Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex

Jesse Gomez [1,2]*, Michael Barnett [3] and Kalanit Grill-Spector [1,4,5]

The functional organization of human high-level visual cortex, such as the face- and place-selective regions, is strikingly consistent across individuals. An unanswered question in neuroscience concerns which dimensions of visual information constrain the development and topography of this shared brain organization. To answer this question, we used functional magnetic resonance imaging to scan a unique group of adults who, as children, had extensive visual experience with Pokémon. These animal-like, pixelated characters are dissimilar from other ecological categories, such as faces and places, along critical dimensions (foveal bias, rectilinearity, size, animacy). We show not only that adults who have Pokémon experience demonstrate distinct distributed cortical responses to Pokémon, but also that the experienced retinal eccentricity during childhood can predict the locus of Pokémon responses in adulthood. These data demonstrate that inherent functional representations in the visual cortex—retinal eccentricity—combined with consistent viewing behaviour of particular stimuli during childhood result in a shared functional topography in adulthood.

Humans possess the remarkable ability to rapidly recognize a wide array of visual stimuli. This ability is thought to occur as a result of cortical computations in the ventral visual stream[1]: a processing hierarchy that extends from the primary visual cortex to the ventral temporal cortex (VTC). Previous research has shown that VTC responses are key for visual recognition because (1) distributed VTC responses contain information about objects[2–4] and categories[5], and (2) responses in category-selective regions in the VTC, such as the face-, body-, word- and place-selective regions[6–12], are linked to the perception of these categories[13–15]. Distributed VTC response patterns to different visual categories are distinct from one another and are arranged with remarkable spatial consistency along the cortical sheet across individuals[16–23]. For example, peaks in distributed VTC response patterns to faces are consistently found on the lateral fusiform gyrus (FG). Although several theories have been suggested to explain the consistent spatial topography of the VTC[24–26], developmental studies suggest that experience may be key for the normal development of the VTC and recognition abilities. For example, behavioural studies suggest that the typical development of recognition abilities is reliant on viewing experience during childhood[27–32]. However, the nature of childhood experience that leads to consistent spatial functional topography of the VTC—whether it is the way stimuli such as faces or places are viewed, or the image-level statistics of the stimuli themselves—remains unknown.

Several theories have proposed attributes that may underlie the functional topography of human high-level visual cortex. These include: (1) the eccentricity bias of retinal images associated with the typical viewing of specific categories[19,24], for example, face discrimination is thought to require high visual acuity supported by foveal vision, but peripheral vision is believed to be more important for processing places, as in the real world they occupy the entire the visual field; (2) the average rectilinearity of stimuli from particular categories[25,33], for example, faces are curvilinear,

but man-made places tend to be rectilinear; (3) the perceived animacy of stimuli[5,34–37], for example, faces are perceived to be animate whereas places are not; and (4) the real-world size of stimuli[38], for example, faces are physically smaller than places and buildings.

Each of these theories proposes an underlying principle to describe the coarse functional topography of the VTC relative to its cortical macroanatomy. That is, inherent in all these theories is the idea that a physical or perceived dimension of a stimulus maps onto a physical dimension along the cortical surface. For example, in the human VTC, small, curvy, animate and foveal stimuli elicit stronger responses lateral to the mid-fusiform sulcus (MFS), whereas large, linear, inanimate and peripherally extending stimuli elicit stronger responses in the cortex medial to the MFS. However, which of these dimensions drives the development of the functional organization of the VTC is unknown.

Research on cortical plasticity in animals has made two key discoveries related to this question. First, eccentricity representations in the early and intermediate visual cortex are probably established in infancy[39,40], as they may be constrained by both wiring[41] and neural activity that starts in utero[42]. For example, research on ferret and mouse development suggests that retinal waves during gestation and before eye-opening are sufficient to establish eccentricity representations in the visual cortex[40]. An eccentricity proto-architecture is also detectable early in infant macaque development[39]. Second, visual development has a critical period during which the brain is particularly malleable and sensitive to visual experience[32,43–48]. For example, previous research suggests that new category representations in high-level visual cortex emerge with experience only in juvenile macaques, but not if the same experience happens to adult macaques[43]. Furthermore, visual deprivation of a category (for example, faces) in infancy results in a lack of development of a cortical representation for that category[32]. Together, these findings support the following predictions regarding human development: first,

[1]Neurosciences Program, Stanford University School of Medicine, Stanford, CA, USA. [2]Psychology Department, University of California Berkeley, Berkeley, CA, USA. [3]Psychology Department, University of Pennsylvania, Philadelphia, PA, USA. [4]Psychology Department, Stanford University, Stanford, CA, USA. [5]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA. *e-mail: jessegomez@berkeley.edu
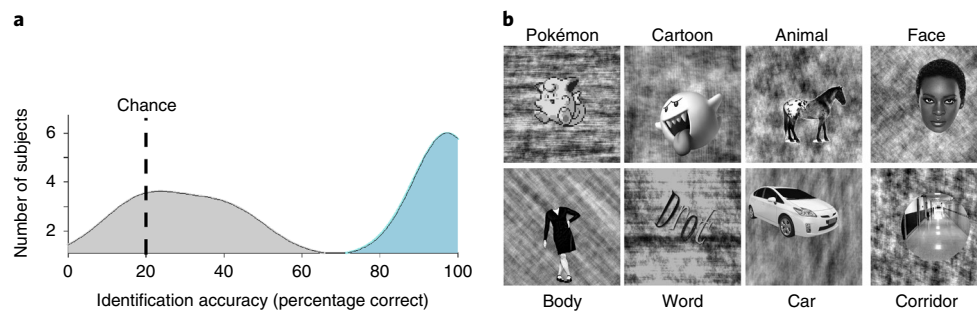
Fig. 1 | Localizer stimuli and behavioural naming performance. a, Distributions of participant accuracies from a five-alternative-choice Pokémon naming task outside the scanner. Experienced participants (blue; $n = 11$) significantly outperformed novices (grey; $n = 9$). **b**, Example stimuli from each of the categories used in the fMRI experiment. In each 4 s trial, participants viewed 8 different stimuli from each category at a rate of 2 Hz while performing an oddball task to detect a phase-scrambled stimulus with no intact object overlaid. Participants completed 6 runs, of 3 min 38 s each, using different stimuli. See https://www.pokemon.com/us/pokedex/ for more general examples of Pokémon and Supplementary Fig. 9 for more examples of the pixelated GameBoy Pokémon stimuli.

if eccentricity representations in high-level visual cortex are present early in development, then eccentricity stands to be a strong developmental constraint for the later emergence of object representations; second, testing theories of VTC development requires the measurement of the effects of childhood experience on the formation of new brain representations.

To investigate the developmental origins of the functional organization of the VTC, the effects of exhaustive childhood experience with a novel visual category (distinct from natural categories along the four aforementioned visual dimensions) must be tested. Performing such an experimental manipulation in a laboratory setting with children would require an unfeasible length of time. However, there are adults who, in their childhood, had prolonged, rewarded, and crucially similar visual experiences with a shared novel stimulus category. In 1996, Nintendo released a popular videogame, Pokémon, alongside a hand-held playing device, the GameBoy. Children as young as 5 yr old played this game, in which individuating the animal-like creatures of Pokémon was integral to optimal game performance. Furthermore, game playing conditions were nearly identical across individuals: children held the device at a similar (arm's length) viewing distance, repeatedly for hours a day, over a period of years. Importantly, Pokémon were rendered with large pixels in a small 2.5 cm × 2.5 cm region of the GameBoy screen, making Pokémon, compared to faces or places, a unique visual category that is of a constant size and viewed with foveal vision, but possesses strong linear features. Although many Pokémon characters resemble animals, and children saw Pokémon animations, they were never encountered in the real world. Thus, the animacy and real-world sizes of Pokémon are inferred attributes rather than physical ones. As Pokémon are different from both faces and places along these visual dimensions, we can use them to answer fundamental questions: first, does prolonged experience individuating Pokémon result in a novel information representation in the VTC? Second, is the location of this response predicted by a specific visual dimension?

To answer these questions, we used functional magnetic resonance imaging (fMRI) to scan 11 adult participants who began playing Pokémon between the ages of 5 and 8 yr old and 11 age-matched adult novices who had no childhood experience of playing Pokémon. During fMRI, participants viewed stimuli from eight categories: faces, animals, cartoons, bodies, words, cars, corridors and Pokémon (Fig. 1b). We performed multi-voxel pattern analysis (MVPA) of VTC responses to these stimuli in each participant, combined with decoding analyses, to assess (1) whether prolonged experience results in an informative representation for Pokémon in experienced versus novice participants, and (2) whether distributed

VTC representations form a consistent spatial topography across the VTC in experienced versus novice participants. If experienced individuals demonstrate an informative and spatially consistent representation for Pokémon compared to other categories, it would allow us to ask whether the topography of this representation across the VTC is predicted by one of the four visual dimensions (eccentricity bias, rectilinearity, animacy and perceived size), which we quantified for Pokémon, faces and places.

Results

Childhood experience with Pokémon results in distinct and reproducible information across the VTC. Experienced participants were adults ($n = 11$, mean age 24.3 ± 2.8 yr, 3 female) initially chosen through self-reporting, who began playing Pokémon between the ages of 5 and 8 yr. Experienced participants were included in the study if they continued to play the game throughout childhood and revisited the game as adults. Novice participants were chosen as similarly aged and educated adults who never played Pokémon ($n = 11$, mean age 29.5 ± 5.4 yr, 7 female). We validated their self-reported experience with Pokémon with data from a behavioural experiment, in which participants viewed 40 Pokémon images from the original Nintendo game and identified each image by name (from 5 choices). Experienced participants ($n = 11$) significantly outperformed novices ($n = 9$) in their naming ability of Pokémon (Student's t-test, $t(18) = 18.2$, $P < 0.001$, Cohen's $d = 8.18$; Fig. 1a). Despite not being able to name Pokémon, novices are capable of visually distinguishing and individuating Pokémon characters (Supplementary Fig. 1).

All participants underwent fMRI while viewing faces, bodies, cartoons, pseudowords, Pokémon, animals, cars and corridors (Fig. 1b). Cartoons and animals were chosen to create a strict comparison to the Pokémon stimuli and the other categories were included as they have well-established and reproducible spatial topography across the VTC[49]. Stimuli were randomly presented at a rate of 2 Hz in 4 s blocks, each containing 8 images from a category. Participants performed an oddball detection task to ensure continuous attention throughout the scan. Participants completed six runs with different stimuli from these categories.

We first examined whether childhood experience affects the representation of category information in the VTC, which was anatomically defined in each participant's native brain (see the MVPA section in the Methods). Therefore, at the individual level, we measured the representational similarity among distributed VTC responses to the eight categories across runs. Each cell in the representational similarity matrix (RSM) is the voxelwise correlation between the distributed VTC responses to different images of
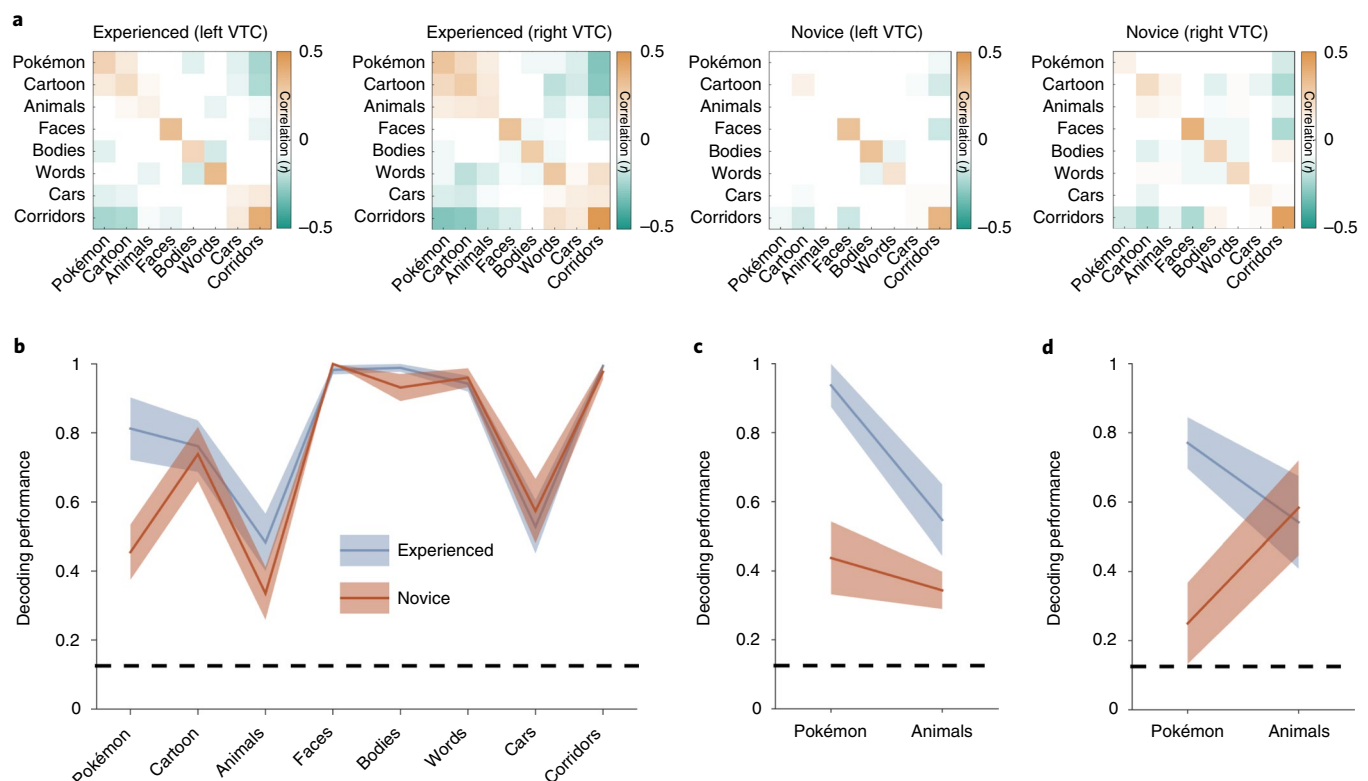
Fig. 2 | Experienced participants demonstrate a consistent and distinct representation for Pokémon compared to novices. a, RSMs calculated by correlating distributed responses (z-scored voxel betas) from an anatomical VTC ROI across split halves of the fMRI experiment. Positive values are presented in orange, negative values in green and near-zero values in white (see the colour scale, which applies to all four RSMs). **b**, The decoding performance from the winner-takes-all classifier trained and tested on split halves of the fMRI data from the bilateral VTC. The shaded region shows s.e.m. across participants within a group (experienced participants, $n = 11$; novices, $n = 11$). The dashed line indicates the chance level performance; decoding performance is represented as a fraction of 1, with 1 corresponding to 100% decoding accuracy. **c,d**, The decoding performance from distributed bilateral VTC responses for experienced ($n = 4$) and novice ($n = 5$) participants in the original oddball task (**c**) and when brought back to undergo an additional fMRI experiment with an attention-demanding two-back task (**d**). The same participants are shown in **c** and **d**.

the same category (diagonal) or different categories (off-diagonal) across split halves of the data. Then, we averaged the RSMs across the participants of each group and compared across groups to examine the representational structure of distributed VTC responses in experienced and novice participants. Averaging across participants of each group allowed us to visualize consistency within a given group. We then used decoding approaches to quantify these representational structures in individual participants, described below.

We hypothesized that the representation similarity of distributed VTC responses will have one of four outcomes. The first is the null hypothesis: Pokémon will not elicit a consistent response pattern in the VTC in any group and will have near-zero correlation with other items of this and other categories. Second, the animate hypothesis: Pokémon, which have faces, limbs and resemble animals to some extent, will have positive correlations with animate categories, such as faces, bodies and animals. Third, the expertise hypothesis: if Pokémon are processed as a category of expertise, then distributed responses to Pokémon will be most correlated with distributed responses to faces, as the expertise hypothesis predicts that expert stimuli are processed in face-selective regions[50,51]. Fourth, the distinctiveness hypothesis: as Pokémon constitute a category of their own, they will elicit a unique response pattern. Thus, correlations among distributed responses to different Pokémon will be positive and substantially higher than the correlation between Pokémon and items from other categories.

Experienced participants differ markedly from novices in their distributed VTC response patterns to Pokémon (see the RSM in

Fig. 2a). Unlike novices, who demonstrate little to no reproducible pattern for Pokémon in the VTC, consistent with the null hypothesis (mean Pearson correlation ± s.d., $r = 0.1 \pm 0.06$, $n = 11$), experienced participants demonstrate a significantly more reproducible response pattern for Pokémon ($r = 0.27 \pm 0.11$; significant between-group difference: $t(20) = 4$, $P < 0.001$, $d = 1.8$). Furthermore, distributed responses to Pokémon were distinct from those of other categories in experienced participants. We quantify this effect by calculating the mean dissimilarity ($D = 1 - r$) of distributed responses to Pokémon from other categories. Distributed responses to Pokémon are significantly more different from distributed responses to the other categories in experienced participants than controls ($t(20) = 4.4$, $P < 0.001$, $d = 1.8$). Interestingly, in experienced participants, Pokémon response patterns are significantly dissimilar (all $t(20) < 4.2$, all $P < 0.001$, all $d > 0.89$) from those of faces ($D \pm$ s.d. $= 0.97 \pm 0.08$), bodies (1.1 ± 0.08) and animals (0.9 ± 0.07) despite Pokémon having faces, bodies and animal-like features themselves. In contrast, when excluding Pokémon, groups do not have a significantly different D between distributed responses to other pairs of categories ($t(20) = 0.52$, $P = 0.6$, $d = 0.19$).

To quantify the information content in the VTC and test whether it varies with experience, we constructed a winner-takes-all classifier[52] trained and tested on independent halves of the data to determine whether the category of the stimulus the participant is viewing can be classified from distributed VTC response patterns and whether this performance depends on childhood experience. Although classification in both groups was above chance for all categories,
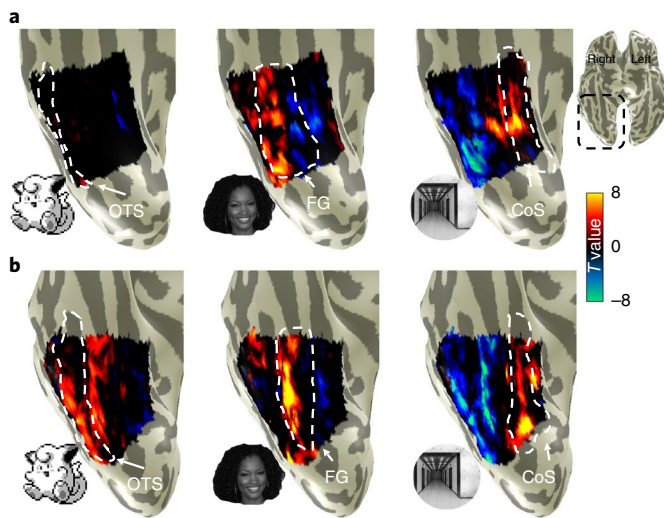
Fig. 3 | Distinct cortical representation for Pokémon in experienced participants. a,b, Unthresholded parameter maps displayed on the inflated ventral cortical surface zoomed on VTC (see inset for the location on a whole-brain map) in an example novice participant (26-year-old female; **a**) and an example experienced participant (26-year-old male; **b**) for the contrasts of Pokémon, faces and corridors, each versus all other stimuli. Dashed lines delineate cortical folds; OTS, occipitotemporal sulcus; FG, fusiform gyrus; CoS, collateral sulcus.

performance varied by category and group (Fig. 2b). An ANOVA with the factors stimulus category and participant group (experienced/novices) on classification performance revealed significant effects of stimulus category (F-test, $F(1,7)=36.5$, $P<0.001$, $\eta^2=0.61$) and group ($F(1,7)=6$, $P=0.05$, $\eta^2=0.036$) and a significant interaction between group and category ($F(1,7)=3.1$, $P=0.004$, $\eta^2=0.12$). Importantly, this interaction was driven by the differential classification of brain responses to Pokémon across groups: in experienced participants, performance was $81.3\pm9\%$ (mean ± s.e.m.), which was significantly higher (post hoc t-test, $P=0.001$, $d=1.5$) than the $45\pm8\%$ in novices (almost double; Fig. 2b). Classification performance was also numerically higher for bodies and animals in experienced compared to novice participants, but this difference—along with classification performance for other stimuli—was not significantly different between groups (Fig. 2c). Lastly, although there were more male ($n=8$ male, $n=3$ female) experienced participants and more female novices ($n=4$ male, $n=7$ female), behavioural naming of Pokémon as well as decoding accuracy were not higher in experienced males than in experienced females (Supplementary Fig. 2b,c). These analyses suggest that childhood experience with Pokémon generates a reliable and informative distributed representation of Pokémon in the VTC.

Can differences in attention account for this pattern of results? Although it has been argued that attention can boost signals to the category of expertise[53], others have argued that attention to the expert category does not explain the enhanced cortical activity to viewing it[54]. Furthermore, it is unclear whether attention alone could induce a distinct distributed response that could be reliably decoded. Nonetheless, to test the possibility that selective attention to Pokémon in experienced participants is driving the improved classification performance, we invited a subset of experienced and novice participants to perform an additional fMRI experiment with the same stimuli, using a different task that was equally attention-demanding across all categories. Here, participants performed a two-back task, indicating whether an image repeated after an intervening image. In this second fMRI experiment, the classification of Pokémon from distributed VTC responses was again significantly

higher in experienced ($77\pm14\%$) than novice participants ($25\pm23\%$; between-group difference: $t(7)=5.6$, $P<0.001$, $d=3.9$; Fig. 2d). In contrast, controlling attention generated a similar classification of animals across groups (Fig. 2d). These data suggest that differential attention to Pokémon across groups is not the driving factor leading to the distinct and reproducible representation of Pokémon stimuli in the VTC of experienced participants.

Visual features make different predictions for the emergent location of cortical responses. Results of the previous MVPA suggest that intense childhood experience with a novel visual category results in a reproducible distributed response across the human VTC that is distinct from other categories. An open question is whether Pokémon generate distributed response patterns with similar topographies across experienced participants. Therefore, we generated statistical parametric maps that contrast the response to each category versus all others (units: T values) and compared across groups. In typical adults, stimulus dimensions such as eccentricity[19,55], animacy[37,56], size[26] and curvilinearity[25] are mapped to a physical, lateral−medial axis across the VTC[49]. As the responses to faces on the lateral VTC and places on the medial VTC generate the most differentiated topographies, we analysed the properties of Pokémon stimuli for these attributes relative to faces and places. Thus, we used these metrics to generate predictions for the emergent topography of a Pokémon representation in experienced participants.

As expected, in both groups, unthresholded contrast maps demonstrating preferences for faces and places showed the typical topography in relation to major anatomical landmarks. That is, despite both anatomical and functional variability between participants, preference for faces was found in the lateral FG and preference for places in the collateral sulcus (CoS), as illustrated in Fig. 3. However, striking differences can be observed when examining VTC responses to Pokémon. In novices, Pokémon do not elicit preferential responses in the VTC, as with faces or corridors (Fig. 3a). In contrast, an example experienced participant demonstrated a robust preference for Pokémon in the lateral FG and occipito-temporal sulcus (OTS; Fig. 3b). This pattern was readily observable in all the other experienced participants (Supplementary Fig. 3). Given that these data suggest childhood experience with Pokémon results in a spatially consistent topography for Pokémon across individuals, we next asked: what attributes of Pokémon drive this topography?

The stimuli of faces, corridors and Pokémon used in the localizer experiments were submitted to a variety of analyses with the goal of ordering these categories linearly along different feature spaces (see the Image statistics analyses section in the Methods). Stimuli were analysed for physical attributes of foveal bias, that is, the retinal size of images when fixated on across a range of typical viewing distances, and rectilinearity, using the Rectilinearity Toolbox[25], which evaluates the presence of linear and curved features at a range of spatial scales. Stimuli were also rated for attributes of size and animacy, by independent raters, described below.

We first evaluated image-based attributes. To estimate retinal size, we assumed that people foveate on the centre of an item, and evaluated retinal image sizes as one would interact with these stimuli on a daily basis. On a GameBoy, Pokémon tend to span the central 2° when foveated on; the distribution of Pokémon retinal sizes partially overlapped with, but was significantly smaller and thus more foveally biased than that of faces ($t(286)=15.7$, $P<0.001$, $d=1.85$; Fig. 4a). Both Pokémon and faces generate significantly smaller retinal images than corridors ($ts>20$, $Ps<0.001$, $ds>2.5$), which often occupy the entire visual field and thus generate large retinal images that extend to the peripheral visual field even when foveated on. This metric predicts that, if retinal image size ranking (from foveally biased to peripherally biased: Pokémon, faces,
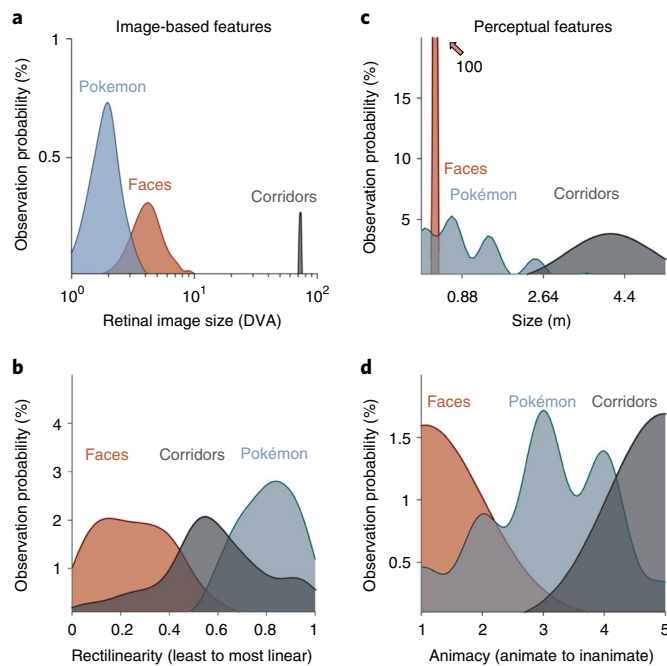
Fig. 4 | Different visual feature statistics predict different cortical locations for Pokémon. a, Distributions of retinal image sizes produced by Pokémon (blue), faces (orange) and corridors (grey) in a simulation that varied viewing distance across a range of sample stimuli. DVA, degrees of visual angle. *X* axis shows log-scale DVA. **b**, Distributions of the relative rectilinearity scores of faces, corridors and Pokémon, as measured using the Rectilinearity Toolbox[25] (0, least linear; 1, most linear). **c**, Distributions of the perceived physical size of Pokémon (from 28 raters) and of the physical sizes of faces and corridor stimuli. The distributions of face and corridor size were produced using Gaussian distributions with standard deviations derived from either anatomical or physical variability within the stimulus category (see Methods). The face distribution extends to a value near 100% (the natural variation of face size is very narrow compared to other stimuli). **d**, Distributions of the scores of perceived animacy collected from a group of 42 independent raters who rated the stimuli of faces, Pokémon and corridors for how 'living or animate' these stimuli were perceived to be (1, animate; 5, inanimate).

corridors) drives the generation of distributed responses, Pokémon representations should be the most foveal representations and corridors the most peripheral ones. As the representation of retinal eccentricity in the VTC moves from lateral (most foveal) to medial (most peripheral), this attribute predicts that emergent activations to Pokémon in experienced participants' VTCs should lie on the OTS, largely lateral to, but partially overlapping face-selective regions on the FG (Fig. 4a).

Analyses of the rectilinearity of Pokémon, faces and corridors (see the Image statistics analyses section in the Methods) show that faces are the least linear of these 3 categories and the original Pokémon stimuli, which were constructed from large square pixels, are the most rectilinear when compared to both faces ($t(383) = 37.5$, $P < 0.001$, $d = 3.9$) and corridors ($t(298) = 10.7$, $P < 0.001$, $d = 1.23$). As the rectilinearity cortical axis is also arranged from curvy (lateral VTC) to rectilinear (medial VTC), rectilinearity predicts that preference for Pokémon in the VTC should lie medial to both the face-selective cortex and place-selective cortex, potentially in the CoS or the parahippocampal gyrus (Fig. 4b). We also evaluated perceived rectilinearity in a separate group of novices (Supplementary Fig. 4). While perceived rectilinearity situates Pokémon between faces (the most curvy) and

corridors (the most linear), Pokémon are still perceived to be significantly more linear than faces ($t(49) = 6.9$, $P < 0.001$, $d = 1.44$). Thus, both image-based and perceived rectilinearity predict that Pokémon selectivity will be medial to face selectivity.

Next, we generated predictions on the basis of perceived visual features (see the Image statistics analyses section in the Methods), using independent raters. To obtain an estimate of how perceived features may affect an untrained visual system, which would mimic the state of experienced participants when they first began playing the game, we chose raters who were not heavily experienced with Pokémon. The first was an estimation of real-world size, which is different from retinal size. For example, a corridor viewed from far away is perceived as large despite subtending a small retinal image. Although the real-world size of faces and corridors are readily estimable, Pokémon do not exist in the real world. For Pokémon, size was evaluated in two ways. First, the game provides information on the physical size of each Pokémon. Second, we evaluated the participants' perception of the sizes of Pokémon. Twenty-eight participants, who did not take part in the main experiment, were asked to rate the size of various Pokémon using a 1–7 scale, with each number corresponding to reference animals of increasing size. Possible sizes ranged from much smaller than faces (for example, 1 cm; ant) to larger than a corridor (for example, 7 m; dinosaur). The raters' perceptions of Pokémon size were largely consistent with the game's provided physical sizes (average game size: 1.2 ± 0.95 m; average rater size: 0.82 ± 0.28 m). Raters perceive Pokémon to be larger than faces ($t(176) = 26.3$, $P < 0.001$, $d = 5.4$) but smaller than corridors ($t(176) = 16.9$, $P < 0.001$, $d = 3.48$); see Fig. 4c. As real-world size also shows a lateral (small) versus medial (large) topography, perceived size predicts that voxels preferring Pokémon in the VTC should lie between face-selective regions on the FG and the place-selective cortex in the CoS, that is, on the medial FG.

Lastly, we evaluated the perceived animacy of our stimuli. A separate group of 42 independent participants rated the perceived animacy of stimuli using a scale of 0–5 (0 most inanimate; 5 most animate). Pokémon are perceived as intermediately animate (Fig. 4d), less animate than faces ($t(82) = 13.2$, $P < 0.001$, $d = 2.88$), but more animate than corridors ($t(82) = 14.5$, $P < 0.001$, $d = 3.18$). As animate stimuli are localized to the lateral VTC and inanimate stimuli to the medial VTC, this metric predicts that the locus of preferential responses for Pokémon lies, as for perceived size, between the face-selective cortex and place-selective cortex on the medial FG.

Location of novel cortical responses in experienced participants supports eccentricity bias theory. To test these predictions, we produced contrast maps for Pokémon versus all other stimuli in each participant. Using cortex-based alignment (CBA), we transformed each participant's map to the FreeSurfer[57] average cortical space, where we generated a group average Pokémon-contrast map. For visualization, we projected the map onto an individual experienced participant's cortical surface. These results revealed four main findings. First, in experienced, but not novice participants, we observed that preference for Pokémon reliably localized in the OTS. As illustrated in the average experienced participants' Pokémon contrast map (Fig. 5a), higher responses to Pokémon versus other stimuli were observed in the OTS and demonstrated two peaks on the posterior and middle portions of the sulcus. Second, we compared Pokémon activations to those of faces by delineating the peaks of face selectivity in the average contrast maps for faces in each group (Fig. 5a, white outlines). This comparison reveals that for experienced participants, Pokémon-preferring voxels partially overlapped face-selective voxels on the lateral FG and extended laterally to the OTS, but never extended medially to the CoS, where place-selective activations occur (Fig. 3). Third, we compared the volume of category selectivity for Pokémon

in each group. Pokémon-selective voxels were any voxels within the VTC that were above the threshold ($T > 3$) for the contrast of Pokémon versus all other stimuli. Pokémon-selective voxels were observed in the lateral FG and OTS of all 11 individual experienced participants. In contrast, although some scattered selectivity could be observed for Pokémon in four novice participants, it was not anatomically consistent. Thus, it did not yield any discernible selectivity for Pokémon in the average novice selectivity map (Fig. 5a). An ANOVA run with the factors group and hemisphere on the volume of Pokémon-selectivity in the VTC revealed a main effect of group ($F(1,1) = 32.75$, $P < 0.001$, $\eta^2 = 0.45$), but no effects of hemisphere, nor any interaction ($Fs(1,1) < 0.67$, $Ps > 0.41$, $\eta^2s < 0.016$). The median volume in experienced participants was sixfold larger than novices (Fig. 5a), with most (7 of 11) novice participants having close to zero voxels selective for Pokémon. This difference in volume between groups was not driven by gender differences (Supplementary Fig. 2a). Fourth, we compared the lateral–medial location of Pokémon selectivity relative to face and place selectivity to directly assess theoretical predictions. Therefore, we partitioned the VTC in each experienced participant into four anatomical bins from lateral (OTS) to medial (CoS); see the inset in Fig. 5b. Within each bin we extracted the mean T value for the contrast of either Pokémon, faces or corridors. Curves fitted to average T values across bins demonstrate that (1) peaks in these curves are the most lateral for Pokémon, intermediately lateral for faces and medial for corridors (Fig. 5b) and (2) Pokémon-selectivity peaks are located significantly more laterally in the VTC than those of faces ($t(20) = 2.88$, $P = 0.009$, $d = 1.23$). Together, this pattern of results is consistent with only the predictions of the eccentricity bias theory for the development of VTC topography (Fig. 4a).

Another prediction of the eccentricity bias hypothesis is that the region displaying Pokémon selectivity in experienced participants should have a foveal bias. A foveal bias predicts that population receptive field (pRF) centres will cluster closer to, and result in greater coverage of, the centre of the visual field compared to the periphery. This coverage is in contrast to the tiling of the visual field observed in the primary visual cortex[31,58], where pRF centres spread across the visual field. To test this hypothesis, six experienced participants participated in a retinotopic mapping experiment (see the pRF mapping section in the Methods). Using pRF modelling[59], we identified the region of the visual field that drives the responses of each voxel. We then generated visual field coverage maps showing how the set of pRFs in the Pokémon-selective cortex in each participant tiles the visual field. Experienced participants demonstrate the typical lateral–medial gradient of foveal to peripheral bias in the VTC (Supplementary Fig. 5). In all experienced participants, we find that the Pokémon-selective cortex shows a more prominent coverage of central portions of the visual field compared to the periphery. For example, in the right hemisphere of experienced participants, the Pokémon-selective cortex shows a contralateral coverage of the visual field, whereby the peak density of visual field coverage is within 4° of the centre of the visual field. With the exception of one participant, the region with the highest pRF density includes the fovea (Fig. 6a). Compared to the eccentricity bias of other categories, such as faces and corridors, overall Pokémon-selective voxels tend to be the most foveally biased, followed by face-selective voxels and lastly corridor-selective voxels, which are the most peripherally biased (Fig. 6b). Examining individual participants, two show mean pRF eccentricities in Pokémon-selective voxels that are more foveally biased than the face-selective cortex, whereas the remaining four participants show no meaningful difference (Supplementary Fig. 6). Overall, the foveal bias of Pokémon-selective pRFs provides evidence consistent with the idea that the eccentricity of retinal images produced by Pokémon biases their cortical responses to emerge in the lateral VTC where foveal pRFs exist.
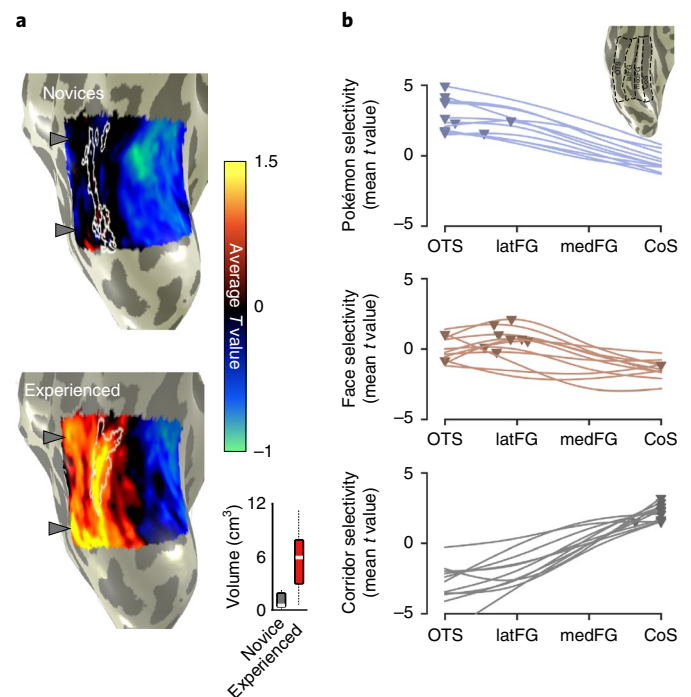
Fig. 5 | Average contrast maps for Pokémon; and anatomical localization reveals lateral VTC responses in experienced subjects. a, Average contrast maps for Pokémon in novice and experienced participants. For each participant, T-value maps were produced for the contrast of Pokémon versus all other stimuli. These maps were aligned to the FreeSurfer average brain using cortex-based alignment (CBA). On this common brain surface we generated a group-average contrast map by averaging maps across all novice participants and all experienced participants. Group-average maps are shown on an inflated right hemisphere of one of our participants, zoomed in on the VTC. White outlines show group-average face-preferring voxels (average $T > 1$) from each respective group. Grey arrows show two peaks in the Pokémon-selectivity maps of experienced participants; the same arrows are shown next to the novice map for comparison. Inset: box plots show the mean (white line), 25% and 75% quartiles (boxes) and range (black dotted line) of the selectivity volume in novices and experienced participants. **b**, Curves fitted to the mean selectivity for Pokémon, faces or corridors, averaged in one of four anatomically defined regions extending from the lateral to medial VTC (illustrated in the inset for an example participant). Each line represents a participant and the triangles show the peak selectivity values. The peaks for the Pokémon-selectivity curves are significantly more lateral than the peaks for face selectivity. The most lateral ROI is the OTS extending from the inferior temporal gyrus (ITG) to the medial aspect of the OTS. The lateral FG (latFG) ROI includes the lateral FG and ends medially at the MFS; the medial FG (medFG) bin extends from the MFS to the lateral edge of the CoS; the CoS bin includes the CoS up to the lateral edge of the parahippocampal gyrus.

How does experience affect the amplitude of responses to Pokémon?. To further understand how novel childhood experience has impacted cortical representations in the VTC, we asked two questions. First, are the emergent responses to Pokémon in experienced participants specific to the Pokémon characters that participants have learned to individuate, or will similar patterns emerge for any Pokémon-related stimulus from the game? Second, how does visual experience change the responsiveness of the OTS to visual stimuli?

To address the first question, a subset of experienced participants ($n = 5$) participated in an additional fMRI experiment in which they viewed other images from the Pokémon game (see the Pokémon
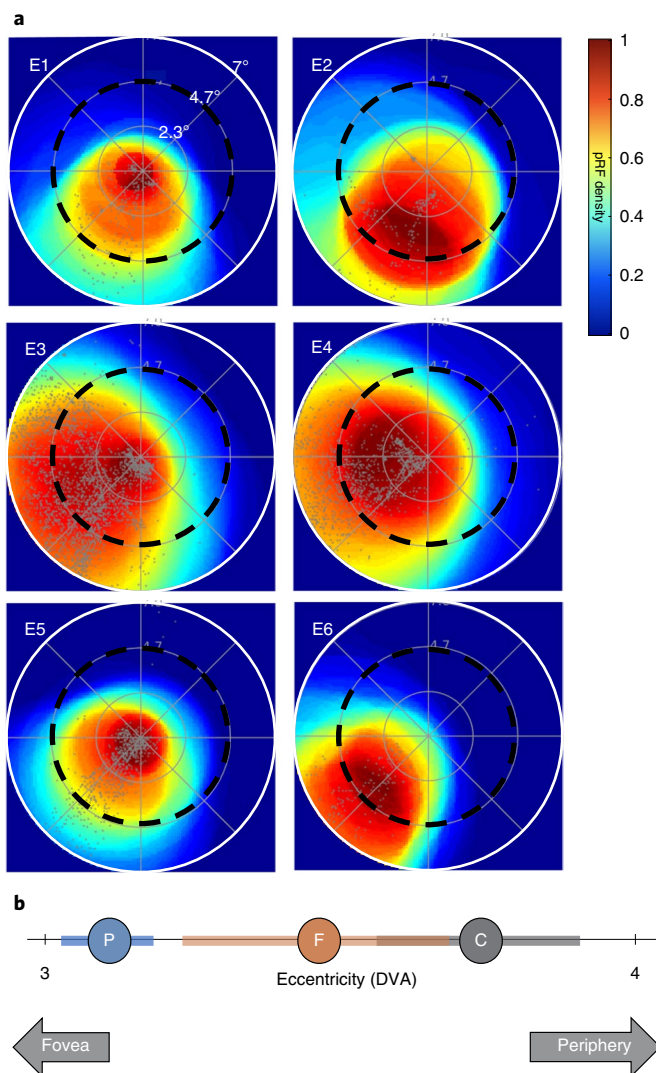
a

b

Fig. 6 | pRF modelling reveals that the Pokémon-selective cortex is foveally biased. a, Density plots (see colour scale) representing the visual field coverage by pRFs of Pokémon-selective voxels from the right hemisphere in 6 experienced participants. Each plot shows data from a single participant (E1–E6). Density is normalized to the maximum in each participant. Grey dots show pRF centres and black dashed circles represent 4.7° of eccentricity. Only voxels that had more than 10% variance explained by the pRF model were included. **b**, The mean (circles) and s.e.m. (coloured bars) of pRF eccentricity averaged within the Pokémon- (blue), face- (brown) and corridor-selective (grey) cortex across participants. Positions towards the left of the line are closer to the centre of the visual field (fovea), measured in DVA.

scenes and pixelated faces fMRI control experiment section in the Methods). In this experiment, participants completed a blocked experiment with two categories of images: images of places (for example, navigable locations) from the Pokémon game as well as downsampled face stimuli (to resemble 8 bit game imagery). Images were presented in 4 s blocks at a rate of 2 Hz while participants performed an oddball task. Results show that places from the Pokémon game drive responses in the CoS, not the OTS or the lateral FG. In other words, they produce the typical pattern of place-selective activations (Fig. 7). In contrast, Pokémon-selective voxels in each participant (Fig. 7, black contours) have minimal to no selectivity for places from the Pokémon game, further demonstrating the specificity of Pokémon-selective voxels to Pokémon characters.
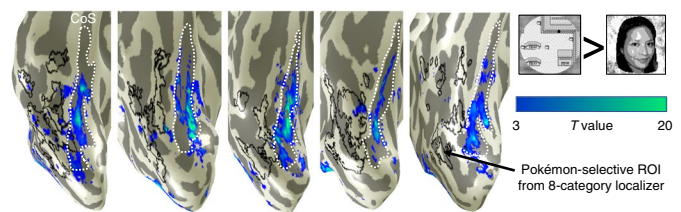

Fig. 7 | Places from the Pokémon game elicit typical place-selective activations in experienced participants. The maps show higher responses to Pokémon scenes than pixelated faces in 5 experienced participants from a follow-up fMRI experiment in which participants viewed downsampled face stimuli (resembling 8 bit game imagery) and scenes from the Pokémon games. The colour bar indicates the T value at each voxel from the threshold ($T > 3$) over the range denoted in the colour scale. Pokémon-selective voxels (outlined in black) do not preferentially respond to Pokémon scenes versus faces; instead, voxels preferring Pokémon scenes versus faces are in the typical location for place selectivity[18], namely, the CoS (outlined in white).

To answer the second question, namely, how experience shapes responses in the OTS, we quantified the response amplitude of Pokémon-selective voxels in both experienced and novice participants. To ensure that the region of interest (ROI) was defined independently from the individual's data, we employed a leave-one-out approach[60] in which we produced a group-defined Pokémon ROI from ten experienced participants by transforming ROIs from individuals to the FreeSurfer cortical average using CBA and then using CBA to project the group ROI to the left-out individual's brain and examining its responses. This procedure was repeated for each experienced participant. For novice participants, we transformed the group ROI produced from all the experienced participants into each individual novice's brain. To ensure that we did not extract a signal from cortex that was already selective for another category, we removed any voxels that were selective for other categories from the group Pokémon ROI for each participant. From this independently defined Pokémon ROI, we extracted the percent signal change from the eight-category fMRI experiment.

As expected, experienced participants have higher responses to Pokémon compared to other categories (Fig. 8a, Supplementary Fig. 7), with lower responses to cartoons and animals and even lower responses to faces, bodies and other stimuli. However, in this putative Pokémon-selective region, novice participants show the highest amplitude of response to animals, then to cartoons and words and lower responses to the other stimuli (Fig. 8a). Although it may be tempting to conclude that Pokémon-selective voxels emerge from voxels with a pre-existing preference for animals, average cortical maps of animal selectivity (Supplementary Fig. 8) reveal similar animal selectivity both lateral and medial to the face-selective cortex in both novice and experienced participants. Furthermore, perceived animacy ratings (Fig. 4d) suggest that Pokémon should have emerged between the face-selective cortex and place-selective cortex, which would correspond to the portion of animal selectivity that is medial to the face-selective cortex. However, contrary to these predictions, Pokémon selectivity is observed overlapping and lateral to the face-selective cortex (Fig. 5). These results suggest that animacy alone is not driving the emergent locus of Pokémon selectivity.

To test the expertise hypothesis[50,51,54], and because we found that preference for Pokémon stimuli in experienced participants partially overlapped the face-selective cortex, we also measured responses to our stimuli in the face-selective regions in the posterior fusiform gyrus (pFus-faces) and medial fusiform gyrus (mFus-faces)[61]. We used a split-halves analysis, defining face-selective voxels on the lateral FG from the even runs and extracting the percent signal change from the odd runs, to ensure
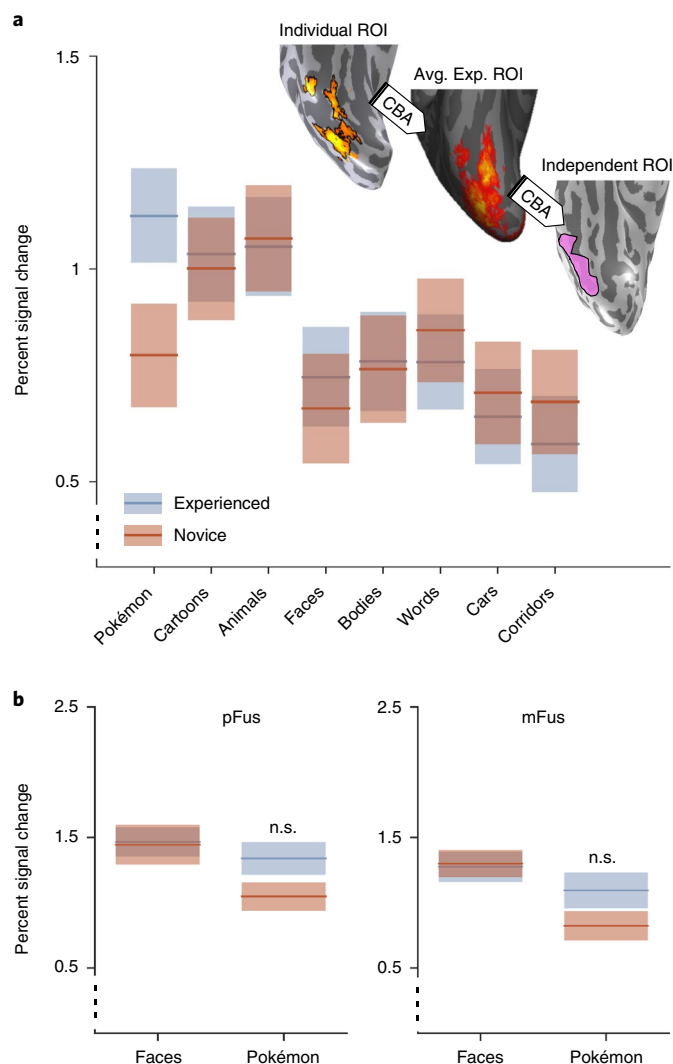
Fig. 8 | Response properties of the VTC vary with childhood experience with Pokémon. a, fMRI responses (percent signal change) measured from an independent definition of Pokémon-selective cortex in both experienced and novice participants. ROI, region of interest. **b**, Responses from bilateral pFus- and mFus-faces to faces and Pokémon in experienced and novice participants. Face-selective voxels were defined using odd runs of the fMRI experiment, and percent signal change was calculated from the even runs. Experienced and novice participant responses to both faces and Pokémon in pFus-faces are not significantly different. Responses to faces and Pokémon between groups in mFus are also non-significant. The black dotted *y* axis denotes that the axis does not begin at zero. Shaded regions are s.e.m. Avg. exp. ROI, average experienced ROI; n.s., non-significant.

independence of data. Experienced and novice participants did not exhibit differences in their response to faces in either pFus- or mFus-faces, bilaterally (Fig. 8b). An ANOVA with the factors group and stimulus type revealed no significant interaction ($F_{(1,1)} = 1.4$, $P = 0.31$, $\eta^2 = 0.03$) for functional responses in pFus-faces. Likewise, a similar ANOVA run for responses in mFus-faces also revealed a non-significant interaction ($F_{(1,1)} = 1.36$, $P = 0.25$, $\eta^2 = 0.02$). Response amplitudes to Pokémon in pFus- and mFus-faces were numerically higher in experienced compared to novice participants, but not significantly so (Fig. 8b). Together, analyses of response amplitude illustrate that extensive experience of Pokémon during childhood results in higher responses to Pokémon in adulthood relative to other stimuli leading to the

emergence of Pokémon selectivity in and around the OTS and largely outside other regions of category selectivity.

Discussion

By examining cortical representations in adults who have had visual experience with a specific, artificial visual category since childhood, we found that participants who have extensive experience with Pokémon characters, beginning as early as 5 yr old, demonstrate distinct response patterns in high-level visual cortex that are consistent across participants. Category-selective responses to Pokémon in all experienced participants occupied the posterior and middle extent of the OTS, largely lateral to face-selective cortex, and responded selectively to learned Pokémon characters rather than general imagery from the Pokémon game. Note that we do not assert that our results should be interpreted as a new Pokémon functional module in the OTS of experienced participants on par with the face-selective cortex[7]. Instead, our data underscore that prolonged experience starting in childhood can lead to the emergence of a new representation in the VTC for a novel category with a surprisingly consistent functional topography across individuals. We demonstrated that this topography is consistent with the predictions of the eccentricity bias theory for two reasons: the small stimuli that required foveal vision during learning biased the emergent representations towards the lateral VTC; and Pokémon-selective voxels in experienced participants show a foveal bias. Together, our data show that shared, patterned visual experience during childhood, combined with the inherent retinotopic representation of the visual system, results in the shared brain organization observed in the adult high-level visual cortex.

The nature of Pokémon as a stimulus category is an interesting one, because it could be seen as similar to other ecological stimuli such as faces or words: the game entails repeated, prolonged and rewarded experience individuating visually similar, but semantically distinct exemplars. This is similar to other stimulus categories for which there is ecological pressure, or interest, to individuate among a visually homogeneous category such as faces, birds or cars[50]. Our data suggest that individuals who have had life-long experience individuating Pokémon characters develop a novel representation for this learned category, demonstrating the plasticity of high-level visual cortex outside of the face-selective regions. This was supported by robust decoding results and a consistent spatial topography of selectivity for Pokémon in experienced but not novice participants.

Our findings suggest that early childhood visual experience shapes the functional architecture of high-level visual cortex, resulting in a unique representation whose spatial topography is predictable. One should exercise caution when comparing the current findings with the effects of visual expertise that was acquired in adulthood[62–64], as the current study focuses on participants whose visual experience began at a young age. There are several differences that distinguish our current results from such previous research on expertise. First, childhood experience led to the development of a new representation for Pokémon that was anatomically consistent across participants and was coupled with increased responses and selectivity for Pokémon in the OTS. As the same piece of cortex did not show selectivity in novices, this suggests that extensive experience individuating a novel stimulus beginning in childhood is necessary and sufficient for the development of a new representation in the VTC. Although previous investigations of category training in adults also demonstrated increases in voxel selectivity for the learned category[63,65], different from our data, these activations were not anatomically consistent across individuals and occurred either in the object-selective cortex or outside the visual cortex entirely (in the prefrontal cortex). Second, Pokémon elicited numerically, but not significantly, higher responses in the face-selective cortex of our experienced participants than in our novices. Although these

data are consistent with previous reports of increased responses in face-selective areas on the FG to stimuli of expertise gained in adulthood[54,62,64], it is unclear from our data whether these increased responses are due to experience or due to the fact that Pokémon have faces (Supplementary Fig. 9). Third, previous research has shown that learning contextual and semantic features of novel objects in adulthood (for example, this object is found in gardens) can influence VTC representations[66]. Thus, part of the emergent representation for Pokémon in experienced participants may have stemmed not only from visual features, but also contextual and semantic information learned about Pokémon. In other words, the representation of Pokémon may include additional semantic and contextual information, such as its habitat and characteristics, that can be investigated in future studies. Lastly, developmental work in humans with parametrically morphed stimuli has shown that improved perceptual discrimination among face identities in adulthood is linked with increased neural sensitivity (lesser adaptation) to face identity in the face-selective cortex from childhood to adulthood[67]. Higher neural sensitivity is thought to be due to narrower neural tuning. Future research can test whether childhood experience with Pokémon also affects neural tuning by measuring adaptation to parametrically morphed Pokémon in experienced versus novice participants[68].

Our findings also have interesting parallels with research on the development of reading abilities in children. First, visual experience with Pokémon began between the ages of 5 and 8 yr old, similar to the ages during which reading ability rapidly improves[69]. Second, it is interesting that, like looking at Pokémon, reading words requires foveation and words typically subtend small retinal images. Third, similar to our findings, research on the development of reading has shown that the word-selective cortex emerges during childhood in the lateral OTS and distributed representations in the OTS and the lateral FG become more informative from childhood to adulthood with increasing reading experience[52,70]. Thus, our data together with the research on the development of reading, suggest that the critical window for sculpting unique response patterns in the human VTC extends to at least school age.

Our results converge with previous research in macaques that offers compelling evidence for three important developmental aspects of high-level visual cortex. First, its organization is sensitive to the timing of visual experience[43]. In macaques, learning new visual categories in juveniles but not adults resulted in new category-selective regions for the trained stimuli. This suggests that there may be a critical developmental period for cortical plasticity in high-level visual cortex. Second, in macaques, early visual experience led to the formation of category selectivity in consistent anatomical locations[33] and deprivation of visual experience with stimuli such as faces results in no face-selective cortex[32]. Third, eccentricity may be a strong prior that constrains development in high-level visual cortex[19,31]. For example, in the macaque visual cortex, a proto-eccentricity map is evident early in infant development[39,40].

However, our data also highlight key developmental differences across species. First, the critical window of cortical plasticity in high-level visual cortex may be more extended in humans than macaques. In humans, extensive discrimination training in adults results in changes in amplitudes[62,71] and distributed representations[63] in high-level visual cortex, but in adult macaques, responses[43] and distributed representations[72] do not change, even as the monkeys become behaviourally proficient at the task (but see ref. [73], which shows increases in the number of inferotemporal neurons responsive to trained stimuli in adult macaques). Second, the anatomical locus of the effects of childhood experience differs across species. In humans, the most prominent functional developments have been reported in the FG[20,74–78] and OTS[52,77], but in macaques they are largely around the superior temporal sulcus and adjacent gyri[32,33]. Notably, the FG is a hominoid-specific structure[79], which

underscores why development and training effects may vary across species. Third, the features of visual stimuli and how they interact with cognitive strategies that sculpt the brain during childhood may differ across species. That is, developmental predictions about the perceived animacy or size of a visual stimuli is readily queried only in humans.

The unique opportunity presented by Pokémon as a stimulus is the manner in which they vary from other visual stimuli in their physical (retinal image size, rectilinearity) and perceived (animacy, size) properties. Furthermore, the topography of the responses in experienced cortices was consistent across individuals, allowing us to ask which potential dimension of Pokémon visual features, either perceived or physical, may determine the anatomical localization of Pokémon responses in the VTC. The lateral location of this emergent representation, and its foveally biased pRFs, suggests that the act of foveating on images that subtend a small retinal image during childhood biases input towards regions in the lateral VTC that have pRFs that overlap the fovea. We posit that individuals experienced with Pokémon had enough patterned visual experience for this biased input to result in category selectivity.

Several aspects of our data indicate that retinal eccentricity is the dominant factor in determining the functional topography of the VTC. Although this by no means invalidates observations of other large-scale patterns describing the functional topography of high-level vision[5,25,26,35], our data suggest that retinal eccentricity is a key developmental factor in determining the consistent functional topography of the VTC across individuals, for several reasons. First, analyses of the physical and perceived properties of Pokémon characters suggest that the attribute of Pokémon stimuli that best predicts the location of peak selectivity for responses to Pokémon in the VTC was the experienced retinal eccentricity of Pokémon during childhood. Second, pRFs of VTC voxels selective for Pokémon in experienced participants showed a foveal bias. Future research examining how these representations emerge during childhood as participants learn these stimuli will be important for verifying that the foveal bias in the OTS precedes the emerging selectivity to Pokémon. Indeed, although Pokémon have almost linear statistics at the image level, individuals are capable of integrating holistically over a large number of pixels to perceive a curve. The extent to which this may be modulated by experience and potentially impact pRFs during learning is an interesting focus for future research. Third, although the representation of animacy also exhibits a lateral–medial organization in the human VTC, Pokémon are perceived as less animate than faces, but their representation appeared lateral to face-selective regions. In other words, if the animacy axis is continuous across the VTC, with animate representations in the OTS and inanimate on the CoS, one would expect Pokémon-selective voxels to be located medial to the face-selective cortex. Although we observe that such a medial region is capable of responding to animate stimuli (Supplementary Fig. 8), it does not become selective for Pokémon across development. Instead, Pokémon-selective voxels were found in the OTS, lateral to the face-selective cortex. Although high-level visual cortex is capable of distinguishing animate from inanimate stimuli[23], it might not be a continuous graded representation across the cortical sheet per se. Thus, the framework that human high-level visual cortex has a representation of retinal eccentricity[24,80], probably inherited from retinotopic input in earlier visual field maps, offers a parsimonious explanation of the development of Pokémon representations in the VTC. Future research can examine how becoming a Pokémon expert in childhood, which probably entails learning optimal fixation patterns on Pokémon stimuli, may further sculpt pRFs throughout development as observed in face- and word-selective cortex[31].

In conclusion, these findings shed light on the plasticity of the human brain and how experience at a young age can alter cortical representations. An intriguing implication of our study is that a

common extensive visual experience in childhood leads to a common representation with a consistent functional topography in the brains of adults. This suggests that how we look at an item and the quality with which we see it during childhood affects the way that visual representations are shaped in the brain. Our data raise the possibility that if people do not share common visual experiences of a stimulus during childhood, either from disease, as is the case in cataracts[28,81], or cultural differences in viewing patterns[82], then an atypical or unique representation of that stimulus may result in adulthood, which has important implications for learning disabilities[83,84] and social disabilities[85].

Overall, our study underscores the utility of developmental research, showing that visual experience beginning in childhood results in functional brain changes that are qualitatively different from plasticity in adulthood. Future research to examine the amount of visual experience necessary to induce distinct cortical specialization and determine the extent of the critical window during which such childhood plasticity is possible will further deepen our understanding of the development of the human visual system and its behavioural ramifications.

Methods

Participant details. Human participants were between the ages of 18 and 44 yr, with a mean and s.d. of 26.8 ± 4.8 yr. Participants were split into two groups: experienced ($n = 11$; age 24.3 ± 2.8 yr, 3 females) and novice participants ($n = 11$; age 29.6 ± 5.4 yr, 7 females). No statistical methods were used to predetermine sample sizes but group sizes are similar to those reported in previous publications[50,63,65]. The former group was selected initially through self-reporting, with the inclusion criteria that participants (1) began playing the original Nintendo Pokémon games between the ages of 5 and 8 yr on the hand-held GameBoy device, (2) continued to play the game and its series heavily throughout their childhood and (3) either continued to play the game into adulthood or revisited playing the game at least once in adulthood. Novice participants were chosen as individuals who had never played the Pokémon game and had little to no interaction with Pokémon otherwise. All participants completed a five-choice naming task designed to test the naming ability of the participants; experienced participants scored significantly better than novices (Fig. 1a). All participants provided informed, written consent to participate in the experiment per Stanford University's Internal Review Board.

Participants who completed behavioural rating experiments (separate from the 11 experienced and 11 novice participants) were undergraduates of Stanford University. All participants provided informed consent (per Stanford University's Internal Review Board) before completing the experiments (described below) and were compensated with extra class credits.

Behavioural naming task. Participants completed a 5-choice naming task with 40 randomly selected Pokémon. The experiment was self-paced and participants were told to choose the correct name of each Pokémon. Performance was evaluated as the percent correct accuracy, with experienced participants significantly outperforming novices (two-tailed t-test $t(18) = 17.3$, $P < 0.001$). The behavioural identification task was run after scanning had been completed to minimize the exposure of novice participants to Pokémon stimuli. Two novices were unable to complete the behavioural testing and were therefore excluded from the behavioural analyses in Fig. 1.

fMRI eight-category experiment. All participants underwent fMRI, completing six runs of the experiment with different images across runs. Each run was 218 s in duration and, across the 6 runs, and stimuli were presented in counterbalanced 4 s blocks, each containing 8 stimuli from a category shown at a rate of 2 Hz. Categories included faces, headless bodies, 8 bit Pokémon sprites from the original Nintendo game, animals, popular cartoon characters from the television of the late 1990s and early 2000s, pseudowords, cars and corridors. Stimuli, as illustrated in Fig. 1b, were presented on a textured background produced by phase scrambling a stimulus from another category[86]. Stimuli and backgrounds were counterbalanced so that all stimuli appeared with all other phase-scrambled categories as a background. This was done such that the Fourier amplitude spectra of all images across categories was matched as closely as possible, as in previous publications[31,67,86]. For the present stimulus set (144 images per category, 8 categories), we observe no significant difference in the summed Fourier amplitude between stimulus categories in an ANOVA with the grouping variable of stimulus category ($F(1,7) < 0.001$, $P = 1$). Stimuli were presented approximately in the centre of the background square but were jittered in size and position to minimize differences in the average image between categories. Stimuli were run through the SHINE toolbox[87] to remove luminance differences between categories or potential outlier stimuli. Animals (including insects) were chosen as

an animate stimulus category that closely resembles Pokémon, as most Pokémon characters were designed to resemble an animal or insect. Cartoons were chosen because they are another animated category that we hypothesize was experienced by both groups predominantly in childhood and was somewhat recognizable by both groups. Experience with cartoons was not quantified.

Pokémon scenes and pixelated faces fMRI control experiment. To evaluate whether the information we observed in the OTS in experienced participants for Pokémon was selective for the learned game characters and not just any stimulus from the game, a subset of experienced participants ($n = 5$, 3 of whom also completed the pRF mapping detailed in the following section) was invited back to the laboratory to complete a follow-up fMRI experiment containing two categories of images. The first consisted of scene stimuli from the Pokémon game. These scenes were map-like images extracted from the game through which the player has to navigate their character, including towns, natural landscapes and indoor spaces. These scenes were free of any images of people or Pokémon. The second category of stimuli was human faces (from the main experiment) downsampled to visually resemble the 8 bit pixelated images from the Pokémon game to match for low-level visual features. This was accomplished by resizing face stimuli to 70×70 pixels (approximately the original pixel size of Pokémon characters) and then resizing again using bicubic sampling to their original size of 768×768 pixels. Images were presented in 4 s blocks at a rate of 2 Hz and were interleaved with a blank baseline condition. There were 12 blocks of each condition in a run; each run was 162 s in duration and participants completed 3 runs. Data were analysed similarly to the main category localizer. Participants were instructed to press a button when a blank image appeared within a block of images (oddball detection task).

pRF mapping. To evaluate the visual field coverage of receptive fields within Pokémon-selective cortex, five experienced participants from the main experiment were invited to complete a pRF mapping experiment. Each participant completed 4 runs of a sweeping bar stimulus that traversed the 7×7 degrees of visual angle (DVA) screen in 8 different directions, as implemented in previous work[31]. For each voxel, the pRF model[59] was fitted using a circular Gaussian with a position (x, y), size (sigma) and a compressive spatial summation[88] to account for neural nonlinearities beyond the primary visual cortex. For each participant, the Pokémon-selective cortex was defined as any voxel in the lateral VTC demonstrating a T value of three or greater for the contrast of Pokémon versus all other stimuli from the localizer data. To derive pRF density maps to understand how the visual field is covered by the pRFs of each participant's Pokémon-selective ROI (Fig. 6), pRFs were plotted as circles using the fit size (standard deviation of the fit pRF Gaussian) and the visual field was coloured pointwise according to pRF density, where 1 corresponds to the maximum pRF overlap in that participant.

Behavioural two-back task. To ensure that novice participants are capable of visually detecting differences between Pokémon characters, and thus demonstrate that they do not perceive all Pokémon as an indistinguishable homogeneous object, we invited a separate group of novice participants ($n = 36$) to the laboratory to complete a two-back repetition detection task. Participants saw blocked images from one of three categories (faces, Pokémon, corridors) and were instructed to press a button when they detected an image repeat with an intervening image. Stimuli were presented at 2 Hz and 8 images were presented within a block, as was done during functional MRI. There could be 0 or 1 repeats within a given block and 50% of blocks contained a repeat; there were 40 blocks per category. Blocks with repeats were randomly assigned and counterbalanced across categories; category and stimulus order were randomized.

Anatomical MRI. T1-weighted anatomical volumes of the entire brain were collected for each participant at a resolution of 1 mm isotropic at the end of each functional scanning session. Anatomical scans were acquired with a T1-weighted BRAVO pulse sequence with the following parameters: inversion time, 450 ms; flip angle, 12°; field of view, 240 mm. Anatomical volumes were processed using FreeSurfer[89] (https://surfer.nmr.mgh.harvard.edu), version 6.0. Volumes were segmented to grey and white matter to produce a reconstruction of the cortical surface as well as a definition of the cortical ribbon (grey matter). Functional data were restricted to the cortical ribbon.

fMRI acquisition. T2*-weighted data were collected on a 3 T GE Discovery scanner using a gradient echo simultaneous multi-slice acquisition protocol for three simultaneous slice readouts with CAIPI-z phase shifting to improve the signal-to-noise ratio of the reconstructed image[90]. Slices (16 prescribed, 48 in total after simultaneous multi-slice) were aligned parallel to the parieto-occipital sulcus in each participant to guarantee coverage of the occipital and ventral temporal lobes. Parameters were as follows: field of view (FOV), 192 mm; repeat time (TR), 2; echo time (TE), 0.03 s; voxel size, 2.4 mm isotropic, no inter-slice gap.

fMRI data processing. Functional data were motion-corrected within and between scans and aligned to an anatomical reference volume per the standard Vistasoft pipeline[78]; see https://github.com/vistalab/vistasoft. Localizer data were

unsmoothed and always analysed in native participant space (unless producing average cortical maps as described below). Functional data were corrected for within- and between-scan motion; all participants were well trained and moved less than a voxel through their scans. For each voxel, the time course was transformed from arbitrary scanner units to percent signal change by dividing every time point by the mean response across an experiment. Localizer data were fitted with a standard general linear model by convolving stimulus presentation times with a difference-of-Gaussians haemodynamic response function implemented in SPM (www.fil.ion.ucl.ac.uk/spm), version 8. The GLM (general linear model) was also used to perform contrast mapping between different stimulus conditions.

MVPA. We performed multivoxel pattern analyses (MVPA) on VTC voxel data. The VTC was defined in each participant's cortical surface using the following cortical folds: laterally, it was bounded by the OTS up to the beginning of the ITG; medially, by the CoS; anteriorly, by the anterior tip of the MFS; and posteriorly, by the posterior transverse CoS. The multi-voxel patterns (MVPs) in response to each category (betas resulting from the GLM) were represented as a vector, the values of which were z-scored by subtracting the voxel's mean response and dividing by $\sqrt{\text{residual GLM variance}/\text{d.f.}}$. We then computed all pairwise correlations between MVPs of one category to every other to produce the 8×8 RSMs in Fig. 2a. Each cell is the average correlation from split-half combinations of the six runs (for example, $1-2-3/4-5-6$, $2-3-4 \times /1-5-6 \ldots$). MVPAs were run on in-plane data using the original fMRI data in the acquired resolution.

Classification. To quantify the information present in MVPs of the VTC, we constructed a winner-takes-all classifier. The classifier was trained on one half of the data and tested on how well it could predict the held-out half of data. For a given split, we used one half of the data as a training set and the other half as the testing set. The winner-takes-all classifier computes the correlation between the MVP of given test data (MVP for an unknown stimulus) and each of the MVPs of the labelled training data. It classifies the test data on the basis of the training data that yielded the highest correlation with the test data. For a given test MVP, correct classification yields a score of one and an incorrect choice yields a score of zero. Performance was computed for each category for both permutations of training and testing sets, yielding possible scores of 1 (both classifications were correct), 0.5 (only 1 was correct) or 0 (both incorrect). For each category, we averaged classification scores from all split-half combinations (ten) per participant and then averaged classification performance across participants within a group to produce the values in Fig. 2b,c.

Measurement of mean ROI response amplitudes. To evaluate the fMRI response from the Pokémon-selective cortex in experienced participants, independently from the data used to define the ROI, we used FreeSurfer to produce an average Pokémon-selective ROI that was independent of each participant's individual data. We implemented a leave-one-out approach, defining for the nth participant an average Pokémon ROI from $n-1$ participants. In 10 experienced participants, we transformed voxels selective for Pokémon (Pokémon versus all other stimuli, $t > 3$, voxel level) using CBA to the FreeSurfer average cortical surface (an independent average of 39 individuals). We then generated a group average probability map of the location of Pokémon on the FreeSurfer average brain and thresholded maps from these 10 experienced participants to include only vertices that were consistent across at least 30% of participants. This threshold was used in previous research and demonstrates the optimal point in Dice coefficients for predicting the ROI location in human cortices[78,91]. Furthermore, the thresholding ensures that no one participant can influence the group average ROI. This thresholded group ROI was then transformed back into the left-out participant, from which we then extracted the percent signal change reported in Fig. 8a. This procedure was repeated 11 times, for each of the experienced participants. In a similar manner, we created a group Pokémon ROI from all 11 experienced participants, thresholded at the same 30% overlap level, on the FreeSurfer average brain. This group ROI was transformed using CBA into each of the novice participants' brains, from which we extracted the percent signal change to our stimuli. Because CBA is not as accurate as defining fROIs from data in each participant's brain, we excluded voxels that showed significant selectivity to another category

Definition of pFus- and mFus-faces. Face-selective cortex was defined in each participant's brain as in previous work using cortical folds in the VTC as anatomical landmarks. mFus-faces was a cluster of face-selective (T values greater than 3 for the contrast of faces versus all other stimuli) voxels located on the lateral FG, aligned to the anterior end of the MFS. pFus-faces was defined with the same contrast and was located on the lateral FG 1–1.5 cm posterior to mFus-faces.

Evaluating the anatomical peak location of face, Pokémon and corridor selectivity in the VTC. The goal of this analysis was to determine the location of peak responses to a given category on a linear (lateral–medial) space to test the theoretical predictions illustrated in Fig. 4. Thus, in each experienced participant, we defined four anatomical bins in the VTC arranged from lateral to medial, whose posterior and anterior extent were constrained by the posterior transverse CoS and the

anterior tip of the MFS, respectively. The lateral and medial edges of each bin are as follows: the lateral-most bin extended from the ITG through the OTS and ended at the FG; the next bin included the lateral FG and ended at the pit of the MFS; the next bin extended from the MFS and included the medial FG, ending at the CoS; the most medial bin included the CoS and ended at the parahippocampal gyrus. Example bins are shown in the inset of Fig. 5b. From each bin, and within each individual participant, we extracted the mean selectivity (T-values) for a given contrast (faces, Pokémon or corridors). For each category we then fitted spline curves across these four values to produce the curves shown in Fig. 5b. Spline curves were fitted to allow for smoother spatial information and also aided in more accurately identifying peak locations. From these curves, we identified the maxima for each participant and compared the lateral–medial coordinate values across categories.

Image statistics analyses. To generate predictions for the locus of Pokémon-selective voxels in the VTC compared to face- and place-selective voxels, we ran a number of analyses on images of Pokémon, faces and corridors. We evaluated two physical properties of the stimuli, eccentricity of the retinal image and curvilinearity of the images, and two perceived properties, physical size and animacy of the stimuli. Corridors were chosen as they represent a category of scene stimuli that all participants probably have equal visual experience of, and because they are matched to Pokémon stimuli in that they possess linear features. This provides a larger dynamic range between faces and scenes to more readily assess where Pokémon exist in visual feature space between these stimuli classes.

Eccentricity bias. To evaluate the average eccentricity of the retinal image, we simulated retinal images (144 per category) in a representation of the visual field spanning 150×150 DVA (75° in one hemifield) and used a distribution of probable stimuli viewing distances to encapsulate the variability of viewing distances in real life and individual differences in viewing behaviour. For corridors and rooms, we assumed that the image nearly always occupied the entire visual field. That is, corridor/room stimuli were simulated to occupy anywhere from 95 to 100% of the visual field by assigning a random value of 95–100% of the size of the 150° simulated visual field (75° radius) to produce a distribution for the retinal image size of places. For stimuli of faces and Pokémon, we simulated Gaussian distributions of retinal image sizes. For Pokémon, viewing distance is relatively consistent, as they are viewed on the GameBoy screen (characters are presented in a 2 cm square and on average occupy 1.6 cm on the GameBoy screen) and limited to a maximum of arm's length. We assumed a relaxed position in which a participant would, on average, hold the device 32 cm (ref. [92]) from the eyes with a standard deviation of 10 cm. Pokémon were also viewed in childhood in the cartoon show, which aired on television once a week for a duration of 12–14 min. Thus, we estimate that much less time was spent watching Pokémon on television than playing the GameBoy game. B. Lechner of RCA laboratories measured average television viewing distance at which individuals watched television (national average: 9 feet, or 2.74 m) and produced optimal viewing distances for television screens of a given resolution. Given Lechner's distance for average television viewing and the size of televisions in the late 1990s and early 2000s, the retinal image produced by Pokémon from the cartoon show is comparable to the closely held GameBoy. For faces, which have an average size of 20 cm and an average viewing distance of 150 cm (about 5–6 feet)[93], we used a standard deviation of 50 cm in viewing distance. Although individuals may fixate on a face that is further away or at very close distances, these viewing times are probably much shorter than the conversational viewing distance of a face, which probably comprises the majority of face-viewing time. The average size of a face was taken to be 20 cm. Each stimulus of each category was then simulated as a retinal image from which we calculated its eccentricity in degrees of visual angle, assuming that participants fixated on the relevant item randomly sampled from the distributions described above. The distributions of these eccentricity values are displayed in Fig. 4a.

Rectilinearity. The lines composing a given image can be quantified in terms of their rectilinearity, that is, how straight (linear) or rounded (high curvature) the composite features of an image are. We employed the Rectilinearity Toolbox[25], which convolves a series of wavelet filters varying in orientation (22.5–360° in 22.5° steps), position, angle (30°, 60°, 90°, 120°, 150° and 180°) and scale (1/5, 1/9, 1/15 and 1/27 cycles per pixel) to assess the rectilinearity of our images. All Pokémon, face and corridor stimuli were analysed using this toolbox, whereby each image is given a score relative to other stimuli. The output of this toolbox is a ranking of all provided stimuli from least to most linear. Histograms based on the distribution of stimulus categories in this linear ranking are plotted in Fig. 4b.

Perceived attributes. *Size.* Faces and places have distinct and measurable sizes in the environment and producing distributions around the mean size of a face or place is relatively straightforward. For places, we used corridors in our simulations as their size, compared to something open air such as a forest or beach, is readily estimable. Given that open air places are probably larger than the numbers simulated here, these simulations thus represent a conservative estimate for the size of place stimuli. For faces and places, we simulated distributions of their real-world sizes and for Pokémon, which have no real-world size, raters

report their perceived size and compare these values to those of faces and places. Assuming an average anatomical head height of 20 cm, we made a Gaussian distribution with a standard deviation of 2 cm around that point. The distribution sample size was 150. The average wall height in a standard storied building is 2.4 m, but can be up to > 3 m. Corridors and building interiors probably make up a large percentage of place interactions for our participants, however, we also account for some larger scenarios that might be encountered in larger buildings or skyscrapers. We thus use a skewed distribution with a tail towards larger sizes; no value for corridor size was allowed to go below 2.43 m. The distribution sample size was 150. Pokémon, which are modelled to have animal-like properties in the game, have explicit sizes in the game itself. For example, a given Pokémon is accompanied by information about its physical features, including its size. While viewers experience the Pokémon on a small screen (similar to viewing images of faces and corridors on a small screen), there is an inherent understanding of what its physical size would be in the real world. To evaluate the physical size of Pokémon we took two approaches. First, we used the actual in-game sizes as the distribution of physical sizes. Second, we had an independent group of participants ($n = 28$) estimate the size of ($n = 50$) Pokémon randomly selected from the 144 stimuli using a 1−7 Likert scale. Each integer corresponded to a reference animal of increasing size (ant, mouse, cat, dog, gorilla, horse, dinosaur). Raters' choices were converted from a Likert digit (2, corresponding to a mouse) to the average metric size of that animal (mouse = 7.5 cm). We found that independent raters perceived Pokémon to be of sizes strikingly similar to those within the game. Each rating (Likert integer) was converted to metres using the average size of the reference animal and a mean size was calculated for each participant. Participants' ratings were then compared to the distribution of face sizes ($n = 150$) and corridor sizes ($n = 150$) from above.

Animacy. To quantify animacy for our images, another independent group of raters ($n = 42$) evaluated the perceived animacy of Pokémon, faces and corridors on a scale of 1 to 5, with increasing values corresponding to higher animacy. Participants were shown one image at a time and asked, for each image: 'How animate, or living, do you perceive this image to be?' Participants rated 50 stimuli each of faces, Pokémon and corridors; stimuli from each category were randomly selected from the total 144 stimuli used in the localizer experiment for each category. For each participant, their mean rating of animacy was calculated by averaging their scores for faces, Pokémon or words. Distributions were then made over the raters' mean values for each stimulus category.

Rectilinearity. To quantify how individuals may perceive different stimuli as more or less rectilinear, we conducted another behavioural experiment in an independent group of raters ($n = 50$). In the experiment, they viewed images of faces, Pokémon characters and corridors and evaluated on a scale from 1 (very curvy) to 7 (very linear/boxy) how linear they perceived the stimuli to be. The experiment was self-paced and participants were instructed to examine the lines, edges and shapes that made up a particular image and evaluate overall how linear they perceived the particular object to be. Participants saw 40 images from each category. From each participant we derived a mean rating for each category; violin plots depicting the distribution across participants for the ratings of faces, Pokémon and corridors are plotted in Supplementary Fig. 4.

Statistics. For all ANOVAs, grouping factors are described throughout the test. For statistical comparisons, data were assumed to be normal in distribution, but this was not formally tested. Individual participant data and distributions are represented in all relevant figures. The number of participants chosen was similar to that in other studies comparing groups of developmentally unique individuals in high-level visual cortex[52,94–96]. All *t*-tests performed were two tailed. Data collection and analysis were not performed blind to the conditions of the experiments. No participants were excluded for data quality issues. Two novice participants were unable to perform the behaviour experiment in Fig. 1 and were thus not included in that behavioural comparison between groups.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from the corresponding author on request. There are no restrictions on the sharing of the data, apart from allowing sufficient time to curate and send them on request.

Code availability
Code used to preprocess and analyse MRI data in this experiment can be found at Vistasoft (https://github.com/vistalab/vistasoft). Remaining code used to further process the data in this experiment can be found at https://www.gomezneuro.com/code.

References

1. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J. et al.) 549–586 (MIT Press, 1982).
2. DiCarlo, J. J. & Cox, D. D. Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007).
3. Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062 (1984).
4. Logothetis, N. K., Pauls, J. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563 (1995).
5. Hanson, S. J., Matsuka, T. & Haxby, J. V. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *NeuroImage* **23**, 156–166 (2004).
6. Grill-Spector, K., Weiner, K. S., Kay, K. N. & Gomez, J. The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.* **3**, 167–196 (2017).
7. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
8. Aguirre, G. K., Zarahn, E. & D'Esposito, M. An area within human ventral cortex sensitive to 'building' stimuli: evidence and implications. *Neuron* **21**, 373–383 (1998).
9. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
10. Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K. & Wandell, B. A. Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cereb. Cortex* **17**, 1604–1611 (2007).
11. McCandliss, B. D., Cohen, L. & Dehaene, S. The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* **7**, 293–299 (2003).
12. 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).
13. Parvizi, J. et al. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14915–14920 (2012).
14. Megevand, P. et al. Seeing scenes: topographic visual hallucinations evoked by direct electrical stimulation of the parahippocampal place area. *J. Neurosci.* **34**, 5399–5405 (2014).
15. Hirshorn, E. A. et al. Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl Acad. Sci. USA* **113**, 8162–8167 (2016).
16. Grill-Spector, K., Weiner, K. S., Kay, K. N. & Gomez, J. The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci* **3**, 167–196 (2016).
17. Weiner, K. S. & Grill-Spector, K. Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychol. Res.* **77**, 74–97 (2013).
18. Weiner, K. S. et al. Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *NeuroImage* **170**, 373–384 (2017).
19. Malach, R., Levy, I. & Hasson, U. The topography of high-order human object areas. *Trends Cogn. Sci.* **6**, 176–184 (2002).
20. Golarai, G., Liberman, A. & Grill-Spector, K. Experience shapes the development of neural substrates of face processing in human ventral temporal cortex. *Cereb. Cortex* **27**, bhv314 (2015).
21. Kanwisher, N. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl Acad. Sci. USA* **107**, 11163–11170 (2010).
22. Haxby, J. V. et al. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
23. Kriegeskorte, N. et al. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).
24. Hasson, U., Levy, I., Behrmann, M., Hendler, T. & Malach, R. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* **34**, 479–490 (2002).
25. Nasr, S., Echavarria, C. E. & Tootell, R. B. Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *J. Neurosci.* **34**, 6721–6735 (2014).
26. Konkle, T. & Caramazza, A. Tripartite organization of the ventral stream by animacy and object size. *J. Neurosci.* **33**, 10235–10242 (2013).
27. de Heering, A. & Maurer, D. Face memory deficits in patients deprived of early visual input by bilateral congenital cataracts. *Dev. Psychobiol.* **56**, 96–108 (2014).
28. Gandhi, T. K., Singh, A. K., Swami, P., Ganesh, S. & Sinha, P. Emergence of categorical face perception after extended early-onset blindness. *Proc. Natl Acad. Sci. USA* **114**, 6139–6143 (2017).
29. McKyton, A., Ben-Zion, I., Doron, R. & Zohary, E. The limits of shape recognition following late emergence from blindness. *Curr. Biol.* **25**, 2373–2378 (2015).

30. Dehaene, S. et al. How learning to read changes the cortical networks for vision and language. *Science* **330**, 1359–1364 (2010).
31. Gomez, J., Natu, V., Jeska, B., Barnett, M. & Grill-Spector, K. Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* **9**, 788 (2018).
32. 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).
33. Srihasam, K., Vincent, J. L. & Livingstone, M. S. Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* **17**, 1776–1783 (2014).
34. Sha, L. et al. The animacy continuum in the human ventral vision pathway. *J. Cogn. Neurosci.* **27**, 665–678 (2015).
35. Wiggett, A. J., Pritchard, I. C. & Downing, P. E. Animate and inanimate objects in human visual cortex: evidence for task-independent category effects. *Neuropsychologia* **47**, 3111–3117 (2009).
36. Warrington, E. K. & Shallice, T. Category specific semantic impairments. *Brain* **107**, 829–854 (1984).
37. Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. Neural correlates of category-specific knowledge. *Nature* **379**, 649–652 (1996).
38. Konkle, T. & Oliva, A. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* **74**, 1114–1124 (2012).
39. Arcaro, M. J. & Livingstone, M. S. A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife* **6**, e26196 (2017).
40. Huberman, A. D., Feller, M. B. & Chapman, B. Mechanisms underlying development of visual maps and receptive fields. *Annu. Rev. Neurosci.* **31**, 479–509 (2008).
41. Osher, D. E. et al. Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cereb. Cortex* **26**, 1668–1683 (2016).
42. Shatz, C. J. Emergence of order in visual system development. *J. Physiol.* **90**, 141–150 (1996).
43. Srihasam, K., Mandeville, J. B., Morocz, I. A., Sullivan, K. J. & Livingstone, M. S. Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron* **73**, 608–619 (2012).
44. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nat. Rev. Neurosci.* **6**, 877–888 (2005).
45. Wiesel, T. N. & Hubel, D. H. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* **26**, 1003–1017 (1963).
46. Hubel, D. H. & Wiesel, T. N. The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J. Physiol.* **206**, 419–436 (1970).
47. Shatz, C. J. Impulse activity and the patterning of connections during CNS development. *Neuron* **5**, 745–756 (1990).
48. Espinosa, J. S. & Stryker, M. P. Development and plasticity of the primary visual cortex. *Neuron* **75**, 230–249 (2012).
49. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
50. Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* **3**, 191–197 (2000).
51. James, T. W. & James, K. H. Expert individuation of objects increases activation in the fusiform face area of children. *NeuroImage* **67**, 182–192 (2013).
52. Nordt, M., Gomez, J., Natu, V. S. & Jeska, B. Learning to read increases the informativeness of distributed ventral temporal responses. Preprint at *bioRxiv* https://doi.org/10.1101/257055 (2018).
53. Harel, A., Gilaie-Dotan, S., Malach, R. & Bentin, S. Top-down engagement modulates the neural expressions of visual expertise. *Cereb. Cortex* **20**, 2304–2318 (2010).
54. McGugin, R. W., Newton, A. T., Gore, J. C. & Gauthier, I. Robust expertise effects in right FFA. *Neuropsychologia* **63**, 135–144 (2014).
55. Levy, I., Hasson, U., Avidan, G., Hendler, T. & Malach, R. Center-periphery organization of human object areas. *Nat. Neurosci.* **4**, 533–539 (2001).
56. Connolly, A. C. et al. The representation of biological classes in the human brain. *J. Neurosci.* **32**, 2608–2618 (2012).
57. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *NeuroImage* **9**, 195–207 (1999).
58. Wandell, B. A. & Winawer, J. Computational neuroimaging and population receptive fields. *Trends Cogn. Sci.* **19**, 349–357 (2015).
59. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *NeuroImage* **39**, 647–660 (2008).
60. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification* 2nd edn (Wiley, 2001).
61. Weiner, K. S. & Grill-Spector, K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage* **52**, 1559–1573 (2010).
62. McGugin, R. W., Gatenby, J. C., Gore, J. C. & Gauthier, I. High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proc. Natl Acad. Sci. USA* **109**, 17063–17068 (2012).
63. Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J. & Kanwisher, N. G. Discrimination training alters object representations in human extrastriate cortex. *J. Neurosci.* **26**, 13025–13036 (2006).
64. Martens, F., Bulthé, J., van Vliet, C. & Op de Beeck, H. Domain-general and domain-specific neural changes underlying visual expertise. *NeuroImage* **169**, 80–93 (2018).
65. Jiang, X. et al. Categorization training results in shape- and category-selective human neural plasticity. *Neuron* **53**, 891–903 (2007).
66. Clarke, A., Pell, P. J., Ranganath, C. & Tyler, L. K. Learning warps object representations in the ventral temporal cortex. *J. Cogn. Neurosci.* **28**, 1010–1023 (2016).
67. Natu, V. S. et al. Development of neural sensitivity to face identity correlates with perceptual discriminability. *J. Neurosci.* **36**, 10893–10907 (2016).
68. Jiang, X., Chevillet, M. A., Rauschecker, J. P. & Riesenhuber, M. Training humans to categorize monkey calls: auditory feature- and category-selective neural tuning changes. *Neuron* **98**, 405–416 (2018).
69. Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K. & Wandell, B. A. The development of cortical sensitivity to visual word forms. *J. Cogn. Neurosci.* **23**, 2387–2399 (2011).
70. 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 Biol.* **16**, e2004103 (2018).
71. Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.* **2**, 568–573 (1999).
72. Op de Beeck, H. P., Deutsch, J. A., Vanduffel, W., Kanwisher, N. G. & DiCarlo, J. J. A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cereb. Cortex* **18**, 1676–1694 (2008).
73. Kobatake, E., Wang, G. & Tanaka, K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* **80**, 324–330 (1998).
74. Golarai, G. et al. Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.* **10**, 512–522 (2007).
75. Scherf, K. S., Behrmann, M., Humphreys, K. & Luna, B. Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev. Sci.* **10**, 15–30 (2007).
76. Peelen, M. V., Glaser, B., Vuilleumier, P. & Eliez, S. Differential development of selectivity for faces and bodies in the fusiform gyrus. *Dev. Sci.* **12**, 16–25 (2009).
77. Cantlon, J. F., Pinel, P., Dehaene, S. & Pelphrey, K. A. Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cereb. Cortex* **21**, 191–199 (2011).
78. Gomez, J. et al. Microstructural proliferation in human cortex is coupled with the development of face processing. *Science* **355**, 68–71 (2017).
79. Weiner, K. S. & Zilles, K. The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia* **83**, 48–62 (2016).
80. Hasson, U., Harel, M., Levy, I. & Malach, R. Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* **37**, 1027–1041 (2003).
81. Lewis, T. L. & Maurer, D. Multiple sensitive periods in human visual development: evidence from visually deprived children. *Dev. Psychobiol.* **46**, 163–183 (2005).
82. Blais, C., Jack, R. E., Scheepers, C., Fiset, D. & Caldara, R. Culture shapes how we look at faces. *PLoS One* **3**, e3022 (2008).
83. Biscaldi, M., Fischer, B. & Aiple, F. Saccadic eye movements of dyslexic and normal reading children. *Perception* **23**, 45–64 (1994).
84. Olulade, O. A., Napoliello, E. M. & Eden, G. F. Abnormal visual motion processing is not a cause of dyslexia. *Neuron* **79**, 180–190 (2013).
85. Dalton, K. M. et al. Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* **8**, 519–526 (2005).
86. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal processing capacity in high-level visual cortex is domain specific. *J. Neurosci.* **35**, 12412–12424 (2015).
87. Willenbockel, V. et al. Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).
88. Kay, K. N., Winawer, J., Mezer, A. & Wandell, B. A. Compressive spatial summation in human visual cortex. *J. Neurophysiol.* **110**, 481–494 (2013).
89. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9**, 179–194 (1999).
90. Feinberg, D. A. & Setsompop, K. Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. *J. Magn. Reson.* **229**, 90–100 (2013).
91. Rosenke, M. et al. A cross-validated cytoarchitectonic atlas of the human ventral visual stream. *NeuroImage* **170**, 257–270 (2018).

92. Bababekova, Y., Rosenfield, M., Hue, J. E. & Huang, R. R. Font size and viewing distance of handheld smart phones. *Optom. Vis. Sci.* **88**, 795–797 (2011).

93. McKone, E. Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Res.* **49**, 268–283 (2009).

94. Weiner, K. S. et al. The face-processing network is resilient to focal resection of human visual cortex. *J. Neurosci.* **36**, 8425–8440 (2016).

95. Gomez, J. et al. Functionally defined white matter reveals segregated pathways in human ventral temporal cortex associated with category-specific processing. *Neuron* **85**, 216–227 (2015).

96. Furl, N., Garrido, L., Dolan, R. J., Driver, J. & Duchaine, B. Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *J. Cogn. Neurosci.* **23**, 1723–1740 (2011).

Author contributions

J.G. and M.B. designed and conducted the study. J.G. analysed the data. K.G.-S. oversaw the study and data analyses. J.G. and K.G.-S. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0592-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Jesse Gomez (jessegomez@berkeley.edu)

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Data was collected using MATLAB (R2015b, https://www.mathworks.com/products/matlab.html), and stimuli were presented with Psychtoolbox (http://psychtoolbox.org).
Data analysis	MRI data was analyzed with the following open source software packages: Vistaosft (https://github.com/vistalab/vistasoft), FreeSurfer (https://surfer.nmr.mgh.harvard.edu), and SPM8 (https://www.fil.ion.ucl.ac.uk/spm/software/spm8/)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Jesse Gomez: jessegomez@berkeley.edu

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was based off previous, single-subject fMRI paradigms in which multiple scans are collected on single subjects to ensure data reproducibility. For this reason, twelve subjects in each group, all of whom show a consistent effect, without data smoothing or group averaging, leads to findings we believe will reproduce more effectively at different institutions.
Data exclusions	No data was excluded.
Replication	Each participant completed six runs of the localizer experiment, and analyses were designed to take subsets of the data for training and other subsets of the data for testing. This cross-validation approach ensures the descriptive model is capturing reproducible signal within and across subjects.
Randomization	Randomization is not applicable to this study; groups were chosen specifically for their visual experience in childhood (or lack thereof). Subjects were matched for education and age.
Blinding	Blinding was not possible. However, all data was analyzed through a single pipeline to ensure minimization of influence from the researcher.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	Unique biological materials
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology
☒	Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Human participants were between the ages of 18 and 44, with a mean and standard deviation of 26.8 ± 4.8 years. Participants were split into two groups: experienced (n=11; age 24.3 ± 3.8) and novice participants (n=11; age 29.6 ± 5.4). The former group was selected initially through self-report, with the inclusion criteria that (1) participants began playing the original Nintendo Pokémon games between the ages of 5 and 8 on the handheld GameBoy device, (2) continued to heavily play the game and its series throughout their childhood, and (3) either continued to play the game into adulthood or revisited playing the game at least once in adulthood. Novice participants were chosen as individuals who never played the Pokémon game and had little to no interaction with Pokémon otherwise. All participants completed a 5-choice naming task designed to test the naming ability of participants; experienced participants scored significantly better than novices (Fig 1A). All participants provided written consent to participate in the experiment per Stanford University IRB.
Recruitment	See above

Magnetic resonance imaging

Experimental design

Design type	Block design

Design specifications	All participants underwent functional magnetic resonance imaging (fMRI), completing six runs of the experiment, with different images across runs. Each run was 218 seconds in duration, and across the six runs, counterbalanced 4s blocks presented 8 stimuli at 2Hz. Categories included faces, headless bodies, 8-bit Pokémon sprites from the original Nintendo game, animals, popular cartoon characters from television of the late 1990s and early 2000s, pseudowords, cars, and corridors. Stimuli, as illustrated in Fig 1B, were presented on a

textured background produced by phase scrambling a stimulus from another category80. Stimuli were presented approximately in the center of the background square, but were jittered in size and position to minimize differences in the average image between categories. Stimuli were run through the SHINE toolbox81 to remove luminance differences between categories or potential outlier stimuli. |
| Behavioral performance measures | All participants completed a 5-choice naming task designed to test the naming ability of participants; experienced participants scored significantly better than novices (Fig 1A). |

Acquisition

Imaging type(s)	Functional, structural
Field strength	3 Tesla
Sequence & imaging parameters	Anatomical MRI: T1-weighted anatomical volumes of the entire brain were collected for each subject at a resolution of 1mm isotropic at the end of each functional scanning session. Anatomical scans were acquired with a T1-weighted BRAVO pulse sequence with the following parameters: inversion time = 450ms, flip angle of 12o, field of view = 240mm. Anatomical volumes were processed using FreeSurfer83, https://surfer.nmr.mgh.harvard.edu. Volumes were segmented to gray and white matter to produce a reconstruction of the cortical surface as well as a definition the cortical ribbon (gray matter). Functional data were restricted to the cortical ribbon.
fMRI acquisition: T2*-weighted data were collected on a 3 Tesla GE Discovery scanner using a gradient echo simultaneous multi-slice (SMS) acquisition protocol for 3x simultaneous slice readout with CAIPI-z phase shifting to improve SNR of the reconstructed image84. Slices (16 prescribed, 48 in total after SMS) were aligned parallel to the parieto-occipital sulcus (POS) in each subject to guarantee coverage of the occipital and ventral temporal lobes. Parameters were as follows: FOV = 192mm, TR= 2, TE = 0.03s, voxel size 2.4mm isotropic, no interslice gap.
Area of acquisition	Occipital and temporal lobes
Diffusion MRI ☐ Used ☒ Not used	

Preprocessing

Preprocessing software	Vistasoft, which uses FreeSurfer and SPM8 functions. See above for links to most recent software versions.
Normalization	Normalization (z-scoring) was used to in decoding analysis to minimize effects driven by absolute magnitude and capture patterns across voxels, which were central to hypotheses of the manuscript. For the analysis of finding were Pokemon selectivity overlapped across subjects, data from participants were aligned to the FreeSurfer cortical average. Otherwise data was always analyzed in single-subject space.
Normalization template	FreeSurfer cortical average, using cortex-based aligned.
Noise and artifact removal	Functional data were motion-corrected within and between scans and aligned to an anatomical reference volume per the standard Vistasoft pipeline72, see https://github.com/vistalab/vistasoft. Localizer data were unsmoothed and always analyzed in native subject space
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	Random-effects analyses were used to compare group differences.
Effect(s) tested	Where appropriate, ANOVAs were always run before post-hoc t-tests.
Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both	
Anatomical location(s)	ROIs were defined either using thresholded parameter maps (T-values > 3), or were anatomically defined in the case of ventral temporal cortex defined with the following anatomical landmarks: laterally by the occipitotemporal sulcus, anteriorly by the anterior tip of the mid-Fusiform Sulcus, medially by the collateral sulcus, and posteriorly by the posterior transverse collateral sulcus.
Statistic type for inference (See Eklund et al. 2016)	For each voxel, the timecourse was transformed from arbitrary scanner units to percent-signal change (PSC) by dividing every timepoint by the mean response across an experiment. Localizer data were fit with a standard general linear model by convolving stimulus presentation times with a difference-of-gaussians hemodynamic response function implemented in SPM (www.fil.ion.ucl.ac.uk/spm). The GLM were also used to perform contrast maps between different stimulus conditions.
Correction	Voxelwise, uncorrected.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis