Visual category representations in the infant brain

Authors: Siying Xie[1]*, Stefanie Hoehl[2,3], Merle Moeskops[1], Ezgi Kayhan[2,4], Christian Kliesch[2,5], Bert Turtleton[1], Moritz Köster[1,2]†, Radoslaw M. Cichy[1,6,7,8]*†

[1]Department of Education and Psychology, Freie Universität Berlin; 14195 Berlin, Germany.

[2]Max Planck Institute for Human Cognitive and Brain Sciences; 04103 Leipzig, Germany.

[3]Department of Developmental and Educational Psychology, University of Vienna; 1010 Wien, Austria.

[4]Department of Developmental Psychology, University of Potsdam; 14469 Potsdam, Germany.

[5]Department of Psychology, Johannes-Gutenberg-University Main; 55122 Mainz, Germany.

[6]Berlin School of Mind and Brain, Faculty of Philosophy, Humboldt-Universität zu Berlin; 10117 Berlin, Germany.

[7]Einstein Center for Neurosciences Berlin, Humboldt-Universität Berlin; 10117 Berlin, Germany.

[8]Bernstein Center for Computational Neuroscience Berlin, Humboldt-Universität Berlin; 10115 Berlin, Germany.

*Corresponding author. Emails: rmcichy@zedat.fu-berlin.de (RMC); siying.xie@outlook.com (SX).

†These authors contributed equally to this work.

Abstract

Visual categorization is a human core cognitive capacity that depends on the development of visual category representations in the infant brain. The nature of infant visual category representations and their relationship to the corresponding adult form however remain unknown. Our results clarify the nature of visual category representations in 6- to 8-month-old infants and their developmental trajectory towards adult maturity in the key characteristics temporal dynamics, representational format, and spectral properties. Temporal dynamics change from slowly emerging, developing representations in infants to quickly emerging, complex representations in adults. Despite those differences infants and adults already partly share visual category representations. The format of infants' representations are visual features of low to intermediate complexity, whereas adults' representations also encoded high complexity features. Theta band neural oscillations form the basis of visual category representations in infants, and these representations are shifted to the alpha/beta band in adults.

Introduction

The ability to recognize and categorize visual objects effortlessly and within the blink of an eye is a core human cognitive capacity[1,2] that develops through learning and interaction with the environment. Behavioral research in infants using looking times[3,4] and neural markers of attention provides evidence for visual category processing[5] and learning[6,7] already within the first year of life.

Fundamental research in human and non-human primates has described the nature of the neural representations underlying mature visual categorization abilities in adults, revealing their temporal dynamics [2,8], what features they encode[9–11], their cortical locus[11,12], and how they relate to neural oscillations[13,14].

We require comparable knowledge about visual category representations in the infant brain[15] to understand how the development of neural representations mediates the development of visual categorization abilities in humans.

While visual category representations have been investigated for decades using electroencephalography (EEG), the resulting insights were principally limited in two ways due to methodological shortcomings. One research approach focused on assessing the successful outcome of visual categorization rather than the underlying representations themselves[16]. Thus, the gained insights about representations are indirect. Another research approach did assess underlying representations directly, but was limited to the category of faces[17,18] for which known neural markers exist. Thus, the generalizability from the unique and small stimulus subset to the broad set of visual categories of the visual world remains unclear.

Here we overcome this double impasse to reveal the nature of general visual category representations for various object categories in 6- to 8-month-old infants. We do so by leveraging an integrated multivariate analysis framework of multivariate classification[8] and direct quantitative comparison[19] of infant to adult EEG data and deep learning models of vision[10]. Our results clarify a crucial developmental state of visual category representations in the human brain and its developmental trajectory toward the adult form in the key characteristics temporal dynamics, representational format and spectral properties.

Results

The temporal dynamics of visual category representations

Infant participants (n = 40) viewed 128 images of real-world objects from four categories (i.e., toys, bodies, houses and faces, **Fig. 1a, Supplementary Fig. 1,2**) while we acquired EEG data. Images were presented for 2s every 2.7–2.9s. For direct comparison we acquired EEG data in adult participants (n = 20) viewing the same stimulus set with an adapted experimental design (**Supplementary Fig. 2**). We consider the epoch of –100ms to +1,000ms with respect to stimulus onset in our analyses.

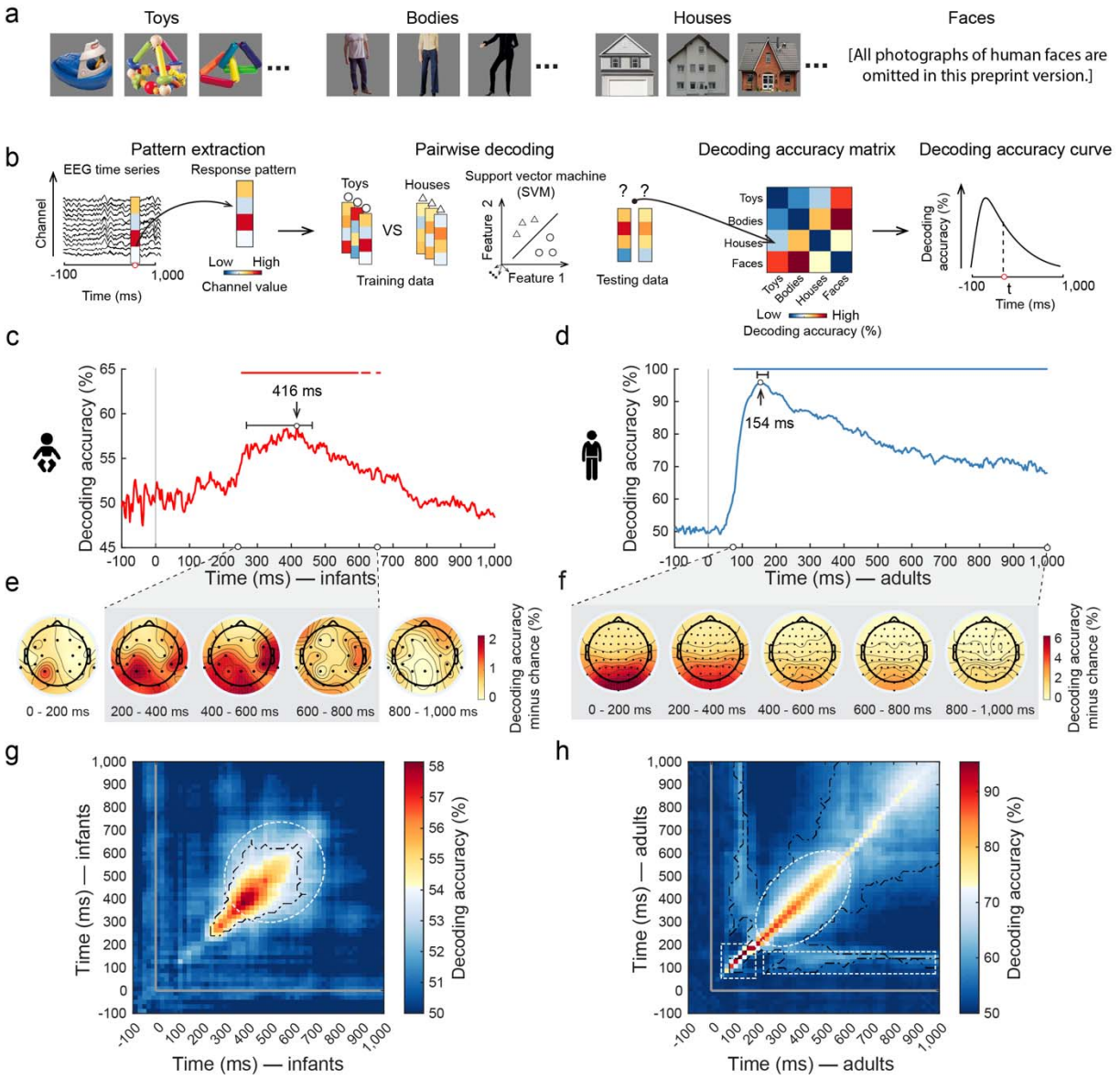

Figure 1 Experimental design and results of time-resolved multivariate analysis. (a) The stimulus set comprised 32 cut-out images from four categories each: toys, headless bodies, houses and faces (full set see **Supplementary Fig. 1**). We omitted all human face images in this version due to insufficient resources on the preprint server to check for consent for the public display of them. **(b)** Time-resolved multivariate analysis on EEG data. First, for every time point in the epoch, we extracted condition-specific EEG sensor activation values and formed them into response vectors. Then, using a leave-one-out cross-validation scheme we trained and tested a support vector machine to pairwise classify visual object category from the response vectors. The results (decoding accuracy, 50% chance level) were aggregated in a decoding accuracy matrix of size 4 × 4, indexed in rows and columns by the conditions classified. The matrix is symmetric along the diagonal, and the diagonal is undefined. Averaging the lower triangular part of the matrix resulted in grand average decoding accuracy as a measure of how well category is discriminated by visual representations at a particular time point. **(c, d)** The grand average time course of visual category decoding in infants **(c)** and adults **(d)**. **(e, f)** Results of category classification in EEG channel-space searchlight analysis for infants **(e)** and adults **(f)**. **(g, h)** Results of time-generalization analysis for infants **(g)** and adults **(h)**. Peak latencies and on- and offsets of significance with 95% confidence intervals (CIs) are listed in **Supplementary Table 1a,b**. Rows of asterisks indicate significant time points (infant n = 40, adult n = 20, right-

tailed sign permutation tests, cluster-defining threshold $P < .005$, corrected significance level $P < .05$). Horizontal error bars indicate 95% CIs. The gray vertical line indicates onset of image presentation.

To reveal the time course with which visual category is discriminated by visual representations we used time-resolved multivariate pattern analysis[8] (**Fig. 1b**). We report peak latency (95% confidence intervals (CIs) in brackets) as the time point during neural processing when category information was most explicit, as well as onset and offset of significance for each group.

In infants (**Fig. 1c**), the classification curve rose gradually from 100ms onwards, reaching significance at 252ms (250–254ms), followed by a broad peak at 416ms (268–462ms) and a gradual decline. This pattern of result did not depend on any particular object or category (**Supplementary Fig. 3a,b),** and emerged equivalently for alternative common analysis schemes (**Supplementary Fig. 3c–e**). In contrast, in adults the classification curve had a different shape (**Fig. 1d**). It emerged earlier (significant at 72ms (72–74ms)) and faster, peaking at 154ms (144–176ms) than in infants ($P < .001$, bootstrap test, **Supplementary Table 1a**). Searchlight analysis in EEG channel space revealed that information about visual category representations was highest in EEG channels overlying occipitoparietal cortex in both infants (**Fig. 1e**) and adults (**Fig. 1f**).

This multivariate approach constitutes a novel analytical access point to visual category representations in infants from EEG data and shows that visual category representations emerge in similar cortical structures, but slower and later in infants than adults.

The rise and fall of the classification curves in a few hundred milliseconds might indicate rapid changes in the underlying visual representations[2,8], or slow ramping up of persistent representations, or a combination of both. To investigate this, we assessed the temporal stability of visual representations using time-generalization analysis[20]. We determined how well classifiers trained on predicting visual category from EEG at one time point perform when tested at other time points. Lack of generalization across time indicates transience of the underlying visual representation, whereas generalization across time indicates persistence.

In both infants (**Fig. 1g**) and adults (**Fig. 1h**) classification accuracy was highest along the diagonal (i.e., similar time points for training and testing) with a broadening over time (white dotted ellipse). This suggests common neural mechanisms of a rapid sequence of processing steps that results in a final outcome that is held online for further use, indicated by rapidly changing transient representations at earlier time points and more slowly changing persistent representations at later time points respectively.

In addition to this general similarity between infants and adults, there were two notable differences, indicative, respectively, of incomplete development of feedforward and feedback information processing in infants. For one, early after stimulus onset when neural processing is dominantly feed-forward, in adults we observed high classification accuracy trailing the diagonal narrowly (**Fig. 1h**, 50–200ms, dotted square), indicating rapidly changing representations. Infants did not exhibit such signals. This pattern indicates incomplete development of feedforward visual information processing mechanisms in infants. Secondly, in adults the classifier generalized well for the time point combination of 100–200ms and 200–1,000ms (**Fig. 1h**, white striped rectangle). This suggests highly persistent representations, likely emerging in early visual cortex[8]. There were no such signals in infants. This indicates incomplete

development of neural structures for recurrent processing that maintain visual information online for long stretches of time.

Shared visual category representations between infants and adults

The results presented so far show that we are able to identify visual category representations in both infants and adults. We then investigated the similarity of the representations between the two groups. Identification of shared representations between infants and adults is challenging, due to differences in the time course over which the representations emerge, and because the EEG channel spaces differ. To overcome these hurdles we used a time-generalization variant of representational similarity analysis (RSA)[19] (**Fig. 2a**). In short, we abstracted multivariate signals from the incommensurate infant and adult EEG channel spaces to a common representational dissimilarity space, in which we compared the signals across all time point combinations.

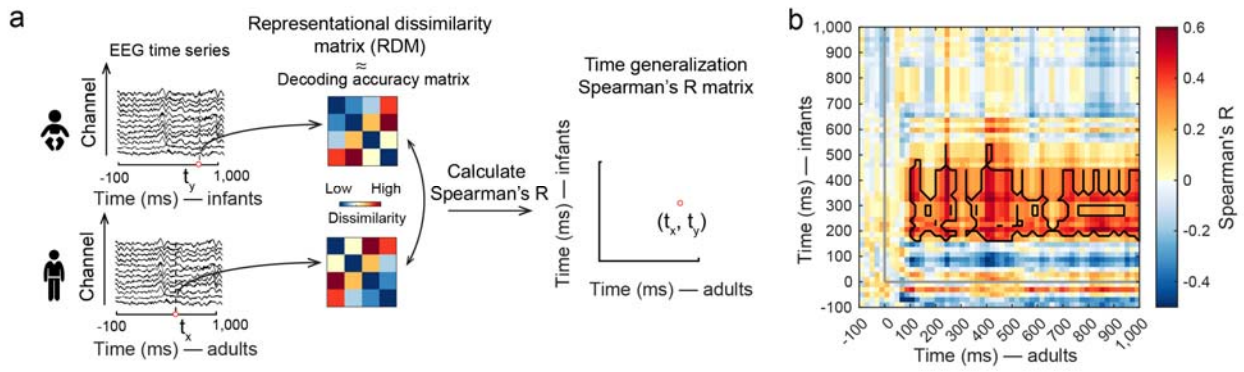

Figure 2 Category representations shared between infants and adults. (a) We used RSA to relate category representations in infants and adults. We interpret decoding accuracy as a dissimilarity measure on the assumption that the more dissimilar two representations are, the better the classifier performs. This allowed us to use time-resolved decoding accuracy matrices as representational dissimilarity matrices (RDMs) that summarize representational similarities between category representations. We compared RDMs (Spearman's *R*) in infants (average across participants) and adults (for each participant separately) for all time point combinations (t_x, t_y), assigning the values to a time-generalization matrix indexed in rows and columns by the time in adults (t_x) and infants (t_y). **(b)** Average time-generalization matrix relating category representations in infants and adults over time. For statistical details, see **Supplementary Table 1c**. The gray lines indicate image onset. Black outlines indicate time point combinations with significant correlation (n = 20, right-tailed sign permutation tests, cluster-defining threshold $P < .005$, corrected significance level $P < .05$).

We observed similarity in visual category representations between infants and adults at the time point combinations of 160–540ms in infants, and 100–1,000ms in adults (**Fig. 2b**, peak latency in infants: 200ms (200–360ms); in adults: 120ms (120–1,000ms)). This result was similarly achieved for alternative processing and data aggregation choices (**Supplementary Fig. 4**). Our findings establish quantitatively and directly that infants and adults share visual category representations.

The format of visual category representations

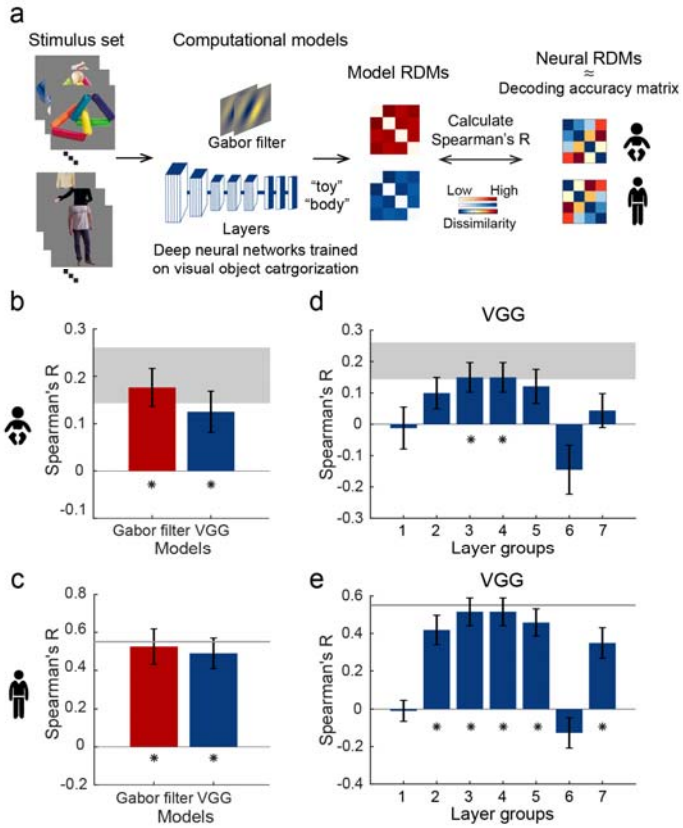

Figure 3 The format of category representations in infants and results. (a) We characterized what type of visual features are encoded in category representations in infants and adults by relating their category representations to computational models using RSA. For this we ran the stimulus images through a Gabor filter bank and the VGG-19 deep neural network trained on object categorization. We constructed RDMs from their unit activation patterns (visualized in **Supplementary Fig. 5**). We then compared model RDMs to infant and adult neural RDMs (Spearman's *R*), constructed as the average of RDMs over time from time-resolved category classification in the 95% CIs around peak latency (see **Fig. 1c,d**). **(b, c)** Results for infants **(b)** and adults **(c)** at the whole-model level. **(d, e)** Results for infants **(d)** and adults **(e)** at the DNN layer level. Error bars represent standard errors of the mean. Asterisks indicate significant correlations between model RDMs and neural RDMs (infant $n = 40$, adult $n = 20$, two-tailed sign-permutation tests, $P < .05$, FDR-corrected). Statistical details (i.e., correlations and *P*-values) are in **Supplementary Table 3**.

The time-resolved multivariate pattern analysis revealed the presence and dynamics of visual category representations in the infant and adult brain, but by itself is unable to specify their format, i.e., what type of visual features they encode. We hypothesized that adults would encode visual features represented at all levels of the visual processing hierarchy from low- to high complexity[10,2]. Instead, infants would encode visual feature rather of low- and mid-complexity, as predicted from visual behavior[4,16] and anatomical development patterns[21] of the infant visual brain.

To determine the format of category representations we related them to computational models of vision (**Fig. 3a**). We probed two types of models: a Gabor wavelet pyramid model as a model of simple visual features[22], and the deep neural network VGG-19 model[23] trained on object

categorization, which exhibits a hierarchy of low-to-high complexity features along its layers, and predicts activity along the visual processing hierarchy of the adult human brain well[10,24].

Assessing first the Gabor model and an aggregated summary of the VGG model across layers, we found similar representations between both models and infant and adult visual representations (**Fig. 3b,c**). This indicates that low complexity, simple features contribute to category representations. We then conducted a finer investigation of VGG at the level of layers. Considering each layer separately, we found that in infants middle layers predicted brain activity best, with layer group 3 and 4 being significant (**Fig. 3d**). In contrast, in adults, layers at all stages were significantly predictive (**Fig. 3e**). This pattern of results was also achieved for other types of deep neural network architectures (**Supplementary Fig. 6**) and independent of data selection choices (**Supplementary Fig. 7**), demonstrating the robustness of the result.

Together, this shows that in infants category representations are in the format of low- and intermediate-complexity features, whereas in adults, category is additionally discriminated by representations encoding features of high complexity.

Spectral properties of visual category representations

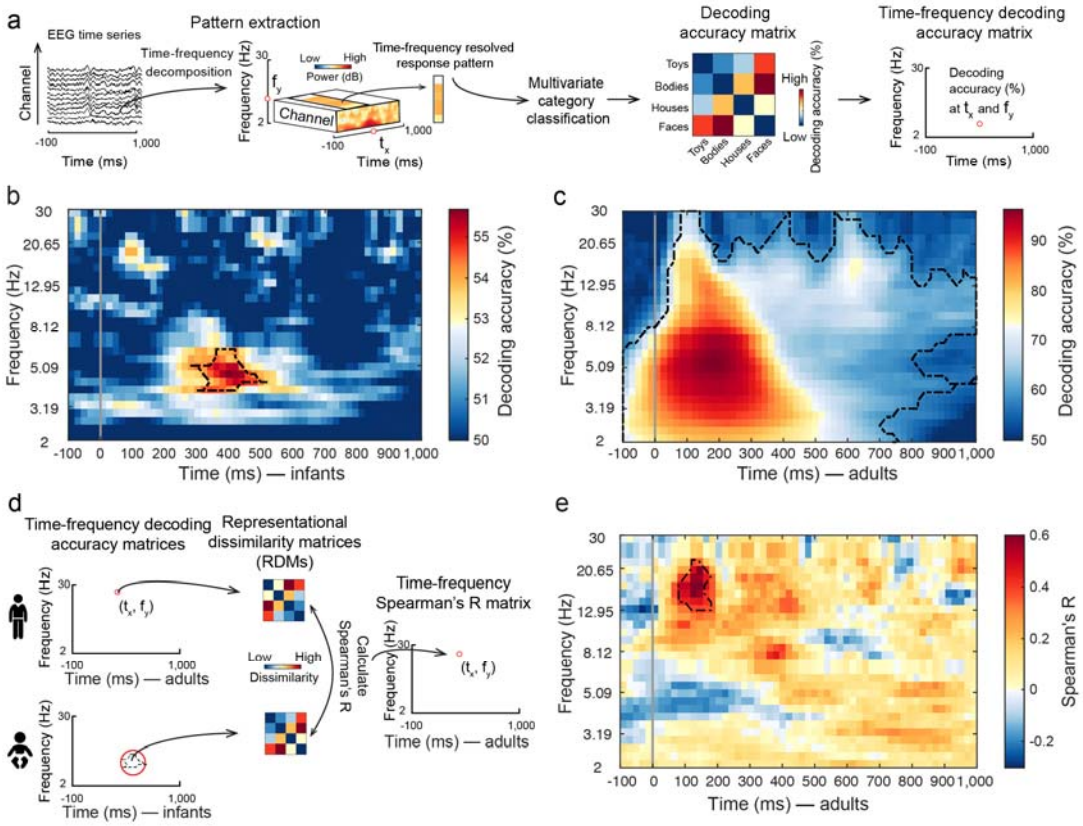

Figure 4 Spectral characterization of infant and adult category representations. (a) Category classification based on frequency-resolved EEG data. We first decomposed EEG data in time and frequency using *Morlet* wavelets for each trial and each channel, yielding a trial-wise representation of induced oscillatory power. We then conducted time-resolved multivariate classification of category separately on each frequency bin. This yielded a 4 × 4 matrix of decoding accuracies at each time point and frequency bin, which we either averaged to obtain grand average category classification **(b, c)** or used as an RDM in RSA **(d, e)**. **(b, c)** Results of time- and frequency-resolved MVPA for infants **(b)** and adults **(c)**. **(d)** RSA procedure linking oscillation-based visual category

representations in infants and adults. We first created a single aggregate infant oscillatory RDM by averaging decoding accuracy matrices based on the extent of the cluster in the infant data. We compared (Spearman's *R*) this aggregate infant RDM to time- and frequency- resolved RDMs for each participant in the adult sample. This yielded a two-dimensional matrix indicating in which frequency range and when category representations are similar between infants and adults. **(e)** Similarity between infant theta-based category representations and adult category representations resolved in time and frequency. The gray vertical lines indicate onset of image presentation. Black outlines indicate time point combinations with significant results (infants $n = 40$, adults $n = 20$, right-tailed permutation test, cluster-defining threshold $P < .005$, corrected significance level $P < .05$). For statistical details see **Supplementary Table 3.**

Neural oscillations underlie the formation and communication of visual representations[14,25–27]. Here we determined the neural oscillatory signature of visual category representations in infants. For this we resolved EEG data in distinct frequency bins from 2 to 30Hz and performed time-resolved visual category classification on each bin separately (**Fig. 4a**). In infants (**Fig. 4b**) we observed significant category classification accuracy in a specific cluster in the theta band with a peak at 4.63Hz (2.91–6.73Hz) and at 400ms (160–580ms). This reveals theta oscillations as the oscillatory basis of visual category representations in infants. In contrast, in adults (**Fig. 4c**) the cluster extended across the whole frequency range and time course investigated. This shows that the oscillatory basis of visual category representations in adults is broadband. Note that the observed differences in infants and adults are not a trivial consequence of differences in EEG power spectra, as those were similarly broadband in both infants and adults (**Supplementary Fig. 8**).

The observed pattern of results is consistent with two alternative hypotheses about the relationship between the oscillatory basis of visual category representations in infants and adults. One hypothesis is that there is a direct match in frequency, suggested by the fact that peak classification in infants and adults is at similar frequencies (i.e., at 4.63Hz and 5.59Hz respectively). Another hypothesis is that there is an upward shift across age, made plausible by the observations that brain rhythms increase in frequency during infant development[28].

To arbitrate between those hypotheses, we determined at which frequency band and at which time points category representations in adults were similar to infant category representations identified in the theta band (**Fig. 4d**. We found a cluster of significant correlations with a peak at 17.13Hz (9.78–20.65 Hz) at 120ms (120–360ms) (**Fig. 4e).** A supplementary data aggregation scheme yielded an equivalent pattern (**Supplementary Fig. 9**). This demonstrates in a direct and quantitative way an upward shift in the oscillatory basis of visual category representations from the theta range in infants to the alpha/beta range in adults.

Discussion

We investigated visual category representations in 6- to 8-month-old infants from EEG data using an integrated multivariate analysis framework. Our results revealed a developmental trajectory from infant to adult visual category representations in terms of their temporal dynamics, feature format and spectral properties. Time-resolved classification analysis[8,20] showed that visual category representations in infants emerged later and slower than in adults, and were incompletely developed in both feedforward and feedback processing mechanisms. Nevertheless, infants and adults already partly shared visual category representations. Computational model comparison[10] uncovered that the format of visual category representations in infants was low- to intermediate-complexity visual features, not including features of high

complexity as in adults. Content-sensitive spectral analysis showed that visual category representations in infants have an oscillatory basis in the theta range and indicated a shift of this basis to the alpha/beta range in adults.

Leveraging the high sensitivity of multivariate analysis on EEG data[8,20] allowed us to dissect the temporal dynamics of general category representations in infants. Our approach goes beyond previous work in developmental visual neuroscience in three ways. First, rather than relying on indirect inference from attentional markers indicating successful categorization[16], our approach assessed representations directly as they emerge with millisecond resolution. Second, our approach is not limited to the face category and face specific EEG components[17,18], but allows the study of potentially any visual category. Third, our approach enabled a new kind of quantitative comparison[19] of infant and adult visual category representations.

The emerging picture is one of not yet fully developed dynamics of adult-like visual category representations in infants. Representations in infants emerged later, slower, and lacked particular components of feedforward and recurrent processing, possibly related to immature myelination[29] and synaptic connectivity[30]. Nevertheless, representations in infants and adults shared large-scale temporal dynamics that encoded visual category information similarly.

Our results make direct predictions for the detailed developmental trajectory of visual category representations[21]. We expect category representations to emerge increasingly earlier and with faster temporal dynamics with increasing age, with additional feedforward and feedback components appearing at critical stages until a mature adult-like system emerges. Our approach makes these predictions immediately testable in future studies.

More broadly, our approach demonstrates a novel access point to largely unmapped neural representations in the infant brain, with strong potential to inform theories of cognitive development for cognitive capacities that emerge in the first year of life, such as object learning[3,6,7], speech processing[31] and core knowledge systems[32].

Our investigation of representational format using deep learning models of vision[10] revealed that infants' visual category representations encoded features of low to intermediate complexity. This is consistent with the fact that such features enable computational category classification[33], predict brain activity of adult visual cortex[10,24] and govern its large-scale topographical organization[34]. Visual category representations in adults instead included also features of high complexity, suggesting that features of increasing complexity emerge through development.

Previous research investigating the format of infant visual category representations tested hypotheses one by one through experimental manipulation, for example determining whether infants are sensitive to stimulus inversion[18] or tolerant to changes in viewing conditions[35]. Our approach instead allows comparison of any number of hypotheses as captured in explicit, image-computable computational models to increasingly accurately capture infant visual category representations in the future. To speed this process, we make the data publicly available.

Our findings about representational format of infants' visual representations suggest interesting constraints for artificial intelligence research. The engineering of deep learning models has been inspired by the biological brain, but the models' learning has remained biologically

unrealistic[10,24,36] and is perceived as a major impediment to building better models. Our results provide first constraints on the developmental trajectory of representations, which models should take for increased biological realism and better performance.

Investigating the oscillatory basis of visual category representations, we observed a shift from the theta range in infants to the alpha/beta range in adults. Evidence for this shift was particularly strong as the link between frequency bands relied on similarity in the detailed contents represented, overcoming ambiguities of previous approaches[37].

One interpretation of these findings is that in infants, category representations are processed in neural networks for learning and memory associated with the theta rhythm[38,39], whereas in adults, equivalent category representations are processed quickly in fully developed semantic networks associated with the alpha/beta rhythms[40,41]. Alternatively, the frequency shift might be due to more efficient axonal transmission as a result of improved myelination[29] that may enable higher neural oscillations to emerge in the same neural circuits, consistent with increases in the prevalent frequency in the EEG across development[28]. On this account our finding suggests a novel general developmental trajectory for neural communication channels in the human brain: specific information-processing mechanisms working at low frequency in infants are shifted up in frequency gradually across development, and the size of this shift depends on the differences in neural circuit myelination.

In sum, our results reveal the nature and developmental trajectory of infant to adult visual category representations, from infancy to adulthood, that underlies the development of fast and efficient visual categorizations skills.

Methods

Participants

Two independent pools of participants took part in this study: 6–8 months old infants and young adults. The infant sample was assessed at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. It comprised 48 participants, of which 8 were excluded due to insufficient data, yielding a final sample analyzed of 40 infant participants (gender: 19 female, age: mean ± SD: 214.9 ± 14.76 days). The adult sample was assessed at the Freie Universität Berlin, Germany. It comprised 20 participants of which none was excluded (gender: 11 female, age: mean ± SD: 26.1 ± 3.81 years). Caregivers of all infants and all adult participants gave written informed consent. The study was conducted according to the Declaration of Helsinki and the infant and adult protocols were approved by the respective local ethic committees.

Stimuli and experimental paradigm

The stimulus set consisted of 32 object images in each of the four categories included: houses, toys, faces, and bodies. This yielded a total set of 4 × 32 = 128 object images. All object images were cut-out from color photographs (**Supplementary Fig. 1**). We analyzed the data at the level of category.

Experimental procedure in the infant sample

In the infant experiment, participants were presented with 272 trials divided into four blocks. Each block had the same basic structure. At the beginning of each block four stimuli (one stimulus per object category) were separately presented three times in randomized order. Thereafter participants were presented with a random sequence of images comprising the same four stimuli seven more times, intermixed with 28 other images (seven images per category) presented only once. This experimental design was chosen because it allows assessing the effect of object image repetition of infant brain responses, but this question is orthogonal to the ones pursued here and will be reported separately.

Each trial consisted of a fixation dot presented for a variable duration of 700–900ms, followed by a stimulus presented for 2,000ms at the center of the screen (**Supplementary Fig. 2a**). To capture the attention of the infants and direct their gaze to the screen we implemented two measures. First, we presented a yellow duck image and duck sound for 1,000ms at the beginning of each block and thereafter every 10 trials. Second, each stimulus was presented together with one of ten arbitrary sounds that were assigned randomly at each trial.

During the assessment, infants sat on their care giver's lap at a viewing distance of about 80 cm from a 17-in. CRT screen. The object images were presented at the center of the screen, subtending a visual angle of approximately 5.0°. To monitor infants' gaze, we recorded videos of infants' faces throughout the experiment.

Experimental procedure in the adult sample

We adapted the experimental design for the adult sample. In short, all images were shown equally often, with higher number of repetitions, at shorter presentation times and higher presentation rates that in the infant study (**Supplementary Fig. 2b**).

The first 3 participants were presented with 1,280 trials divided into 5 runs. In each run each object image was presented twice. The other 17 participants were presented with 3,840 trials divided into 10 runs. In each run each object image was presented three times. In each run,

images were presented in random order, and runs were separated by breaks that were self-paced by the participants.

Each trial consisted of the presentation of a fixation cross with a variable duration of 600–800ms, followed by a stimulus presentation for 500ms. Stimuli were presented at the center of the screen at a visual angle of approximately 7.0°.

Participants were instructed to keep fixation on the center of the screen throughout the experiment. To ensure that participants attended to the stimuli and to avoid contamination of the relevant recording times with blink artefacts, participants were instructed to press a button and blink their eyes in response to a paper clip image that was shown randomly every 4 to 6 trials (average 5 trials). Paper clip trials were excluded from all further analysis.

EEG acquisition and preprocessing

Infant sample

EEG data for the infant sample were recorded in a shielded room using 30 Ag/AgCl ring electrodes and a TMSi 32-channel REFA amplifier at a sampling rate of 500 Hz. Electrodes were placed according to the standard 10-20 system. Electrodes V+Fp2 and V– recorded the vertical electrooculogram (VEOG), and electrodes H-F9 and HF+10 recorded the horizontal electrooculogram (HEOG), Cz served as the online reference. We conducted preprocessing using the Fieldtrip toolbox[42]. The continuous EEG data was segmented for each trial into epochs. For subsequent time-resolved multivariate analysis we extracted the epoch from –200ms to +1,000ms with respect to image onset. For analysis that was additionally resolved in frequency we used longer epochs to allow better estimation at lower frequencies from –500ms to +1,000ms.

We removed all trials during which participants did not gaze at the screen for 1,000ms after stimulus onset as assessed by visual inspection of the video recordings. In this reduced trial set (mean ± SD: 139.6 ± 47.76 trials) we removed noisy channels (mean ± SD: 1.25 ± 1.32) and replaced them by interpolated data from adjacent electrodes. We further conducted independent component analysis (ICA) and removed components related to eye-movement and muscle artifacts as identified by visual inspection.

Adult sample

EEG data for the adult sample were recorded using an EASYCAP 64-channel system and a Brainvision actiCHamp amplifier at a sampling rate of 1,000Hz. Data were filtered online between 0.3 and 100 Hz. Electrodes were placed according to the standard 10-10 system. Electrode Fz served as the online reference. We conducted preprocessing using the Brainstorm 3 toolbox[43]. Up to two noisy channels were removed for each participant as identified by visual inspection. We conducted ICA to identify and remove eye-movement and muscle artifact components by visual inspection of independent components. The continuous EEG data were then segmented for each trial into epochs from –200ms to +1,000ms (for time-resolved analysis) and from –500ms to +1,000ms (for time- and frequency resolved analysis).

Time-frequency decomposition

We decomposed the EEG time series into frequency-specific components by convolving the data with complex Morlet wavelets separately for each trial and sensor. The wavelets had a constant length of 2,600ms and were logarithmically spaced in 30 frequency bins between 2Hz and 30Hz. We obtained the absolute power values for each time point and frequency bin by taking the

square root of the resulting time-frequency coefficients. We normalized these power values to reflect relative changes (expressed in dB) with respect to the pre-stimulus baseline (–300ms to –100ms with respect to stimulus onset). We downsampled the time-frequency representations to a temporal resolution of 50 Hz (by averaging data in 20ms-bins) to increase the signal-to-noise ratio of subsequent analyses. This yielded for each trial a power value for each time point and frequency bin.

Multivariate classification of visual category from EEG data

To characterize the temporal dynamics with which visual category representations emerge in infant and adult brains we conducted multivariate EEG classification using linear support vector machines (SVMs). We analyzed the infant and the adult data set separately and equivalently.

We conducted two common variants of multivariate EEG classification: time-resolved EEG analysis[8,44] and time-generalization analysis[20]. We conducted the analysis separately on the adult and infant sample, and separately for each participant. All analyses employed binary c-support vector classification (C-SVC) with a linear kernel as implemented in the LIBSVM toolbox[45]. The details of the time-resolved and the time-generalization analysis are as follows.

Time-resolved classification

We used time-resolved multivariate pattern analysis on EEG data (**Fig. 1b**) to determine the time course with which visual category representations emerge in infant and adult brains. For each time point of the EEG epoch (from –200ms to +1,000ms), we extracted trial-specific EEG channel activations (i.e., 25 in infants and 63 in adults) and arranged them into pattern vectors for each of the four category conditions (i.e., face, house, body, and toy) of the stimulus set. To increase the signal-to-noise ratio (SNR), we randomly assigned raw trials into four bins of approximately equal size each and averaged them into four pseudo-trials. We used a leave-one-pseudo-trial-out cross validated classification approach. We trained the SVM classifier to pairwise decode any two conditions using three of the four pseudo-trials for training. We used the fourth left-out pseudo-trial for testing, yielding classification accuracy (chance level 50%) as a result. The procedure was repeated 100 times, each time with a new random assignment of trials to pseudo-trials. The resulting decoding accuracy was averaged across repetitions and assigned to a decoding accuracy matrix of size 4 × 4, with rows and columns indexed by the conditions classified. The matrix is symmetric across the diagonal, with the diagonal undefined. This procedure yielded one decoding matrix for every time point.

In addition to classifying visual category from broadband responses (i.e., single trial raw unfiltered waveforms), we classified object categories from oscillatory responses. This analysis followed the same rationale as the classification analysis described above, with the only difference that classification was conducted on power value patterns instead of raw activation value patterns. The analysis was conducted separately for each frequency bin separately. This resulted in a decoding accuracy matrix of size 4 × 4 as defined above for every time point and every frequency bin.

Time generalization analysis

We used time-generalization classification analysis[20] to determine how visual representations emerging at different time points during the dynamics of visual perception relate to each other. For time and memory efficiency, we downsampled the EEG data to a sampling rate of 50 Hz by averaging the raw EEG data in 20ms bins. The procedure was equivalent to the time-resolved

classification analysis with the only difference that classifiers trained on data from a particular time point were not only tested on left out data from the same time point, but iteratively on data from the same and all other time points. The idea is that successful classifier generalization across time points indicates similarity of visual representations over time. This analysis yielded thus a size 4 × 4 decoding accuracy matrix indexed in rows and columns by the conditions compared for all time point combinations from –200 to +1,000ms. We averaged the entries of the decoding accuracy matrix at each time point, yielding a temporal generalization matrix indexed in rows and columns by training and testing time.

Sensor-space searchlight analysis

We performed a sensor-space searchlight analysis[46,47] to localize in EEG channel space which channels contributed to the classification of category. For each EEG channel we defined a neighborhood as a sphere of the 10 (for adults) or 5 (for infants) closest EEG channels. For each EEG channel we then performed time-resolved category classification analysis, limiting data entering the analysis to its neighboring channels. Averaging across all pairwise category classifications yielded one decoding accuracy for each time point and for each EEG channel. We further averaged the results in 200ms bins, yielding a single EEG channel searchlight map of grand average decoding accuracy for each time bin.

Comparing visual representations in infants and adults

We determine whether infants and adults have similar visual category representations using representational similarity analysis (RSA)[48,49]. The idea is that infants and adults share representations of category if they treat the same categories as similar or dissimilar. We determined this in a two-step process.

In a first step, for each age group independently condition-specific multivariate activity patterns (adults: 63 electrodes; infants: 25 electrodes) were compared for dissimilarity. Dissimilarity was determined for all pairwise combinations of conditions, and dissimilarity values were aggregated in so-called representational dissimilarity matrices (RDMs) indexed in rows and columns by the conditions compared (here: 4 × 4 RDMs indexed by the 4 object categories). RDMs thus provide a statistical summary of the similarity and thus representational relations between visual category representations. The RDMs gained from the infant and adult sensor space separately have the same definition and dimensionality and are thus directly comparable. Thus, in a second step, the infant RDM and the adult RDMs are related to each other by determining their similarity.

We applied RSA to two different types of data: evoked responses (i.e., recorded voltage signals) and oscillatory responses (i.e., spectral power). In both cases we re-used the results of the classification analysis described above for the definition of RDMs. Classification accuracy can be interpreted as a dissimilarity measure on the assumption that the more dissimilar activation patterns are for two conditions, the easier they are to classify[8,50]. We detail the different RSA procedures below. To reduce visual complexity of the analysis we subsampled the results of the classification analysis by binning them in 10ms bins.

Relating visual category representations in infants and adults based on raw broadband time courses

We investigated whether infants and adults share common visual representations based on broadband responses. As visual representations in adults and infants likely emerge with different time courses, we related their visual representations in a representational similarity time-

generalization analysis. As RDMs we used time-point specific decoding accuracy matrices **(Fig. 2a**). We first averaged infant RDMs across all participants to increase SNR, resulting in one average infant RDM per time point. We then correlated (Spearman's *R*) the average infant RDM to each adult (n = 20) RDM across all time point combinations. This yielded 20 correlation matrices, indexed in rows and columns by the time points compared (rows: infant time; columns: adult time), indicating when infants and adults share category representations.

Relating visual category representations in infants and adults based on frequency-specific power time courses

We investigated whether infants and adults share visual representations in particular frequency bands. As RDMs we used decoding accuracy matrices from the classification analysis based on time-frequency resolved power values. As in this analysis we could neither assume similar time courses, nor similar roles for particular frequencies across infants and adults, we related infant and adult representations in a time-and-frequency-generalization analysis. To do this, we first defined a single aggregate infant RDM by averaging decoding accuracy matrices based on the extent (time and frequency) of the significant cluster in the infant data alone (see **Fig. 4b**). To increase signal-to-noise we only included RDMs whose average across entries in single participants was greater than or equal to 50% decoding accuracy. Note that this criterion is orthogonal to the hypotheses tested and thus does not bias the analysis. We compared (Spearman's *R*) this single aggregate infant RDM to time- and frequency- resolved RDMs for each participant of the adult sample (n = 20), separately for each frequency and time point. This yielded 20 correlation matrices, with rows representing time points and columns representing frequency bins, indicating when and at which frequency infants and adults share category representations.

Relating visual representations in infants and adults to computational models

To characterize the format of visual category we related neural representations in infants and results to different computational models using RSA. We constructed model RDMs from computational models (**Fig. 3a**) that represent visual information in different formats. We considered two types of visual computational models: a Gabor wavelet pyramid as a model of low-level feature representations[22,51], and the VGG-19[23] deep convolutional neural network (DNN) trained to categorize object images. Deep neural networks process visual information along a hierarchy of increasing complexity from low to high[52,53] that has been shown to match the processing hierarchy of the human brain[54–56] and predict human and non-human primate brain activity better than other model class[10,57–59].

To construct model RDMs we first ran all visual stimuli in the study (i.e., 128 object images) through the models and extracted their activation values. More specifically, for the Gabor model we extracted a single set of model responses for Gabor wavelets differing in size, position, orientation, spatial frequency and phase. For the DNN we used the MatConvNet toolbox[60] to extract model neuron activation values from the rectified linear units (Relu) for each layer. We z-transformed activation values across stimuli for each stage/layer separately and averaged the transformed values across the 32 stimuli belonging to each of the four categories (i.e., face, body, house and toy), resulting in four category-specific activation values. We formed the patterns into vectors and computed the dissimilarity (1 – Pearson's *R*) between all pairwise combinations of the four category activation vectors, resulting in a 4 × 4 RDM for each DNN layer of each DNN separately, and one model RDM for the Gabor feature model.

To construct neural RDMs that capture category representations well we averaged decoding accuracy matrices from time-resolved category classification (**Fig. 1c,d**) in the 95% confidence intervals around peak latency in time-resolved category classification. Our rationale was that peak latency is the time point when categories were linearly best separable and thus their representations most explicit[61]. To increase signal-to-noise we only included RDMs whose average across entries in single participants was greater than or equal to 50% decoding accuracy. Note that this criterion is orthogonal to the hypotheses tested and thus unbiased. This yielded a single neural RDM for every infant and adult participant. We then related infant and adult neural RDMs to model RDMs using Spearman's *R*, yielding a single correlation value for each model RDM and participant.

To allow assessing the models' predictivity with respect to the noise in the data we calculated an upper and lower bound for the noise ceiling[62], that is the predictions a perfect model may reach given the noise in the data. This procedure was conducted separately for the infant and the adult sample. To estimate the upper bound we correlated (Spearman's *R*) each participant's neural RDM with the mean neural RDM across all participants. To estimate the lower bound we correlated (Spearman's *R*) each participant's neural RDM with the mean neural RDM excluding that participant iteratively for all participants. We averaged the results, yielding estimates of the lower and upper noise ceiling for infants and adults.

Statistical inference

We used non-parametric statistical inference for random-effects inference to avoid assumptions about the distribution of the data[63,64]. We used permutation tests for cluster-size inference, and bootstrap tests for confidence intervals on maxima, cluster onset/offset, and peak-to-peak latency differences. The sample size (*n*) for infants was 40 and for adults 20. Tests were either two- or right-tailed and are indicated for each result separately.

<u>Permutation tests</u>

We tested the statistic of interest (i.e., mean decoding accuracy or correlation coefficient in RSA across participants) using sign permutation tests. The null hypothesis was that the statistic of interest was equal to chance (i.e., 50% decoding accuracy, a Spearman's *R* of 0). Under the null hypothesis, we could permute the category labels of the EEG data, which effectively corresponds to a sign permutation test that randomly multiplies participant-specific data with +1 or –1. For each permutation sample, we recomputed the statistic of interest. Repeating this permutation procedure 10,000 times, we obtained an empirical distribution of the data. We converted the original statistic (i.e., correlation coefficient, the decoding time courses, time-time matrices of correlation coefficients or decoding accuracies, and time-frequency decoding matrices) into *P*-values (correlation coefficients), 1-dimensional (time courses), or 2-dimensional (time-generalization or time-frequency) *P*-value matrices.

We controlled the family-wise error rate using cluster-size inference. We first thresholded *P*-value time courses or maps at $P < .005$ (cluster-definition threshold) to define supra-threshold clusters by contiguity. These supra-threshold clusters were reported significant only if the size exceeded a threshold, estimated as follows: the previously computed permutation samples were also converted to *P*-value time courses/matrices and also thresholded to define resampled versions of supra-threshold clusters. These clusters were used to construct an empirical distribution of maximum cluster size and estimate a threshold of 5% of the right tail of this distribution (i.e., the corrected *P*-values is $P < .05$).

Bootstrap tests

We calculated 95% confidence intervals for the onsets of the first significant cluster and the peak latency of the observed effects. To achieve this, we created 1,000 bootstrapped samples by sampling the participants with replacement. For each bootstrap sample, we determined the peak latency as well as onsets of the first significant cluster and the offset of the last significant cluster. This resulted in empirical distributions of peak, onset and offset latencies on which we determined 95% confidence intervals.

To calculate confidence intervals on mean peak-to-peak latency differences, we created 1,000 bootstrapped samples by sampling the participant-specific latencies with replacement. This yielded an empirical distribution of mean peak-to-peak latencies. If the 95% confidence interval did not include 0, we rejected the null hypothesis of no peak-to-peak latency differences. The threshold $P < .05$ was corrected for multiple comparisons whenever appropriate using FDR correction.

References

1. Potter, M. Meaning in visual search. *Science* **187**, 965–966 (1975).
2. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**, 520–522 (1996).
3. Pascalis, O., Haan, M. de & Nelson, C. A. Is Face Processing Species-Specific During the First Year of Life? *Science* **296**, 1321–1323 (2002).
4. Aslin, R. N. What's in a look? *Dev Sci* **10**, 48–53 (2007).
5. Rakison, D. H. & Poulin-Dubois, D. Developmental origin of the animate–inanimate distinction. *Psychological Bulletin* **127**, 209–228 (2001).
6. Mareschal, D. & Quinn, P. C. Categorization in infancy. *Trends in Cognitive Sciences* **5**, 443–450 (2001).
7. Stahl, A. E. & Feigenson, L. Observing the unexpected enhances infants' learning and exploration. *Science* **348**, 91–94 (2015).
8. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat Neurosci* **17**, 455–462 (2014).
9. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–30 (2001).
10. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat Neurosci* **19**, 356–365 (2016).
11. Grill-Spector, K. & Malach, R. The Human Visual Cortex. *Annual Review of Neuroscience* **27**, 649–677 (2004).
12. Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. Neural correlates of category-specific knowledge. *Nature* **379**, 649–52 (1996).
13. Singer, W. Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annual Review of Physiology* **55**, 349–374 (1993).
14. Fries, P. Rhythms for Cognition: Communication through Coherence. *Neuron* **88**, 220–235 (2015).

15. Aslin, R. N. & Fiser, J. Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences* **9**, 92–98 (2005).
16. Hoehl, S. The development of category specificity in infancy – What can we learn from electrophysiology? *Neuropsychologia* **83**, 114–122 (2016).
17. de Haan, M. & Nelson, C. A. Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology* **35**, 1113–1121 (1999).
18. 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).
19. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Sys. Neurosci.* **2**, 4 (2008).
20. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences* **18**, 203–210 (2014).
21. Johnson, M. H. Functional brain development in humans. *Nature Reviews Neuroscience* **2**, 475–483 (2001).
22. Jones, J. P. & Palmer, L. A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* **58**, 1233–1258 (1987).
23. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]* (2015).
24. Kietzmann, T. C., McClure, P. & Kriegeskorte, N. Deep Neural Networks in Computational Neuroscience. *Oxford Research Encyclopedia of Neuroscience* (2019) doi:10.1093/acrefore/9780190264086.013.46.
25. Berger, H. Über das Elektrenkephalogramm des Menschen. *Archiv f. Psychiatrie* **87**, 527–570 (1929).
26. Buzsáki, G. & Draguhn, A. Neuronal Oscillations in Cortical Networks. *Science* **304**, 1926–1929 (2004).
27. Uhlhaas, P. J., Roux, F., Rodriguez, E., Rotarska-Jagiela, A. & Singer, W. Neural synchrony and the development of cortical networks. *Trends in Cognitive Sciences* **14**, 72–80 (2010).

28. Marshall, P. J., Bar-Haim, Y. & Fox, N. A. Development of the EEG from 5 months to 4 years of age. *Clinical Neurophysiology* **113**, 1199–1208 (2002).

29. Deoni, S. C. L. *et al.* Mapping Infant Brain Myelination with Magnetic Resonance Imaging. *J. Neurosci.* **31**, 784–791 (2011).

30. Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology* **387**, 167–178 (1997).

31. Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606–608 (1992).

32. Spelke, E. S. & Kinzler, K. D. Core knowledge. *Developmental Science* **10**, 89–96 (2007).

33. Oliva, A. & Torralba, A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision* **42**, 145–175 (2001).

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

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

36. Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. Neuroscience-Inspired Artificial Intelligence. *Neuron* **95**, 245–258 (2017).

37. Saby, J. N. & Marshall, P. J. The Utility of EEG Band Power Analysis in the Study of Infancy and Early Childhood. *Developmental Neuropsychology* **37**, 253–273 (2012).

38. Ward, L. M. Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences* **7**, 553–559 (2003).

39. Buzsáki, G. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus* **15**, 827–840 (2005).

40. Jensen, O. & Mazaheri, A. Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Front. Hum. Neurosci.* **4**, (2010).

41. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences* **16**, 606–617 (2012).

42. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* vol. 2011 e156869 https://www.hindawi.com/journals/cin/2011/156869/ (2010).

43. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience* **2011**, 1–13 (2011).

44. Carlson, T., Tovar, D. A., Alink, A. & Kriegeskorte, N. Representational dynamics of object vision: The first 1000 ms. *Journal of Vision* **13**, 1–19 (2013).

45. Chang, C. & Lin, C. *{LIBSVM}: a library for support vector machines*. (2001).

46. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3863–3868 (2006).

47. Kaiser, D., Oosterhof, N. N. & Peelen, M. V. The Neural Dynamics of Attentional Selection in Natural Scenes. *J Neurosci* **36**, 10522–10528 (2016).

48. Kriegeskorte, N. *et al.* Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron* **60**, 1126–1141 (2008).

49. Kriegeskorte, N. & Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences* **17**, 401–412 (2013).

50. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage* **173**, 434–447 (2018).

51. Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural images from human brain activity. *Nature* **452**, 352–5 (2008).

52. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv:1312.6034 [cs]* (2013).

53. Bau, D., Zhou, B., Khosla, A., Oliva, A. & Torralba, A. Network Dissection: Quantifying Interpretability of Deep Visual Representations. *arXiv:1704.05796 [cs]* (2017).

54. Güçlü, U. & Gerven, M. A. J. van. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J. Neurosci.* **35**, 10005–10014 (2015).

55. Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A. & Oliva, A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports* **6**, 27755 (2016).

56. Eickenberg, M., Gramfort, A., Varoquaux, G. & Thirion, B. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage* **152**, 184–194 (2017).

57. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Comput Biol* **10**, e1003915 (2014).

58. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* **111**, 8619–8624 (2014).

59. Schrimpf, M. *et al.* Integrative Benchmarking to Advance Neurally Mechanistic Models of Human Intelligence. *Neuron* **108**, 413–423 (2020).

60. Vedaldi, A. & Lenc, K. MatConvNet - Convolutional Neural Networks for MATLAB. *arXiv:1412.4564 [cs]* (2016).

61. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How Does the Brain Solve Visual Object Recognition? *Neuron* **73**, 415–434 (2012).

62. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol* **10**, e1003553 (2014).

63. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping* **15**, 1–25 (2002).

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

Acknowledgements

R.M.C is supported by Deutsche Forschungsgemeinschaft (DFG) grants (CI241/1-1, CI241/3-1) and by a European Research Council Starting Grant (ERC-2018-StG). S.X. is supported by a scholarship by the Chinese Scholarship Council. M.K. and S.H. were supported by the DFG and FWF jointly (grant numbers: KO 6028/1-1; I 4332-B), and S.H. also by the Max Planck Society. Computing resources were provided by the high-performance computing facilities at ZEDAT, Freie Universitaet Berlin.

Author contributions

Conceptualization: RC, SX, MK, SH, BT. Methodology: SX, RC, MK. Software: SX, MK. Formal analysis: SX, MM. Investigation: EK, CK, Resources: SX, RC, MK. Data curation: SX. Writing - original draft preparation SX, RC. Writing - review and editing: RC, SX, MK, SH, BT. Visualization: SX. Supervision: RC, MK. Project administrations: RC, SH, MK. Funding acquisition: RC, SH and MK.

Additional information

Correspondence and requests for materials should be addressed to R.M.C. rmcichy@zedat.fu-berlin.de or S.X. siying.xie@outlook.com.

Competing interest declaration

The authors have no competing interests.

Data availability

Full data will be available upon publication here: [*OSF link to be inserted upon acceptance*].

Example data for review is available here: https://osf.io/ruxfg/?view_only=919a62ecab944e99a80b9a467951d2e1.

Code availability

Custom code is available here: https://github.com/anonymturtle/VCR_infant.