the peak response and if it varied by category we used the following LMMs:

Peak amplitude ~ 1+ log10(age in days) × category + (1|participant).

in which peak waveform amplitude is the amplitude of the peak response in each individual at a certain time window (400-700 ms), age in days is a continuous fixed factor, category is a categorical fixed factor (faces, limbs, corridors, cars and characters), and the term (1|participant) represents the individual level random intercept. The formula examines fixed effects including the main effects of age and category as well as the interaction between age × category, and the random effect.

As we found a significant interaction between age × category for the peak amplitude, we further assessed development separately for each category, using LMMs that related peak amplitude to age (log transformed, fixed effect) and participant (random intercept) (Supplementary **Table 3**):

Peak amplitude ~ 1+ log10(age in days) + (1|participant).

We used similar LMM to assess the development of the latency of the peak response and the results are reported in the Supplementary **Table 4.**

(3) To examine the development of category distinctiveness, and assess if it varied by category, we used the following LMM:

Distinctiveness value ~ log10(age in days) × category + (1|participant).

As we found a significant interaction, we then fit separate LMMs relating distinctiveness to age separately for each category:

Distinctiveness value ~ logg10(age in days) + (1|participant).

Analyses of adult data

In the study, we also collected data from 20 adults with the same experimental task (**Fig 1**). By using the same number of trials in adults as those obtained from our infant participants (an average of 49 epochs per category per individual) for data analyses, our goal was to test that we had sufficient power to detect categorical responses from infants using the experimental paradigm. We expect temporal and amplitude differences between adults and infants[20] as infants have immature brains and skulls. For example, cortical gyrification, which determines the

orientation of the electrical fields generated in certain part of the brain region on the scalp, still undergoes development during the first two years of infants' lives[75]. Second, adults' skulls are thicker and have lower conductivity than infants' skulls, thus electrical signals on their scalp are lower than infants[76]. Nonetheless, we tested if we could in principle detect category information in adults with the same amount of data as infants. We reasoned that if category information can be detected in adults and signals are stronger in infants then we should have the power to detect category information in infants.

Acknowledgements

This work was supported by grants from the Wu Tsai Neurosciences Institute of Stanford University, the Human Centered Artificial Intelligence Institute of Stanford University to KGS and AMN.

Author contributions

XY designed the experiment, coded the analysis pipeline, collected the data, analyzed data, and wrote the paper. AMN designed the experimental protocol and provided funding and instrumentation. SF and BF recruited the participants. ST, BF, and YC collected the data. ST and BF equally contributed to the study. KGS designed the experiment, oversaw the data collection and analyses, provided funding, and wrote the paper. All authors read the manuscript, provided critical revisions and approved the final paper for submission.

Competing interests

The authors declare no competing interests.

References

1. Kanwisher, N., McDermott, J. & Chun, M. M. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience* **17**, 4302–4311 (1997).
2. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
3. Downing, P. E., Jiang, Y., Shuman, M. & Kanwisher, N. A Cortical Area Selective for Visual Processing of the Human Body. *Science* **293**, 2470–2473 (2001).
4. Golarai, G. *et al.* Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience* **10**, 512–522 (2007).
5. Scherf, K. S., Behrmann, M., Humphreys, K. & Luna, B. Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science* **10**, F15–F30 (2007).
6. Dehaene-Lambertz, G., Monzalvo, K. & Dehaene, S. The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology* **16**, e2004103 (2018).
7. Nordt, M. *et al.* Cortical recycling in high-level visual cortex during childhood development. *Nat Hum Behav* **5**, 1686–1697 (2021).
8. Haxby, J. V. *et al.* Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science* **293**, 2425–2430 (2001).

9. Nordt, M. *et al.* Longitudinal development of category representations in ventral temporal cortex predicts word and face recognition. 2022.12.23.521732 Preprint at https://doi.org/10.1101/2022.12.23.521732 (2022).

10. Mondloch, C. J. *et al.* Face Perception During Early Infancy. *Psychol Sci* **10**, 419–422 (1999).

11. Spriet, C., Abassi, E., Hochmann, J.-R. & Papeo, L. Visual object categorization in infancy. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2105866119 (2022).

12. Quinn, P. C., Eimas, P. D. & Rosenkrantz, S. L. Evidence for Representations of Perceptually Similar Natural Categories by 3-Month-Old and 4-Month-Old Infants. *Perception* **22**, 463–475 (1993).

13. Younger, B. A. & Fearing, D. D. Parsing Items into Separate Categories: Developmental Change in Infant Categorization. *Child Development* **70**, 291–303 (1999).

14. Conte, S., Richards, J. E., Guy, M. W., Xie, W. & Roberts, J. E. Face-sensitive brain responses in the first year of life. *NeuroImage* **211**, 116602 (2020).

15. Farzin, F., Hou, C. & Norcia, A. M. Piecing it together: Infants' neural responses to face and object structure. *Journal of Vision* **12**, 6 (2012).

16. de Heering, A. & Rossion, B. Rapid categorization of natural face images in the infant right hemisphere. *eLife* **4**, e06564 (2015).

17. Bayet, L. *et al.* Temporal dynamics of visual representations in the infant brain. *Developmental Cognitive Neuroscience* **45**, 100860 (2020).

18. Deen, B. *et al.* Organization of high-level visual cortex in human infants. *Nat Commun* **8**, 13995 (2017).

19. Kosakowski, H. L. *et al.* Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology* **32**, 265-274.e5 (2022).

20. Xie, S. *et al.* Visual category representations in the infant brain. *Current Biology* **32**, 5422-5432.e6 (2022).

21. Kanwisher, N. Domain specificity in face perception. *Nat Neurosci* **3**, 759–763 (2000).

22. Dehaene, S. *et al.* How Learning to Read Changes the Cortical Networks for Vision and Language. *Science* **330**, 1359–1364 (2010).

23. Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. & Fabre-Grenet, M. Mother's face recognition by neonates: A replication and an extension. *Infant Behavior & Development* **18**, 79–85 (1995).

24. Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19 (1991).

25. Buiatti, M. *et al.* Cortical route for facelike pattern processing in human newborns. *Proceedings of the National Academy of Sciences* **116**, 4625–4630 (2019).

26. Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R. & Livingstone, M. S. Seeing faces is necessary for face-domain formation. *Nat Neurosci* **20**, 1404–1412 (2017).

27. Livingstone, M. S. *et al.* Development of the macaque face-patch system. *Nat Commun* **8**, 14897 (2017).

28. Fausey, C. M., Jayaraman, S. & Smith, L. B. From faces to hands: Changing visual input in the first two years. *Cognition* **152**, 101–107 (2016).

29. Long, B. *et al. The BabyView Camera: Designing a New Head-mounted Camera to Capture Children's Early Social and Visual Environment*. (2022) doi:10.31234/osf.io/238jk.

30. Barry-Anwar, R., Hadley, H., Conte, S., Keil, A. & Scott, L. S. The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. *Neuropsychologia* **108**, 25–31 (2018).

31. Farzin, F., Hou, C. & Norcia, A. M. Piecing it together: Infants' neural responses to face and object structure. *Journal of Vision* **12**, 6 (2012).

32. Heinrich, S. P., Mell, D. & Bach, M. Frequency-domain analysis of fast oddball responses to visual stimuli: A feasibility study. *International Journal of Psychophysiology* **73**, 287–293 (2009).

33. Liu-Shuang, J., Norcia, A. M. & Rossion, B. An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia* **52**, 57–72 (2014).

34. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *Journal of Neuroscience* **35**, 12412–12424 (2015).

35. Cohen, L. *et al.* The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307 (2000).

36. Peelen, M. V. & Downing, P. E. Selectivity for the Human Body in the Fusiform Gyrus. *Journal of Neurophysiology* **93**, 603–608 (2005).

37. Bugatus, L., Weiner, K. S. & Grill-Spector, K. Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage* **155**, 437–449 (2017).

38. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* **2**, (2008).

39. Margalit, E. *et al.* Ultra-high-resolution fMRI of Human Ventral Temporal Cortex Reveals Differential Representation of Categories and Domains. *J. Neurosci.* **40**, 3008–3024 (2020).

40. Norcia, A. M. & Tyler, C. W. Spatial frequency sweep VEP: Visual acuity during the first year of life. *Vision Research* **25**, 1399–1408 (1985).

41. Norcia, A. M., Tyler, C. W. & Hamer, R. D. Development of contrast sensitivity in the human infant. *Vision Research* **30**, 1475–1486 (1990).

42. Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y. & Malach, R. Cue-Invariant Activation in Object-Related Areas of the Human Occipital Lobe. *Neuron* **21**, 191–202 (1998).

43. Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R. & Ungerleider, L. G. Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron* **22**, 751–761 (1999).

44. Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K. & Tootell, R. B. H. The Representation of Illusory and Real Contours in Human Cortical Visual Areas Revealed by Functional Magnetic Resonance Imaging. *J. Neurosci.* **19**, 8560–8572 (1999).

45. Guy, M. W., Zieber, N. & Richards, J. E. The Cortical Development of Specialized Face Processing in Infancy. *Child Dev* **87**, 1581–1600 (2016).

46. Halit, H., Csibra, G., Volein, Á. & Johnson, M. H. Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry* **45**, 1228–1234 (2004).

47. Goren, C. C., Sarty, M. & Wu, P. Y. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* **56**, 544–549 (1975).

48. Halit, H., de Haan, M. & Johnson, M. H. Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage* **19**, 1180–1193 (2003).

49. Macchi Cassia, V., Kuefner, D., Westerlund, A. & Nelson, C. A. A behavioural and ERP investigation of 3-month-olds' face preferences. *Neuropsychologia* **44**, 2113–2125 (2006).

50. Peykarjou, S. Frequency tagging with infants: The visual oddball paradigm. *Front. Psychol.* **13**, 1015611 (2022).

51. Tzourio-Mazoyer, N. *et al.* Neural Correlates of Woman Face Processing by 2-Month-Old Infants. *NeuroImage* **15**, 454–461 (2002).

52. Rossion, B. Biomarkers of Face Perception in Autism Spectrum Disorder: Time to Shift to Fast Periodic Visual Stimulation With Electroencephalography? *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **5**, 258–260 (2020).

53. Vettori, S. *et al.* Reduced neural sensitivity to rapid individual face discrimination in autism spectrum disorder. *NeuroImage: Clinical* **21**, 101613 (2019).

54. Farran, E. K. *et al.* Characterizing the neural signature of face processing in Williams syndrome via multivariate pattern analysis and event related potentials. *Neuropsychologia* **142**, 107440 (2020).

55. Kobayashi, M., Kakigi, R., Kanazawa, S. & Yamaguchi, M. K. Infants' recognition of their mothers' faces in facial drawings. *Dev Psychobiol* **62**, 1011–1020 (2020).

56. Peelen, M. V., Fei-Fei, L. & Kastner, S. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* **460**, 94–97 (2009).

57. Cantlon, J. F., Pinel, P., Dehaene, S. & Pelphrey, K. A. Cortical Representations of Symbols, Objects, and Faces Are Pruned Back during Early Childhood. *Cerebral Cortex* **21**, 191–199 (2011).

58. Cohen, M. A. *et al.* Representational similarity precedes category selectivity in the developing ventral visual pathway. *NeuroImage* **197**, 565–574 (2019).

59. Jayaraman, S., Fausey, C. M. & Smith, L. B. Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology* **53**, 38–49 (2017).

60. Bannert, M. M. & Bartels, A. Visual cortex: Big data analysis uncovers food specificity. *Current Biology* **32**, R1012–R1015 (2022).

61. Jain, N. *et al.* Selectivity for food in human ventral visual cortex. 2022.05.22.492983 Preprint at https://doi.org/10.1101/2022.05.22.492983 (2023).

62. Pennock, I. M. L. *et al.* Color-biased regions in the ventral visual pathway are food selective. *Current Biology* **33**, 134-146.e4 (2023).

63. Nordt, M. *et al.* Learning to Read Increases the Informativeness of Distributed Ventral Temporal Responses. *Cerebral Cortex* **29**, 3124–3139 (2019).

64. Changizi, M. A., Zhang, Q., Ye, H. & Shimojo, S. The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes. *The American Naturalist* **167**, E117–E139 (2006).

65. Morin, O. Spontaneous Emergence of Legibility in Writing Systems: The Case of Orientation Anisotropy. *Cognitive Science* **42**, 664–677 (2018).

66. Testolin, A., Stoianov, I. & Zorzi, M. Letter perception emerges from unsupervised deep learning and recycling of natural image features. *Nat Hum Behav* **1**, 657–664 (2017).

67. Jacques, C., Retter, T. L. & Rossion, B. A single glance at natural face images generate larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. *NeuroImage* **137**, 21–33 (2016).

68. Norcia, A. M., Pei, F. & Kohler, P. J. Evidence for long-range spatiotemporal interactions in infant and adult visual cortex. *Journal of Vision* **17**, 12 (2017).

69. Kohler, P. J., Barzegaran, E., Norcia, A. M. & McCandliss, B. D. Parietal Contributions to Abstract Numerosity Measured with Steady State Visual Evoked Potentials. 2020.08.06.239889 Preprint at https://doi.org/10.1101/2020.08.06.239889 (2020).

70. Xie, W., McCormick, S. A., Westerlund, A., Bowman, L. C. & Nelson, C. A. Neural Correlates of Facial Emotion Processing in Infancy. *Dev Sci* **22**, e12758 (2019).

71. Victor, J. D. & Mast, J. A new statistic for steady-state evoked potentials. *Electroencephalography and Clinical Neurophysiology* **78**, 378–388 (1991).

72. Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W. & Norcia, A. M. Cue-Invariant Networks for Figure and Background Processing in Human Visual Cortex. *J. Neurosci.* **26**, 11695–11708 (2006).

73. Blair, R. C. & Karniski, W. An alternative method for significance testing of waveform difference potentials. *Psychophysiology* **30**, 518–524 (1993).

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

75. Li, G. *et al.* Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. *J Neurosci* **34**, 4228–4238 (2014).

76. Grieve, P. G., Emerson, R. G., Fifer, W. P., Isler, J. R. & Stark, R. I. Spatial correlation of the infant and adult electroencephalogram. *Clinical Neurophysiology* **114**, 1594–1608 (2003).