Seeing faces is necessary for face-patch formation

Michael J. Arcaro[*], **Peter F. Schade**[*], **Justin L. Vincent**, **Carlos R. Ponce**, and **Margaret S. Livingstone**[*]

Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, USA

Abstract

Here we report that monkeys raised without exposure to faces did not develop face patches, but did develop domains for other categories, and did show normal retinotopic organization, indicating that early face deprivation leads to a highly selective cortical processing deficit. Therefore experience must be necessary for the formation, or maintenance, of face domains. Gaze tracking revealed that control monkeys looked preferentially at faces, even at ages prior to the emergence of face patches, but face-deprived monkeys did not, indicating that face looking is not innate. A retinotopic organization is present throughout the visual system at birth, so selective early viewing behavior could bias category-specific visual responses towards particular retinotopic representations, thereby leading to domain formation in stereotyped locations in IT, without requiring category-specific templates or biases. Thus we propose that environmental importance influences viewing behavior, viewing behavior drives neuronal activity, and neuronal activity sculpts domain formation.

How do we develop the brain circuitry that enables us to recognize a face? The biological importance of faces for social primates and the stereotyped localization of face-selective domains in both humans and monkeys have engendered the idea that face domains are innate neural structures[1,2]. The view that face processing is carried out by an innate, specialized circuitry is supported by very early preferential looking at faces by human[3,4] and non-human primate infants[5] (see, however,[6,7]), and by the fact that faces, more so than other image categories, are recognized most proficiently as upright and intact wholes[8,9]. But genetic specification of something as particular as a face template, or even a response bias for such a high-level category, in inferotemporal cortex (IT) seems inconsistent with the widely accepted view that the visual system is wired up by activity-dependent self-organizing mechanisms[10]. Furthermore, the existence of IT domains for unnatural object categories,

Correspondence should be addressed to MSL at mlivingstone@hms.harvard.edu.
[*]Equal contribution

AUTHOR CONTRIBUTIONS

All authors scanned and reared monkeys; PS trained monkeys; ML, MA and PS analyzed data; ML wrote the paper.

COMPETINGNG FINANCNCIAL INTERESTS

The authors declare no competing financial interests.

DATA & CODE AVAILABILITY

Data will be made available upon request. We used publically available image analysis software and standard Matlab functions for analysis.

such as text[11] and buildings[12], and the induction by training of highly unnatural symbol domains in monkeys[13], all indicate a major effect of experience on IT domain formation. Yet, as pointed out by Malach and colleagues[14,15], both biological and unnatural domains invariably occur in stereotyped locations, and this consistency mandates some kind of proto-architecture. The mystery is what kind of innate organization could be so modifiable by experience as to permit the formation of domains selective for such unnatural categories as buildings[12], tools[16], chairs[17], or text[11], yet result in stereotyped neuroanatomical locations for these categories. Here we show that lack of early visual experience of faces leads to the absence of face-selective, but not other category-selective, domains. Further, we show that preferential face looking behavior precedes the emergence of face domains in monkeys raised with normal face experience. We recently reported that a strong retinotopic proto-organization is present throughout the visual system in newborn macaques, well before the emergence of category domains in IT[18]. Therefore the selectivities and stereotyped locations of category domains could be a consequence of viewing behavior that determines where activity-dependent changes occur on this retinotopic proto-map.

RESULTS

Face patches in control and face-deprived monkeys

To find out whether seeing faces is necessary for acquiring face domains, we raised 3 macaques without any visual experience of faces. These monkeys were hand raised by humans wearing welders' masks. We took extensive care to handle and play with them several times every day, and provided fur-covered surrogates that hung from the cage and moved in response to the animals' activity. We also provided them with colorful interactive toys and soft towels that were changed frequently, so these monkeys had an otherwise responsive and visually complex environment. These are the kinds of interventions that have been shown to prevent depression and support normal socialization[19]. They were kept in a curtained-off part of a larger monkey room so they could hear and smell other monkeys. They were psychologically assessed during the deprivation and were consistently lively, curious, exploratory, and interactive. They have all been returned to social housing post-deprivation, and have all integrated normally into juvenile social groups.

We scanned control and face-deprived monkeys while they viewed images of different categories (Supplementary Fig. 1). We recently reported that, in typically reared macaques, face patches (i.e. regions in the superior temporal sulcus (STS) selectively responsive to faces) are not present at birth, but instead emerge gradually over the first 5–6 months[20]. By 200 days of age, control monkeys all show robust, stereotyped faces>objects patches that are remarkably stable across sessions (ref.[20] and Fig. 1). The control monkeys consistently showed significant faces>objects activations in 4–5 patches distributed along the lower lip and fundus of the STS, as previously described for adult monkeys[21]. The face-deprived monkeys, in contrast, consistently showed weak, variable, or no significant faces>objects activations either in the expected locations along the STS, or anywhere else. The reproducibility of the face patches in the control monkeys across sessions, and the impressive difference in face activations between control and face-deprived monkeys, are clear from the individual session maps in Figure 1.

To quantify the consistency across scan sessions, we correlated the beta contrast maps for faces>objects within an anatomically defined ROI consisting of the STS and its lower lip (black dotted outline on B4 map in Fig. 1). The spatial pattern of (unthresholded) beta values within the STS ROI for faces>objects was highly reproducible across sessions for both control monkeys (mean r=0.86+/−0.02 sem) and face-deprived monkeys (mean r=0.66 +/−0.02 sem); see Supplementary Table 1 for individual monkeys. However, removing object-selective voxels from the analysis almost entirely eliminated any cross-session correlation in the deprived monkeys (mean r=0.02 +/−0.06 sem), suggesting that the consistency of the beta map pattern of faces>objects was driven by object-selective responsiveness, and therefore there was no consistency to the face-selective responsiveness in the face-deprived monkeys. In contrast, faces>objects correlations remained strong in the control monkeys even when the object-selective voxels were removed from the analysis (mean r=0.72+/−0.05 sem). This analysis further demonstrates that the spatial pattern of face-selective activity was consistent across sessions for the control monkeys, but inconsistent across sessions for face-deprived monkeys. Conjunction maps (Fig. 2) further demonstrated this reproducibility by illustrating spatially consistent activity across sessions for all monkeys that were scanned in multiple sessions. Together, these data demonstrate a striking difference between control and face-deprived monkeys specifically in face-selective responsiveness in IT.

Hand>object patches in control and face-deprived monkeys

The surprising result that face-deprived monkeys had weak or no face selectivity in the STS, or anywhere else, motivated us to examine other domains that normally occur in IT. In most monkeys, body domains are adjacent to, and partially overlapping with, face patches[21,22]. Reasoning that the most often seen, and behaviorally significant, body part for the hand-reared face-deprived monkeys would be their own hands, or the gloved hands of laboratory staff, we tested them with interleaved blocks of monkey hands or gloved human hands along with blocks of faces and blocks of objects. The face-deprived monkeys all showed strong hand activations in roughly the same fundal regions of the STS as did control monkeys (Figs. 1 & 2). The spatial pattern of (unthresholded) beta values within the STS ROI for hands>objects was reproducible across sessions for both control monkeys (mean r=0.49 +/− 0.07 sem) and face-deprived monkeys (mean r=0.76 +/−0.01 sem, respectively; see Supplementary Table 1 for individual monkeys). Further, cross-session correlations remained strong for both control and face-deprived monkeys even when object-selective voxels were removed from the analysis (mean r=0.26 +/−0.05 sem, and 0.65 +/−0.03 sem, respectively), demonstrating that the locations of hand-selective responses in the STS were stable across scans. In the face-deprived monkeys, the hands>objects patches were stronger, larger, and more anatomically consistent across sessions than the sparse faces>objects activations, but in control monkeys, the hand patches were smaller or equivalent in size to the face patches.

Hand>object vs face>object patches in control and face-deprived monkeys

The sizes of face and hand patches differed between control and face-deprived monkeys. Figure 2b quantifies these differences between control and face-deprived monkeys in the STS ROI (ROI is shown as a dotted outline in Fig. 2a B4 map). Figure 2b compares the number of 1mm^3 voxels in this ROI in each monkey that were significantly more activated

by faces>objects with the number of voxels more activated by hands>objects (voxels selected using p<0.01 FDR corrected threshold). The number of hand-selective voxels in the STS ROI was larger than the number of face-selective voxels for every scan session in each face-deprived monkey. This difference was significant for every face-deprived monkey (2-tailed, paired-sample t-test across sessions and across hemispheres; B6: $t(17)=-13.5$; $p=2\times10^{-10}$; B9: $t(17)=-8.2$; $p=3\times10^{-7}$; B10: $t(11)=-6.7$; $p=3\times10^{-5}$). The opposite held true for two of the control monkeys, in that there were significantly more face-selective voxels than hand-selective voxels (B4: $t(5)=3.2$; $p=0.02$; B8: $t(5)=6.4$; p-0.001) and in the third control monkey, B5, there was no significant difference between the number of face-selective voxels and the number of hand voxels ($t(5)=-1.0$; $p=0.4$). Although our n was too small for robust group level comparisons, fixed effects analyses converged with our within-subject analyses (Methods: Group-level, Fixed-effects Analyses). Further, the smaller number of face-selective voxels in the face-deprived monkeys cannot be attributed to an overall diminished responsiveness in the STS because the average number of hand-selective voxels was actually larger in the face-deprived monkeys compared to controls (311 vs 202 voxels on average). These analyses confirm the impression from the beta contrast maps in Figures 1&2 that control monkeys showed more face-selective activation in the STS than did face-deprived monkeys, and demonstrate a specific deficit of face-, not hand-, selective activity in the face-deprived monkeys.

Fixation during scanning

Can any artifact in our study design explain the selective deficit in face responsiveness in IT of the face-deprived monkeys? Analysis of responsiveness in V1 (Fig. 2c) and fixation behavior during scanning (Supplementary Fig. 2) indicate that there were no differences in fixation behavior during scanning that could explain the large differences between control and face-deprived monkeys in face-selective activations (see Methods: Fixation During Scanning).

Relative locations of face and hand activations

Because the face-deprived monkeys had slightly larger hand patches than did the control monkeys, we wondered whether this indicated take-over of face territory by hand responsiveness. By inspection of Figures 1&2, although the face and hand patches overlapped, the hand patches were centered slightly deeper in the sulcus compared to the face patches (Fig. 2d). To measure the locations of the face and hand activations along the depth of the STS, we drew two other anatomically-defined ROIs, one in central IT (CIT) and another in anterior IT (AIT). These ROIs spanned from the lower lip through the STS fundus to the upper bank, and thus encompassed (in control monkeys) the face patches on the lower lip of the STS (ML/AL) and in the fundus (MF/AF) in CIT and AIT (ROIs are shown by dotted lines in B5 face map in Fig. 2a). We drew these ROIs to maximally sample the face patches in control monkeys and hand patches in all monkeys in order to compare the relative mediolateral locations in control and deprived monkeys of face and hand responsiveness with respect to the anatomical landmarks, the banks and fundus of the STS. Beta coefficients for faces-minus-objects (solid lines) and hands-minus-objects (dotted lines) were averaged across the AP dimension of the CIT ROI (Fig. 2e) and the AIT ROI (Fig. 2f) for all the control (red) and all the face-deprived (green) monkeys. On average, the face-deprived

monkeys showed enhanced hands-minus-objects responses (cf. green & red dashed lines) in the fundus of the STS. Supplementary Figure 3 shows the raw percent signal changes to faces, hands, and objects for the same CIT and AIT ROIs, from which it can be determined that these regions in face-deprived monkeys do respond to faces, just less than to hands or to objects. Since the peak of the hand selectivity in the face-deprived monkeys was maximum in the depth of the STS, it seems likely that the hand activations in the fundus of the STS in the face-deprived monkeys represents hand responsiveness of what would normally be hand patches, rather than take-over by hands of what would normally be face territory. Nevertheless, our results are not inconsistent with the idea that hand and face selectivities compete for cortical territory.

Scene and body patches in control and face-deprived monkeys

Since the face-deprived monkeys' face-patch organization was grossly abnormal, and they showed strong hand selectivity, we wondered about other IT domains, and so we conducted scan sessions contrasting blocks of scenes and faceless bodies. As shown in Figure 3, the face-deprived monkeys showed scene and body patches in roughly the same locations as did control monkeys, in that all monkeys showed bodies>scenes activations lying along the STS, and two consistent regions of scenes>bodies activations, one in the occipitotemporal sulcus (OTS, outlined in the B4 map in Fig. 3), corresponding to the Lateral Place Patch (LPP)[23,24], a potential homologue of the human PPA[24], and one more dorsally, in dorsal prelunate cortex, a region that corresponds to the Dorsal Scene Patch[25], the potential homologue of the human scene area in the transverse occipital sulcus (TOS) (indicated by white arrowheads in the B5 map in Fig. 3). Since the STS and the OTS are clear anatomical regions (the dorsal scene patch did not have any clear anatomical ROI that did not also include early visual areas), we used two anatomically defined ROIs to compare the scene and body activations: one, the same STS ROI that we used to quantify the face vs hand activations, outlined on the B4 map in Figure 3, and the other, an OTS ROI, drawn to encompass the entire OTS and its ventral lip, also outlined on the same map. Figure 3 (bottom) compares the number of significant bodies>scenes voxels and the number of scenes>bodies voxels (voxels selected at p<0.01 FDR corrected threshold) in each ROI in each hemisphere in each monkey for all body vs scene scan sessions. Group level comparisons converged with these individual subject scenes>body contrast maps (see Methods: Group-level Fixed-effects Analyses). Thus both control and deprived monkeys showed a similar pattern of selectivity for bodies in the STS and for scenes in the OTS. These complementary scene and body activations, as well as the hand activations in the STS, in the deprived monkeys, indicated that their IT was organized into at least several non-face category-selective domains.

Eccentricity organization in control and face-deprived monkeys

The face-deprived monkeys not only had hand, body, and scene patches, but the retinotopic organization of their entire visual system was also normal, even in regions of IT that in control monkeys represent faces. Figure 4 (top) shows average eccentricity maps from phase-encoded retinotopic mapping, for 2 control monkeys (left) and 2 face-deprived monkeys (right). All the monkeys showed a normal adult-like eccentricity organization across the entire visual system, including IT, with a foveal bias lying along the lower lip of the STS, flanked dorsally and ventrally by more peripheral representations, consistent with

previous retinotopic mapping in adult macaques[26–28]. Both control and face-deprived monkeys also showed a separate eccentricity map in the posterior STS, corresponding to the MT cluster[26,28]. The lateral face patches ML and AL in the control monkeys (ventral black outlines) were located in the strip of foveal bias that lies along the lower lip of the STS[18], as previously reported for adult monkeys[27,29], whereas the fundal patches MF and AF (dorsal black outlines) lay in more mid-field biased regions. Both control and face-deprived monkeys showed some hand responsiveness (magenta outlines) in the foveally biased lip of the STS, but greater hand responsiveness in the depth of the sulcus, which is more mid visual field. In both control and face-deprived monkeys, scene selectivity (white outlines) mapped to peripherally biased regions of IT, also consistent with previous results in adult monkeys[29]. To illustrate the visual field biases in these category-selective regions, we plotted the distribution of eccentricity representations within face, hand, and scene ROIs determined by conjunction analysis (Fig. 4c). In control monkeys (Fig. 4c, left), face patches were foveally biased, hand patches covered both central and peripheral visual space, and the scene patches were peripherally biased. In face-deprived monkeys (Fig. 4c, right), the visual field coverage for hand and scene patches was similar to controls. Analysis of the eccentricity representation at the most selective voxel for each ROI in each monkey showed similar visual field biases as the distribution analysis (Figure 4c, dots at the top of graphs). Thus retinotopy, body patches, hand patches, and scene selectivity were not detectably abnormal in face-deprived monkeys, but face selectivity was significantly abnormal. Therefore face deprivation produced a dramatic and selective reduction in face selective clusters in IT.

Looking behavior in control and face-deprived monkeys

The lack of face patches in the face-deprived monkeys reveals a major role for visual experience in the formation, or maintenance, of category domains. Furthermore, the foveal bias of face patches and the peripheral bias of scene patches (Fig. 4), suggest that where in its visual field the animal experiences these categories, may be important in IT domain formation, as has been proposed for humans[14,30]. We therefore asked what young monkeys look at by monitoring their gaze direction when they freely viewed 266 images of natural scenes, lab scenes, monkeys, or familiar humans. In general all monkeys tended to look at image regions with high contrast, especially small dark regions and edges (e.g. Fig. 5a gaze patterns for the hammer and clothes pins), but, in addition, the control monkeys looked disproportionately towards faces, whereas the face-deprived monkeys did not look preferentially at faces, and instead looked more often at regions containing hands. Figure 5a illustrates the differential looking behavior between control and face-deprived monkeys for 9 of the images used in the viewing study. Since there were no systematic differences between control and face-deprived monkeys in where they looked in images that did not contain hands or faces (e.g. the hammer or the clothes pins), we narrowed our analysis to the 19 images in the viewing set that contained prominent faces and hands. Figure 5b quantifies the difference in viewing behavior between control and face-deprived monkeys by plotting the average looking time at regions containing faces or regions containing hands for these 19 images (see Fig. 5a for examples of regions used). This scatter plot shows the strong tendency of the control monkeys to look disproportionately at faces and of the face-deprived monkeys to look less often at faces and more at hands. Across viewing sessions, all 4 control

monkeys looked significantly more at faces than at hands, and all 3 face-deprived monkeys looked significantly more at hands than at faces (2-tailed t-test; controls: B3: t(20)=5.2; p=4×10^{-5}; B4: t(11)=10.3; p=5.5×10^{-7}; B5: t(13)=24; p=4×10^{-12}; B8: t(17)=9.6; p=3×10^{-8}; face-deprived: B6: t(10)= −8.4; p=7×10^{-6}; B9: t(21)= −6.5; p= 2×10^{-6}; B10: t(12)= −13.6; p=1×10^{-8}). Although our n was too small for robust group level comparisons, fixed effects analyses converged with our within-subject analyses (see Methods: Group-level Fixed-effects Analyses). The failure of the face-deprived monkeys to look selectively at faces means that faces must not be inherently interesting or visually salient, so we conclude that the face looking behavior of the control monkeys reflects learning from reinforcement they received during their development, rather than an innate tendency to look at faces. Plotting fixation times to face- or hand-containing areas as a function of age (Fig. 5 c&d) revealed that differential looking behavior, most notably face looking in control monkeys, emerged very early in development, prior to the establishment of category-selective domains, which become detectable by fMRI only by 150–200 days of age[20]. The preferential looking at hands in the 19 viewing images by the face-deprived monkeys need not imply that they normally fixated their hands all day, but rather that they recognized and attended to the hands in these images because they had already learned that hands are the most interesting, animate, or important things in their daily environment.

DISCUSSION

We asked whether face experience is necessary for the development of the face patch system. Monkeys raised without exposure to faces did not develop normal face patches, but they did develop domains for scenes, bodies, and hands, all of which they had had extensive experience with. Furthermore the large-scale retinotopic organization of the entire visual system, including IT, was normal in face-deprived monkeys. Thus, the neurological deficits in response to this abnormal early experience were specific to the never-seen category. These results highlight the malleability of IT cortex early in life, and the importance of early visual experience for normal visual cortical development.

A lack of face patches in face-deprived monkeys indicates that face experience is necessary for the formation (or maintenance) of face domains. Our previous study showing that symbol selective domains form in monkeys as a consequence of intensive early experience indicates that experience is sufficient for domain formation. That experience is both necessary and sufficient for domain formation is consistent with the hypothesis that a retinotopic proto-organization is modified by viewing behavior to become selectively responsive to different categories in different parts of the retinotopic map. Malach and colleagues proposed that an eccentricity organization could precede and drive, by different viewing tendencies for different image categories, the categorical organization of IT[14,15,30]. We recently found that a hierarchical, retinotopic organization is indeed present throughout the entire visual system, including IT, at birth and therefore could constitute a proto-organization for the development of experience-dependent specializations[18]. Here, we further asked whether early viewing behavior acting on a retinotopic proto-map could produce the stereotyped localization of category domains. We found that control monkeys looked disproportionately at faces, even very early in development, before the emergence of face patches[20] (Fig. 5c). A tendency to fixate on faces could contribute to the formation of

face domains in foveally biased parts of IT, without requiring any prior categorical bias. The failure to form face patches in the absence of face experience is also a strong prediction of this retinotopic proto-map model. Consistent with this hypothesis, cortical responses to face or body parts are selective for typically experienced configurations[31–33], faces are optimally recognized in an upright, intact configuration[8,9], and experience shapes face-selectivity during infancy and childhood[34–36], all of which corroborate a role for long-term natural experience in generating cortical domains.

A critic could argue that, alternatively, a lack of face patches in face-deprived monkeys could be explained by withering of an innate face domain. The literature in favor of such an innate categorical organization is extensive, but the most compelling evidence would be finding category-selective domains at birth, or a genetic marker for these domains, neither of which have been reported. Face vs. scene selectivity can be detected (in the same locations that are face selective in adults) in human infants by 5–6 months[37] and macaques as early as 1 month of age[20], but these early 'face' domains are not selective for faces per se, since they respond equally well to faces and objects. This early faces>scenes selectivity has been interpreted as evidence for immature innate domains[37] or as eccentricity-based low-level shape selectivity[20], but the absence of face patches in face-deprived monkeys favors the latter interpretation. Our results limit how strong an innate *category* bias can be; furthermore, the earliest organization cannot be *categorical* alone, given that a robust *retinotopic* organization is present in IT at birth in monkeys[18]. Moreover, an innate retinotopic organization carries with it an organization for receptive-field size, and therefore a bias for features such as curvature[13,38] and scale. Thus the *retinotopy* model could account for a very early selectivity for faces vs scenes, and it predicts a lack of face patches after face deprivation, whereas the *category* model is weakened by these findings, and by the finding of an innate retinotopic organization[18], to such an extent that the *category* model is essentially unfalsifiable.

The retinotopic model does require early category-selective looking behavior, but our finding that face-deprived monkeys do not preferentially look at faces as early as 90 days of age indicates that face looking is also unlikely to be innate, but rather must be learned. An innate face recognition system has been proposed to account for infants' tracking faces or face-like patterns[3], or imitating facial gestures[39,40]; however, other studies have claimed that infants' face looking is based on non-specific low-level features[7,41], and that newborns do not actually imitate facial gestures[42]. The failure of face-deprived monkeys to look preferentially at faces supports the view that face looking by infants is a consequence of low-level perceptual biases and behavioral reinforcement. We showed that there is an innate retinotopic organization[18], that will carry with it a bias for scale and curvature[13,38] in the parts of IT that will become face selective. That, plus a tendency to attend to movement or animacy[43], and the consistent reinforcing value of faces in the infant's environment, could all contribute to preferential face looking behavior, without requiring any innate face-specific mechanisms. Further, a prior study on face-deprived monkeys supports a learned, rather than innate, basis for face viewing behavior: Sugita's study on face-deprived monkeys[44] has been interpreted as showing an innate bias for looking at faces[2] and for face discrimination abilities[1,2,45]. However, although his face-deprived monkeys looked on average more often at faces than at objects, the difference was not significant for his

deprived animals (p 395 lines 29–30), despite being significant for control animals (p 395 line 26). Also the deprived monkeys' face discrimination was also not significant (p 395 lines 55–56), though control animals' face discrimination was (p 395 line 52). Therefore Sugita's study does not support an innate bias for looking at or discriminating faces, though he does show significant effects of experience on face looking and discriminating. A critic nonetheless argues that a lack of preferential face looking by face-deprived monkeys in both our study and Sugita's is consistent with "innate face looking behavior that requires experience of faces to be maintained". However, as for the innateness of face domains, a face looking predisposition that requires face experience is a weak innate predisposition, and difficult to distinguish from learned behavior.

We propose that IT domains develop as follows: *Very early in development, reinforcement affects the ways infants interact with their environment, including what they look at*[46]. *Looking behavior determines where, on an innate retinotopic proto-map*[18], *neuronal activity for different image categories predominates. Synapses are retained or strengthened based on neuronal activity*[47], *and local interconnectivity*[48] *and competition*[49] *favor clustered retention or strengthening of inputs responsive to similar, or co-occurring, features.* This model is parsimonious, and it would mean that the development of IT involves the same kind of activity-dependent, competitive, self-organizing rules that generate in V1 at birth a precise retinotopic map organized into orientation pinwheels[50].

Online Methods

Monkeys

Functional MRI and viewing behavior studies were carried out on 4 control and 3 face-deprived Macaca mulattas, all born in our laboratory, all with excellent health, normal growth, and normal refraction. Animals were housed under a 12-hour light/dark cycle. All procedures conformed to USDA and NIH guidelines and were approved by the Harvard Medical School Institutional Animal Care and Use Committee. The control monkeys (3 male 1 female) were socially housed in a room with other monkeys. The face-deprived monkeys (1 male 2 female) were hand-reared by laboratory staff wearing welders' masks that prevented the monkey from seeing the staff member's face. The only exposures they had to faces of any kind were during scan sessions and viewing sessions, which constituted at most 2 hours per week, with the face exposure being a minor fraction of that. The viewing sessions had some images that contained faces, but they were conducted only after age 90 days, and the scanning with some blocks of face images was started only after age 225 days for B6, 151 days for B9 and 173 days for B10. For scanning, the monkeys were alert, and their heads were immobilized non-invasively using a foam-padded helmet with a chin-strap/bite-bar that delivered juice, to reward them for looking at the screen. The monkeys were scanned in a primate chair that was modified to accommodate small monkeys in such a way that they were positioned semi-upright, or in a sphynx position. They were positioned so they could move their bodies and limbs freely, but their heads were restrained in a forward-looking position by the padded helmet. The monkeys were rewarded with juice for looking at the center of the screen. Gaze direction was monitored using an infrared eye tracker (ISCAN, Burlington, MA).

Stimuli

The visual stimuli were projected onto a screen at the end of the scanner bore, 50 cm from the monkey's eyes. All the images covered 20 degrees of visual field. Faces, hands, and objects were centered on a pink-noise background and subtended a similar maximum dimension of 10°. The face, hand, and body mosaics were on a pink-noise background with faces, hands, or bodies covering the entire 20 degrees, as did the scenes. All images were equated for spatial frequency and luminance using the SHINE toolbox [51]. In some sessions we showed both monkey hands and gloved human hands, and in 3 sessions we showed blocks of gloved hands instead of monkey hands; activations to gloved hand were almost identical to the monkey-hand activations (Fig. 1). Each scan comprised blocks of each image category; each image was presented for 0.5 seconds; block length was 20 seconds, with 20 seconds of a neutral gray screen with a fixation spot between image blocks. Blocks and images were presented in a counterbalanced order. In most of the early scan sessions for the control monkeys, hands were not included in the image sets because we had not yet discovered how strongly the face-deprived monkeys responded to hands. One of the control monkeys, B3, was scanned viewing blocks of hands only once, so this monkey's data are included in figures showing individual scan session data, but not in analyses of across-session statistics.

Scanning

Monkeys were scanned in a 3-T Tim Trio scanner with an AC88 gradient insert using 4-channel surface coils (custom made by Azma Maryam at the Martinos Imaging Center). Each scan session consisted of 10 or more functional scans. We used a repetition time (TR) of 2 seconds, echo time (TE) of 13ms, flip angle (α) of 72°, iPAT = 2, 1mm isotropic voxels, matrix size 96×96mm, 67 contiguous sagittal slices. To enhance contrast[52,53] we injected 12 mg/kg monocrystalline iron oxide nanoparticles (MION; Feraheme, AMAG Pharmaceuticals, Cambridge, MA) in the saphenous vein just before scanning. MION increases the signal-to-noise and inverts the signal[52]; for the readers' convenience we show the negative signal change as positive in all our plots. By using this contrast agent we were able to reliably obtain clear visual response signals in V1 for single blocks, so we could use both eye position data combined with the V1 signal as a quality control, to eliminate blocks or entire scans when the monkey was inattentive or not looking at the stimulus[20] (see Data Analysis below).

Free-Viewing Behavior

Only a subset of 19 images that were used in a larger free-viewing study had clearly visible hands and faces and were used in this analysis. Because we wanted to avoid exposing the face-deprived monkeys to faces early in development we did not show them any of the face-containing images until >90 days. For each session, all of which occurred during the animals' normal lights on period, the images from the larger viewing set were presented in random order for 3 seconds with 200ms of black screen in between presentations; the image set was presented repeatedly until the monkey stopped looking at the screen. Images ranged from 18.9×18.9 to 37.8 × 37.8 degrees of visual angle. Monkeys were seated comfortably in a primate chair with their heads restrained using a padded helmet, as for scanning. Gaze

direction of the monkey's left eye was collected at a rate of 60Hz by an infrared eye tracker (ISCAN, Burlington, MA). The monkey's eye was 40cm from the screen and the eye-tracker camera. A five-point calibration was used at the start of each session. Gaze data were analyzed using Matlab. Data were smoothed, and fixations identified by an algorithm[54] that identifies epochs when eye velocity was less than 20°/s but rejects eye-closed epochs. Fixations shorter than 100ms or with standard deviations larger than 1° were rejected. Fixations that fell more than 1° outside the border of the presented image were not included in the analysis. For the heat maps, in addition, fixations longer than 2 standard deviations greater than the mean were not included. Heat maps were summed over multiple sessions for each image, smoothed by a 0.25° Gaussian, and normalized to the maximum.

Phase-encoded Retinotopic Mapping

Eccentricity and polar angle mapping were performed in monkeys B4, B5 and B6 > 1.5 years and in monkey B9 at 9 months of age. See Arcaro et al.[26] for details. To obtain eccentricity maps, the visual stimulus consisted of an annulus that either expanded or contracted around a central fixation point. The annulus consisted of a colored checkerboard with each check's chromaticity and luminance alternating at the flicker frequency of 4 Hz. The duty cycle of the annulus was 10%; that is, any given point on the screen was within the annulus for only 10% of the time. The annulus swept through the visual field linearly. Each run consisted of seven cycles of 40 s each with 10 s of blank, black backgrounds in between. Blank periods were inserted to temporally separate responses to the foveal and peripheral stimuli. 10–12 runs were collected with an equal split in direction. The eccentricity stimuli are more difficult for the monkeys to fixate than are the category images, and monkey B10 is not yet good enough at fixating to do eccentricity mapping. The polar angle stimulus consisted of a flickering colored checkerboard wedge that rotated either clockwise or counterclockwise around a central fixation point. The wedge spanned 0.5–10° in eccentricity with an arc length of 45° and moved at a rate of 9°/s. Each run consisted of eight cycles of 40 s each. 10–12 runs were collected with an equal split in the direction of rotation.

Fourier analysis was used to identify spatially selective voxels from the polar angle and eccentricity stimuli. For each voxel, the amplitude and phase (the temporal delay relative to the stimulus onset) of the harmonic at the stimulus frequency were determined by a Fourier transform of the mean time series. To correctly match the phase delay of the time series of each voxel to the phase of the wedge/ring stimuli, and thereby localize the region of the visual field to which the underlying neurons responded best, the response phases were corrected for a hemodynamic lag (4 s). The counterclockwise (expanding) runs were then reversed to match the clockwise (contracting) runs and averaged together for each voxel. An F-ratio was calculated by comparing the power of the complex signal at the stimulus frequency to the power of the noise (the power of the complex signal at all other frequencies). Statistical maps were threshold at a $p < 0.0001$ (uncorrected for multiple comparisons). Contiguous clusters of spatially selective voxels were identified throughout cortex in both polar angle and eccentricity mapping experiments. A series of visual field maps were identified in general accordance with previous literature[26,28,55,56]. These included areas V1, V2, V3, V4, V4A, MT, MST, FST, V4t, V3A, DP, CIP1, CIP2, LIP, OTd, PITd, and PITv. Borders between visual field maps were identified based on reversals in the

systematic representation of visual space, particularly with respect to polar angle. Eccentricity representations were evaluated to ensure that phase progressions were essentially orthogonal (nonparallel) to the polar angle phase progression and to differentiate between the MT+ cluster[28] and surrounding extrastriate visual field maps.

Fixation During Scanning

To evaluate whether there were differences in how the two monkey groups looked at the different stimuli, we measured the visual response to each image category for each monkey for each scan session in an ROI covering the central 6–7° of each hemisphere's V1. In 4 monkeys, B5 (control), B6 (deprived), B8 (control), and B10 (deprived), the response in V1 was larger to hands than to faces (2-tailed t-test; B5: $t(5)= -5.9$; $p=0.002$; B6: $t(17)= -7.8$; $p=0.0001$; B8: $t(5)= -3.1$; $p=0.03$; B10: $t(11)= -2.4$, $p=0.04$), probably because the hands were more retinotopically heterogeneous than the faces. In the other 2 monkeys, B4 (control) and B9 (deprived), there was no difference between the V1 response to faces and to hands (2-tailed t-test; B4: $t(5)= -2.0$, $p=0.1$; B9: $t(17)=0.3$, $p=0.74$). To look for group-level effects we performed an ANOVA on the means across sessions with monkey as the random effects factor (including the single session for B3). There were no main effects of stimulus (V1 response to faces vs hands) $F(1,21)=1.97$, $p=1.7$, monkey group $F(1,21)=1.43$; $p=0.24$, or hemisphere $F(1,21)=0.01$; $p=0.93$, and no interaction $F(1,21)<0.42$, $p>0.52$). The similarity between the control and the face-deprived monkeys in the V1 response magnitude to hands and faces indicates that there were no differences in fixation behavior during scanning that could explain the large differences between control and face-deprived monkeys in face-selective activations.

We included in our analyses only scans in which the monkeys looked at the screen (20° across) for more than 80% of the scan. Supplementary Figure 2 shows the percent of time for each monkey for each scan session for each image category separately that the monkeys fixated within 5° of the screen center, which was the area covered on average by the images. Although fixation performance was slightly worse for deprived monkeys in some scan sessions, there was no systematic difference between categories in percent time the monkeys' gaze was <5° from fixation (ANOVA; no main effect of monkey group; no main effect of image category and no interaction; $F(1,8)<1.2$; $p>0.3$). Furthermore, there was no difference between groups in fixation quality for faces vs other categories, and the data were sufficient to reveal stereotyped hand patches and object patches in every individual scan session (Fig. 1). Therefore, it is unlikely that differences in face-selective activity between deprived and controls can be attributed to viewing behavior during scanning. The only other potential artifact we could think of would be that the face-deprived monkeys might not have attended to the faces as well as the control monkeys did. Such effects of attention should be apparent in the magnitudes of the V1 response[57–59], though certainly attentional modulation could be larger in extrastriate areas[59]. However, there were no differences in V1 activity between groups, indicating that a relative lack of attention cannot explain the lack of face patches in face-deprived monkeys. Furthermore robust face selectivity has been reported repeatedly in both humans and in monkeys during passive viewing[60], and even in anesthetized monkeys[61], therefore a dramatic lack of face responsiveness in our face-deprived monkeys cannot be due entirely to differences between groups in attention to faces.

Data Analysis

Functional scan data were analyzed using AFNI[62] and Matlab (Mathworks, Natick MA). Each scan session for each monkey was analyzed separately. Potential "spike" artifacts were removed using AFNI's 3dDespike. All images from each scan session were then motion corrected and aligned to a single reference time-point for that session first by a 6-parameter linear registration in AFNI then a nonlinear registration using JIP (written by Joseph Mandeville, and which accounted for slight changes in brain shape that naturally emerge over the course of a scan session from changes in the distortion field). Scans with movements greater than 0.8 mm were not included in any further analysis. Data were spatially smoothed with a 2mm full width at half maximum kernel to increase signal-to-noise while preserving spatial specificity. Each scan was normalized to its mean. For each scan, signal quality was assessed within a region of interest (ROI) that encompassed the entire opercular surface of visual area V1 (which covered the central 6–7° of visual field, confirmed by retinotopic mapping at a later age). Square-wave functions matching the time course of the experimental design were convolved with a MION function[52]. Pearson correlation coefficients were calculated between the response functions and the average V1 response. We used both gaze direction data and average V1 responses to evaluate data quality. We deleted entire scans if the monkey was asleep or was not looking at the screen for at least 80% of the scan, and censored individual blocks if the V1 response did not fit the MION hemodynamic response convolved with the stimulus time-course with a correlation coefficient greater than 0.5. When individual blocks were censored, conditions were re-balanced by censoring blocks from the other stimulus conditions starting with the last scan, which typically had the poorest data quality and largest animal movement. Importantly, only V1 responses were used for block evaluation. This scan and block selection was necessary for quality control in these young monkeys, and inclusion of bad scans would have given a false perspective on the responses in each session. Each session was always analyzed first using the entire data set after motion censoring, and this always resulted in noisier, more variable maps, and never revealed stronger selectivity for one category vs. another than maps made using only accepted scans and blocks. For Supplementary Figure 2, eye tracking data from each session were averaged over each stimulus-on block type for the graph, and, for the heat maps, smoothed with a 0.5° Gaussian, and normalized to the maximum.

A multiple regression analysis (AFNI's 3dDeconvolve[62]) in the framework of a general linear model[63] was then performed for each scan session separately. Each stimulus condition was modeled with a MION-based hemodynamic response function[52]. Additional regressors that accounted for variance due to baseline shifts between time series, linear drifts, and head motion parameter estimates were also included in the regression model. Due to the timecourse normalization, resulting beta coefficients were scaled to reflect % signal change. Brain regions that responded more strongly to one image category than another were identified by contrasting presentation blocks of each of these image categories. Maps of beta coefficients were clustered (>5 adjacent voxels) and threshold at p<0.01 (FDR-corrected).

Each scan session had enough power to detect category-selective activity: In controls, face patches identified in scan sessions with as few as 7 of each block-type repetitions were virtually identical in location and extent to face patches identified in scan sessions with 27

repetitions. In contrast, in deprived monkeys face patches were undetected even in sessions with as many as 33 repetitions of each block type. Although the scan sessions for face and object stimuli covered equivalent age ranges for deprived and control animals, we did not start using hand stimuli in the (older) control monkeys until after we realized that the face-deprived monkeys had such strong hand patches. Given the consistency in location of face and hand patches across sessions (Figs. 1&2), and the consistency of face domains in control monkeys once they appear during development[20], it is probably safe to assume the hand patches at earlier ages in the control monkeys were similar to these maps.

Conjunction Analysis

A conjunction analysis was performed across sessions to evaluate the consistency of faces>objects and hands>objects beta maps in each monkey. Sessions were registered to the F99 standard macaque template[64] in each monkey, collapsed across hemisphere, and beta contrast maps were binarized at a threshold of p<0.01 (FDR corrected). The proportion of sessions for which each voxel was significantly more activated by faces>objects or by hands>objects were calculated for the 3 control monkeys and three face-deprived monkeys with multi-session data.

Regions of Interest

Several ROIs were used to quantify the imaging data. (1) A V1 ROI was drawn on the anatomy of each monkey covering the entire opercular surface of V1; we confirmed with retinotopic mapping in each monkey (except B10) that this region encompassed the central 6–7° of visual field. (2) An STS-Lower-Lip ROI was drawn on the F99 standard macaque brain[64]. This ROI included the entire lower bank and lip of the STS, from its anterior tip back to the prelunate gyrus and MT and V4 borders (identified by retinotopic mapping) posteriorly. Though the ROI was not constrained by functional data, it was chosen to optimally sample face-selective cortex in the control monkeys. The statistics from each scan session were then aligned to the F99 atlas brain, and the number of mm^3 voxels responsive to faces>object or to hands>objects at p<0.01 FDR corrected were calculated. (3) The OTS ROI was similarly drawn on the F99 atlas brain to include the OTS and its lower lip, and statistics from each scan session were then aligned to the F99 brain, and the number of voxels responsive to bodies>scenes or to scenes>bodies at p<0.01 FDR corrected were calculated. (4) CIT and AIT ROIs spanning the lower LIP of the STS, through the fundus and the upper bank were drawn on the F99 brain based on the average location of middle and anterior face/hand patches. We drew these ROIs such that they maximally sampled the face patches in control monkeys and hand patches in all monkeys in order to compare the relative mediolateral locations of face and hand responsiveness with respect to the anatomical landmarks, the banks and fundus of the STS. In each ROI, the shortest cortical distance was calculated between each data point and the crown of the lower lip of the STS. Data were then sorted into 1mm bins based on distance from the crown, and average face>object and hand>object beta contrasts were plotted for each bin. (5) Face and hand ROIS were calculated from conjunction maps of faces>objects and hands>objects (threshold at p<0.01 FDR corrected), respectively, and an overlap of >40% across scans was set to determine each ROI. The ventral scene ROIs were determined for each monkey based on single session scenes>bodies (threshold at p<0.01 FDR corrected).

Group-level Fixed-effects Analyses

Face vs Hand Patches—To compare face and hand patch sizes between normal and face-deprived monkeys, we performed a 3-way ANOVA on the means across sessions (including the single session for B3), with monkey as the random effects factor. Given the small sample size in each group, hemispheres were treated as separate data points and hemisphere was included as a factor in the analysis. There was a difference in the number of significant (p<0.01 FDR corrected) category-selective voxels (faces>objects or hands>objects) as well as an interaction with monkey group (controls vs deprived); (ANOVA, main effect of category, $F_{(1,21)}=12.56$; p=0.002, interaction between category and monkey group, $F_{(1,21)}=34.0$; p=0.0001, no main effects of monkey group or hemisphere, and no other interactions, $F_{(1,21)}<0.58$; p>0.45). Since we saw a main effect of category and an interaction with monkey group, we performed post-hoc comparisons. The average number of face-selective voxels in the STS ROI was smaller in the face-deprived group compared to controls (2-tailed unpaired t-test averaged over sessions, not hemispheres; $t_{(12)}=-7.6$; $p=6\times10^{-6}$). There was no significant difference in the average number of hand-selective voxels (2-tailed unpaired t-test; $t_{(12)}=1.5$; p=0.14). Though we used a typical number of subjects for a monkey fMRI study, the across-monkey statistics necessarily involve a small N. Nevertheless the results were robust for both across-monkey comparisons, as well as within-animal comparisons, for which we had a much larger N.

Body vs Scene Patches—Across all monkeys (averaged across sessions for B4&B5) the STS ROI had more significant bodies>scenes voxels, and the OTS had more scenes>bodies voxels (ANOVA; main effect of ROI; $F_{(1,33)}=5.89$, p=0.02; no main effect of monkey group $F_{(1,33)}=0.11$; p=0.74, or hemisphere $F_{(1,33)}=0.01$; p=0.95, and no interaction $F_{(1,33)}<0.57$; p>0.46.. Post-hoc t-tests showed that both control and deprived monkeys had more significant body-selective voxels than scene voxels in the STS (2-tailed paired t-test; controls: $t_{(3)}=11.3$; p=0.002; deprived: $t_{(5)}=3.1$; p=0.05), and that both control and deprived monkeys had more significant scene-selective voxels than body voxels in the OTS (2-tailed paired t-test, controls: $t_{(3)}=-14.7$; p=0.0007; deprived: $t_{(5)}=-8.2$; p=0.0004). There was no significant difference for either ROI between controls and face-deprived monkeys in the number of significant body voxels (2-tailed unpaired t-test; STS: $t_{(8)}=1.91$; p=0.09; OTS: $t_{(80}=0.06$; p=0.95) or scene voxels (STS: $t_{(8)}0.981$; p=0.35; OTS: $t_{(8)}=-1.29$; p=0.23).

Free-viewing Behavior—To evaluate looking behavior during free viewing as a function of monkey group, we performed an ANOVA on the means across sessions with monkey as the random effects factor. Mean looking time per monkey across sessions (>90 days) showed an interaction between monkey group (control vs. deprived) and feature category (faces vs. hands) (ANOVA; interaction, $F_{(1,10)}=43.13$, p < 0.0001; no main effects of either monkey group or feature category, $F_{(1,10)}<1.3$, p > 0.28).

Experimental design and statistical analysis

Seven monkeys (4 control, 3 face-deprived) were included in stimulus category and behavior analyses. A subset (2 control, and 2 face-deprived) were included in retinotopic mapping analyses. Individual subject and group-level analyses were performed. Though 3–4 monkeys per group is a small sample size for group-level analyses, this is a typical subject size for

non-human primate studies and effects were apparent in each individual monkey. Investigators were not blind as to the group or identity of each subject. As described in the Data analysis subsection, category selectivity maps for individual monkeys were determined via a linear regression analysis and threshold on the resulting t statistics. To assess the consistency of category responses for individual monkeys across sessions, a conjunction analysis was performed based on individual t statistics maps threshold at $p < 0.01$ (FDR-corrected). Consistency was further evaluated by comparing beta coefficient contrast maps across scans by Pearson correlation. To compare the size of category-selective regions, paired-sample t-tests were conducted in individual monkeys across scan sessions. To evaluate group-level stimulus category selectivity, an ANOVA was conducted on the mean number category-selective voxels (across sessions). Stimulus category, monkey group, and hemisphere were included as factors in the ANOVA. Post hoc t tests were then performed accordingly. Retinotopic maps were measured for individual subjects and group averages. Significance was assessed by an F ratio as described in the Phase-encoded Retinotopic Mapping subsection. To evaluate looking behavior during free viewing in individual monkeys, we performed t-tests between the mean looking time for different stimulus categories across viewing sessions. As reported in the Group-level Fixed-effects Analyses subsection in the methods, looking behavior at the group level was evaluated by an ANOVA on the mean looking time per monkey across sessions. Stimulus category and monkey group were included as factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. McKone E, Crookes K, Jeffery L, Dilks DD. A critical review of the development of face recognition: experience is less important than previously believed. Cogn Neuropsychol. 2012; 29:174–212. [PubMed: 22360676]

2. McKone, E., Crookes, K., Kanwisher, N. The cognitive and neural development of face recognition in humans. In: Gazzaniga, M., editor. The Cognitive Neurosciences. 4. MIT Press; Cambridge, MA: 2009. p. 467-482.

3. Goren CC, Sarty M, Wu PY. Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics. 1975; 56:544–549. [PubMed: 1165958]

4. Johnson MH, Dziurawiec S, Ellis H, Morton J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition. 1991; 40:1–19. [PubMed: 1786670]

5. Mendelson MJ, Haith MM, Goldman-Rakic PS. Face scanning and responsiveness to social cues in infant rhesus monkeys. Developmental Psychology. 1982; 18:222–228.

6. Cassia VM, Turati C, Simion F. Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? Psychol Sci. 2004; 15:379–383. [PubMed: 15147490]

7. Turati C, Simion F, Milani I, Umilta C. Newborns' preference for faces: what is crucial? Dev Psychol. 2002; 38:875–882. [PubMed: 12428700]

8. Tanaka JW, Farah MJ. Parts and wholes in face recognition. Q J Exp Psychol A. 1993; 46:225–245. [PubMed: 8316637]

9. Young AW, Hellawell D, Hay DC. Configurational information in face perception. Perception. 1987; 16:747–759. [PubMed: 3454432]

10. Stryker, MP., Allman, J., Blakemore, C., Greuel, JM., Kaas, JH., Merzenich, MM., Rakic, P., Singer, W., Stent, GS., Loos, Hvd, Wiesel, TN. Group report. Principles of cortical self-organization. In: Rakic, P., Singer, W., editors. Neurobiology of Neocortex. Wiley; New York: 1988. p. 115-136.

11. Cohen L, Dehaene S. Specialization within the ventral stream: the case for the visual word form area. Neuroimage. 2004; 22:466–476. [PubMed: 15110040]

12. Aguirre GK, Zarahn E, D'Esposito M. An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron. 1998; 21:373–383. [PubMed: 9728918]

13. Srihasam K, Vincent JL, Livingstone MS. Novel domain formation reveals proto-architecture in inferotemporal cortex. Nat Neurosci. 2014; 17:1776–1783. [PubMed: 25362472]

14. Levy I, Hasson U, Avidan G, Hendler T, Malach R. Center-periphery organization of human object areas. Nat Neurosci. 2001; 4:533–539. [PubMed: 11319563]

15. Malach R, Levy I, Hasson U. The topography of high-order human object areas. Trends Cogn Sci. 2002; 6:176–184. [PubMed: 11912041]

16. Martin A, Wiggs CL, Ungerleider LG, Haxby JV. Neural correlates of category-specific knowledge. Nature. 1996; 379:649–652. [PubMed: 8628399]

17. Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A. 1999; 96:9379–9384. [PubMed: 10430951]

18. Arcaro MJ, Livingstone MS. A hierarchical, retinotopic proto-organization of the primate visual system at birth. Elife. 2017; 6

19. Harlow HF, Suomi SJ. Social recovery by isolation-reared monkeys. Proc Natl Acad Sci U S A. 1971; 68:1534–1538. [PubMed: 5283943]

20. Livingstone MS, Vincent J, Arcaro M, Srihasam K, Schade PF, Savage T. Development of the macaque face-patch system. Nature Communications. 2017; 8:14897.

21. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB. Faces and objects in macaque cerebral cortex. Nat Neurosci. 2003; 6:989–995. [PubMed: 12925854]

22. Pinsk MA, Desimone K, Moore T, Gross CG, Kastner S. Representations of faces and body parts in macaque temporal cortex: A functional MRI study. Proc Natl Acad Sci U S A. 2005; 102:6996–7001. [PubMed: 15860578]

23. Kornblith S, Cheng X, Ohayon S, Tsao DY. A network for scene processing in the macaque temporal lobe. Neuron. 2013; 79:766–781. [PubMed: 23891401]

24. Arcaro MJ, Livingstone MS. Retinotopic Organization of Scene Areas in Macaque Inferior Temporal Cortex. J Neurosci. 2017

25. Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RB. Scene-selective cortical regions in human and nonhuman primates. J Neurosci. 2011; 31:13771–13785. [PubMed: 21957240]

26. Arcaro MJ, Pinsk MA, Li X, Kastner S. Visuotopic organization of macaque posterior parietal cortex: a functional magnetic resonance imaging study. J Neurosci. 2011; 31:2064–2078. [PubMed: 21307244]

27. Janssens T, Zhu Q, Popivanov ID, Vanduffel W. Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. J Neurosci. 2013; 34:10156–10167.

28. Kolster H, Mandeville JB, Arsenault JT, Ekstrom LB, Wald LL, Vanduffel W. Visual field map clusters in macaque extrastriate visual cortex. J Neurosci. 2009; 29:7031–7039. [PubMed: 19474330]

29. Lafer-Sousa R, Conway BR. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat Neurosci. 2013; 16:1870–1878. [PubMed: 24141314]

30. Hasson U, Levy I, Behrmann M, Hendler T, Malach R. Eccentricity bias as an organizing principle for human high-order object areas. Neuron. 2002; 34:479–490. [PubMed: 11988177]

31. Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI. Cortical representations of bodies and faces are strongest in commonly experienced configurations. Nat Neurosci. 2010; 13:417–418. [PubMed: 20208528]

32. de Haas B, Schwarzkopf DS, Alvarez I, Lawson RP, Henriksson L, Kriegeskorte N, Rees G. Perception and Processing of Faces in the Human Brain Is Tuned to Typical Feature Locations. J Neurosci. 2016; 36:9289–9302. [PubMed: 27605606]

33. Issa EB, DiCarlo JJ. Precedence of the eye region in neural processing of faces. J Neurosci. 2012; 32:16666–16682. [PubMed: 23175821]

34. Golarai G, Liberman A, Yoon JM, Grill-Spector K. Differential development of the ventral visual cortex extends through adolescence. Front Hum Neurosci. 2010; 3:80. [PubMed: 20204140]

35. Natu VS, Barnett MA, Hartley J, Gomez J, Stigliani A, Grill-Spector K. Development of Neural Sensitivity to Face Identity Correlates with Perceptual Discriminability. J Neurosci. 2016; 36:10893–10907. [PubMed: 27798143]

36. Pascalis O, de Haan M, Nelson CA. Is face processing species-specific during the first year of life? Science. 2002; 296:1321–1323. [PubMed: 12016317]

37. Deen B, Richardson H, Dilks DD, Takahashi A, Keil B, Wald LL, Kanwisher N, Saxe R. Organization of high-level visual cortex in human infants. Nat Commun. 2017; 8:13995. [PubMed: 28072399]

38. Ponce CR, Hartmann TS, Livingstone MS. End-Stopping Predicts Curvature Tuning along the Ventral Stream. J Neurosci. 2017; 37:648–659. [PubMed: 28100746]

39. Ferrari PF, Visalberghi E, Paukner A, Fogassi L, Ruggiero A, Suomi SJ. Neonatal imitation in rhesus macaques. PLoS Biol. 2006; 4:e302. [PubMed: 16953662]

40. Meltzoff AN, Moore MK. Newborn infants imitate adult facial gestures. Child Dev. 1983; 54:702–709. [PubMed: 6851717]

41. Turati C, Macchi Cassia V, Simion F, Leo I. Newborns' face recognition: role of inner and outer facial features. Child Dev. 2006; 77:297–311. [PubMed: 16611173]

42. Oostenbroek J, Suddendorf T, Nielsen M, Redshaw J, Kennedy-Costantini S, Davis J, Clark S, Slaughter V. Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. Curr Biol. 2016; 26:1334–1338. [PubMed: 27161497]

43. James, W. Rouben Mamoulian Collection (Library of Congress). The principles of psychology. H Holt and company; New York: 1890.

44. Sugita Y. Face perception in monkeys reared with no exposure to faces. Proc Natl Acad Sci U S A. 2008; 105:394–398. [PubMed: 18172214]

45. Kanwisher N. Functional specificity in the human brain: a window into the functional architecture of the mind. Proc Natl Acad Sci U S A. 2010; 107:11163–11170. [PubMed: 20484679]

46. Bushnell IWR. Mother's face recognition in newborn infants: learning and memory. Infant and Child Development. 2001; 10:67–74.

47. Hebb, DO. The organization of behavior; a neuropsychological theory. Wiley; New York: 1949.

48. Livingstone MS. Ocular dominance columns in New World monkeys. J Neurosci. 1996; 16:2086–2096. [PubMed: 8604053]

49. Wiesel TN, Hubel DH. Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. J Neurophysiol. 1965; 28:1029–1040. [PubMed: 5883730]

50. Blasdel G, Obermayer K, Kiorpes L. Organization of ocular dominance and orientation columns in the striate cortex of neonatal macaque monkeys. Vis Neurosci. 1995; 12:589–603. [PubMed: 7654611]

51. Willenbockel V, Sadr J, Fiset D, Horne G, Gosselin F, tanaka J. The SHINE toolbox for controlling low-level image properties. Journal of Vision. 2010; 10:653.

52. Leite FP, Tsao D, Vanduffel W, Fize D, Sasaki Y, Wald LL, Dale AM, Kwong KK, Orban GA, Rosen BR, Tootell RB, Mandeville JB. Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. Neuroimage. 2002; 16:283–294. [PubMed: 12030817]

53. Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA. Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. Neuron. 2001; 32:565–577. [PubMed: 11719199]

54. Wass SV, Smith TJ, Johnson MH. Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. Behav Res Methods. 2013; 45:229–250. [PubMed: 22956360]

55. Janssens T, Zhu Q, Popivanov ID, Vanduffel W. Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. J Neurosci. 2014; 34:10156–10167. [PubMed: 25080579]

56. Kolster H, Janssens T, Orban GA, Vanduffel W. The Retinotopic Organization of Macaque Occipitotemporal Cortex Anterior to V4 and Caudoventral to the Middle Temporal (MT) Cluster. J Neurosci. 2014; 34:10168–10191. [PubMed: 25080580]

57. Gandhi SP, Heeger DJ, Boynton GM. Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci U S A. 1999; 96:3314–3319. [PubMed: 10077681]

58. Martinez A, Anllo-Vento L, Sereno MI, Frank LR, Buxton RB, Dubowitz DJ, Wong EC, Hinrichs H, Heinze HJ, Hillyard SA. Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci. 1999; 2:364–369. [PubMed: 10204544]

59. Somers DC, Dale AM, Seiffert AE, Tootell RB. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. Proc Natl Acad Sci U S A. 1999; 96:1663–1668. [PubMed: 9990081]

60. Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. Proc Natl Acad Sci U S A. 2008; 105:19514–19519. [PubMed: 19033466]

61. Logothetis NK, Guggenberger H, Peled S, Pauls J. Functional imaging of the monkey brain. Nat Neurosci. 1999; 2:555–562. [PubMed: 10448221]

62. Cox R. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29:162–173. [PubMed: 8812068]

63. Friston KJ, Frith CD, Turner R, Frackowiak RS. Characterizing evoked hemodynamics with fMRI. Neuroimage. 1995; 2:157–165. [PubMed: 9343598]

64. Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH. An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc. 2001; 8:443–459. [PubMed: 11522765]

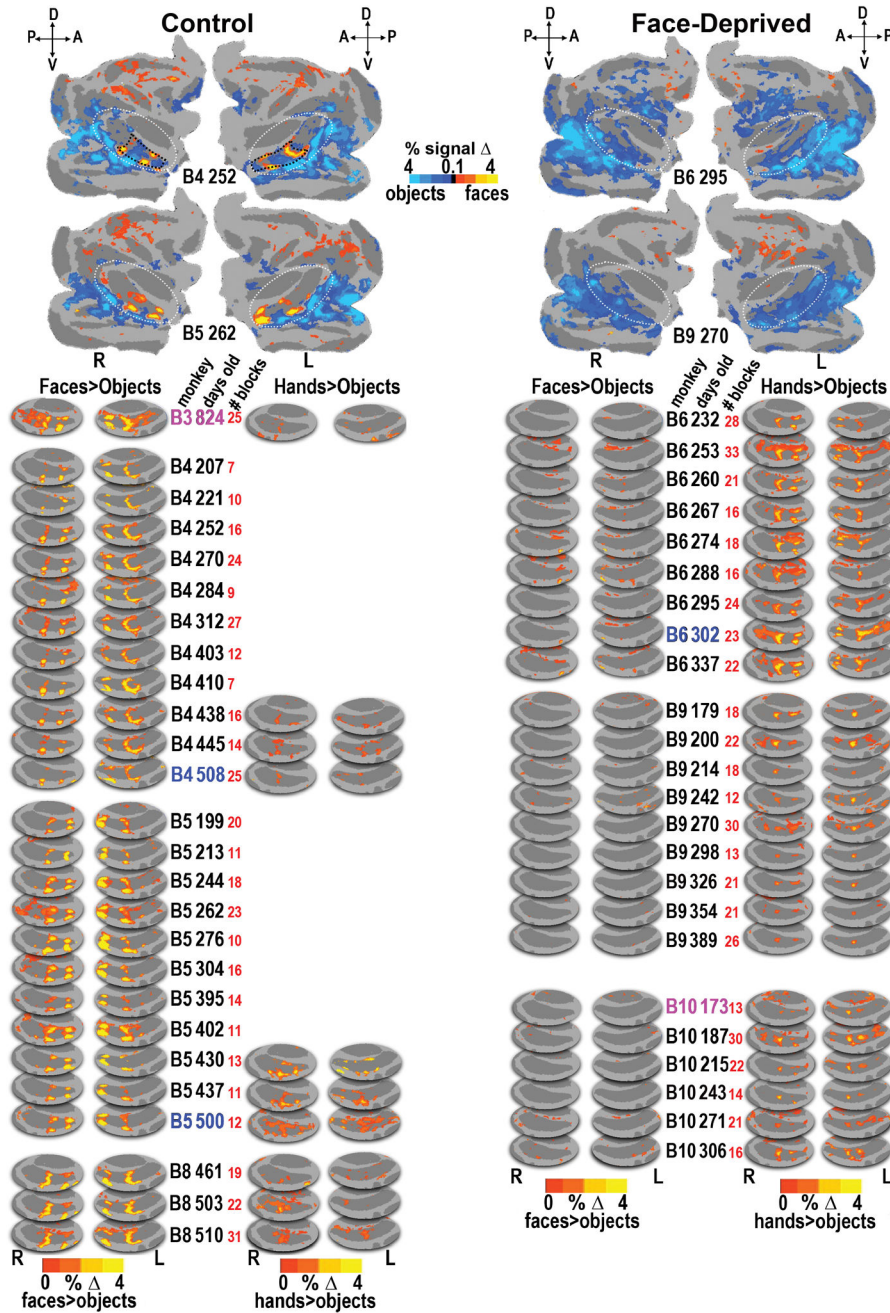**Figure 1.**
Faces>objects and hands>objects activations in control and face-deprived monkeys. (top)
Example maps for the contrast faces-minus-objects aligned onto a standard F99 macaque
flattened cortical surface (light gray=gyri; dark gray=sulci). These examples show % signal
change (beta coefficients) thresholded at p<0.01 (FDR corrected) for one session each for
two control (B4&B5; left) and two face-deprived (B6&B9; right) monkeys at 252, 262, 295,
& 270 days old, as indicated. Dotted white ovals indicate the STS region shown in the
bottom half for all the scan sessions for all 7 monkeys. Dotted black outline in the B4 map
shows the STS ROI used for the correlation analysis described in the text. (bottom)

Faces>objects and hands>objects beta coefficients thresholded at p<0.01 (FDR corrected) for all the individual scan sessions at the ages (in days) indicated for 4 control monkeys (B3,B4,B5&B8; left) and 3 face-deprived monkeys (B6,B9&B10; right). For all but the 5 scan sessions indicated by colored lettering, the images were single large monkey faces, monkey hands, or objects (Supplementary Fig. 1). For the scan sessions B4 508, B5 500, and B6 302 the hands were gloved human hands; for scan sessions B3 824 and B10 173 the stimuli were mosaics of monkey faces, monkey hands, or objects rather than single large images (these were sessions in which scenes were also presented, and mosaics were used to equate visual-field coverage; as in Fig. 3 top). The small red numbers indicate the number of blocks of each category type included in the analysis for each session.
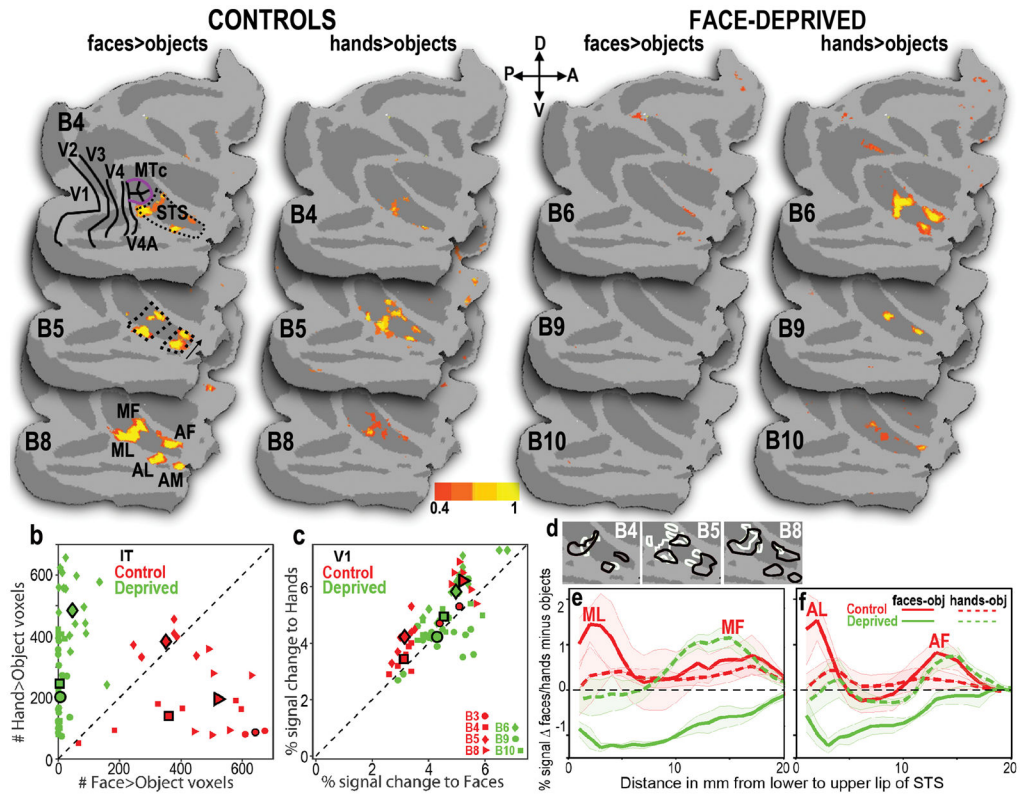
Figure 2.
Quantification of face and hand activations in control and face-deprived monkeys. (a)
Conjunction analysis: Flat maps show the proportion of sessions for each monkey (that was
scanned in multiple sessions) for which each voxel was significantly more activated by
faces>objects or by hands>objects for 3 control monkeys (left) and three face-deprived
monkeys (right). Significance threshold for each session was p<0.01 (FDR corrected). The 5
previously described face patches typically found in control monkeys are labeled in the B8
face map[21,22]. Across sessions, face patch AM was consistently significant in only monkey
B8, though present subthreshold in B4 and B5. In B4 areal borders between V1, V2, V3, V4,
and V4A are indicated in black, and the MT cluster is outlined in purple; areas were
determined by retinotopic mapping at 2 years of age in the same monkey. Graphs b,c,e,f
provide a quantitative comparison of face and hand activations in 4 control and 3 face-
deprived monkeys. (b) Number of voxels in an anatomically-defined entire-STS ROI
(outlined in B4 face map) that were significantly more activated by faces than by objects
(horizontal axis) compared to the number of voxels in the same ROI that were significantly
more activated by hands than by objects (vertical axis); voxels selected at p<0.01 FDR
corrected threshold. Different symbols represent different monkeys (see legend in panel c),
and each symbol represents data from one hemisphere from one scan session for one control
(red) or one face-deprived (green) monkey. Large black-outlined symbols show means for
each monkey across sessions and hemispheres (for B3 only across hemispheres). (c) Percent
signal change in a central V1 ROI (central 6–7° of visual field) in response to faces
(horizontal axis) and to hands (vertical axis). Each symbol represents data from one
hemisphere for one scan session for one monkey; large symbols show means for each

monkey that was scanned multiple times; single-session mean across hemispheres for B3 is highlighted in black. (d) Overlays of face (black) and hand (white) patches for each of the control monkeys from the conjunction analysis. (e) Average percent signal change (shading indicates sem) to faces-minus-objects (solid lines) and hands-minus-objects (dotted lines) as a function of ventrolateral to dorsomedial distance along an anatomically defined CIT ROI extending from the crown of the lower lip of the STS to the upper bank of the STS (left dotted outline on the B5 face map; arrow on map corresponds to x axis on graph). Data were averaged over all sessions and both hemispheres of all three control monkeys that were scanned multiple times (red lines) and all three face-deprived monkeys (green lines). Activations were averaged across the AP dimension of the ROI to yield an average response as a function of mediolateral location. (f) Same as (e) but for an anatomically defined AIT ROI (right dotted outline on the B5 face map).
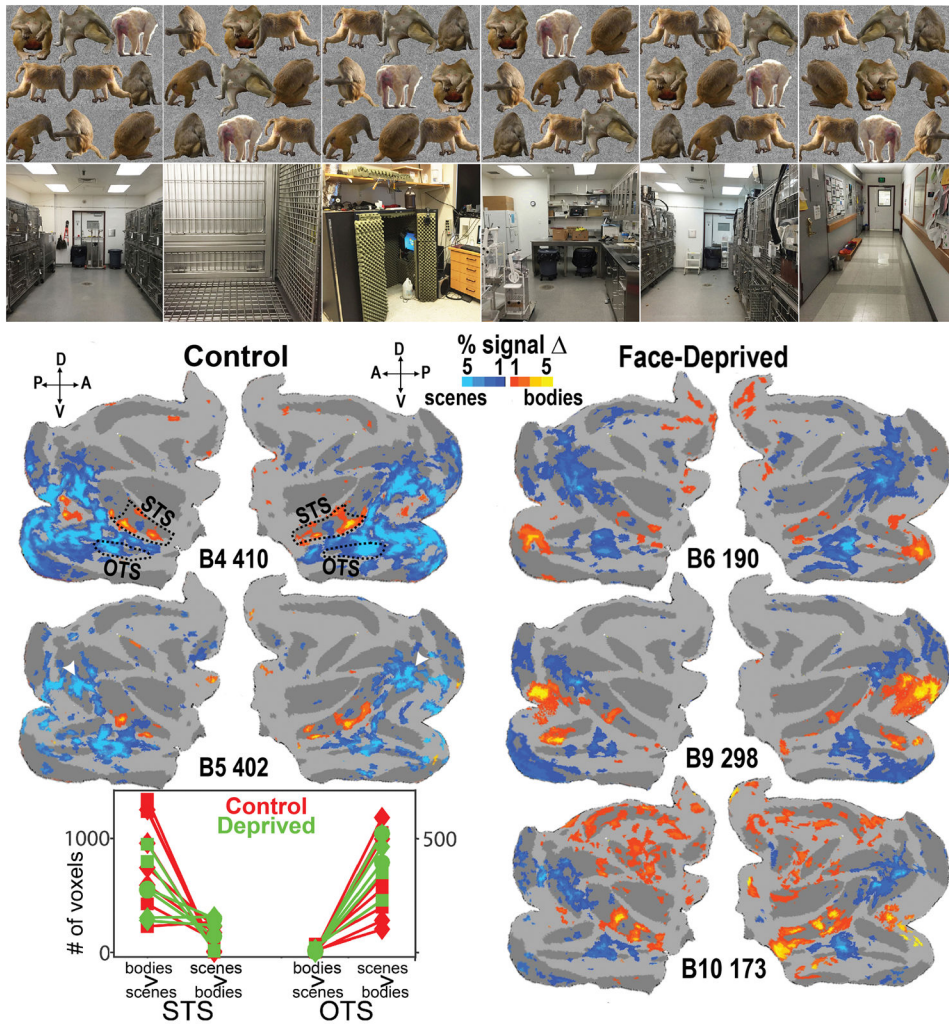
Figure 3.
Activations to bodies vs scenes. (top) Subset of the images used for bodies vs scenes scans. (bottom) Maps of responses (beta values) for bodies vs scenes, thresholded at p<0.01, FDR corrected, for 2 control monkeys and 3 face-deprived monkeys at the ages indicated in days. The two control monkeys were scanned in an additional session, with similar activation patterns (B4 326 and B5 318). Scene activations that corresponded to previously described scene patches were present in all monkeys: the ventral scene region in each monkey corresponds to the Lateral Place Patch (LPP) in the occipitotemporal sulcus[23] (outlined in the B4 map), and the more dorsal region (white arrowheads in B5 map) corresponds to the Dorsal Scene Patch[25], in dorsal prelunate cortex, the potential homologue of human TOS. Note also the stereotyped bodies>scenes activations lying along the STS in all monkeys. The graph shows the number of voxels in each monkey in each session in each hemisphere in each ROI (outlined in the B4 maps) that was significantly more responsive to bodies than to scenes or the reverse, as indicated (voxels selected at a p<0.01 FDR corrected threshold); each symbol corresponds to one monkey, as in Fig. 2c; red symbols controls, green deprived.
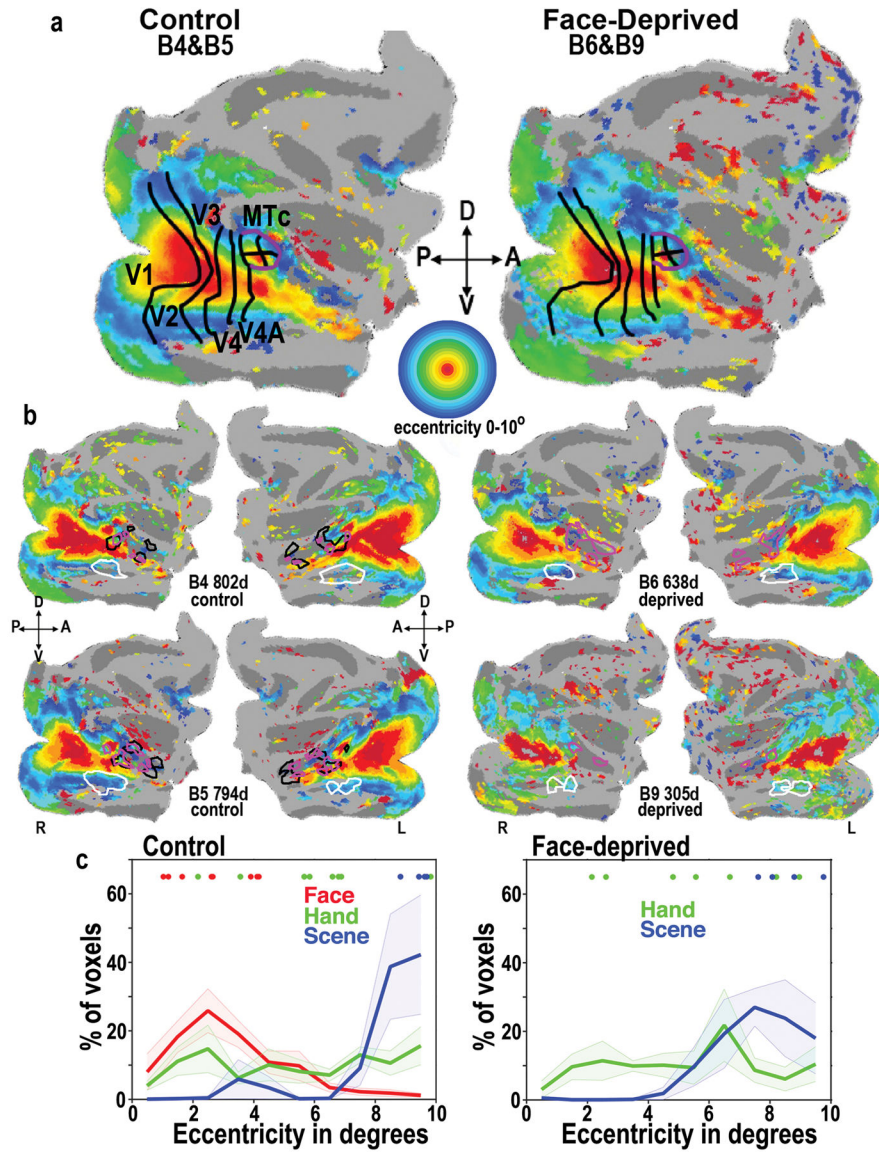
Figure 4.
Eccentricity organization in control and face-deprived monkeys. Eccentricity, from 0° (red) to 10° (blue) as indicated by the circular color scale, was mapped using expanding and contracting annuli of flickering checkerboard patterns. (a) Average eccentricity organization for two control monkeys (B4&B5) on the left, and two face-deprived monkeys (B6&B9) on the right. Thick black lines indicate areal borders of (L to R) V1, V2, V3, V4 V4A, and the purple outline indicates the MT cluster, comprising areas MT, MST, FST< and V4t. Across 19 retinotopic areas[26–28] the mean voxelwise distance between the average control-monkey eccentricity map and the average face-deprived eccentricity map was 0.9°±1°s.d. The correlation (r) between control and face-deprived averaged maps was 0.82. Even considering only the inferotemporal cortical areas the mean difference between control and deprived averaged maps was 1.4°±1.5°s.d.; correlation = 0.66, which is comparable to the consistency between individual control monkeys (1.84°±1.76°s.d.; correlation = 0.54). (b) Individual

maps for each hemisphere for each monkey, thresholded at p<0.01. Outlines of the conjunction ROIs for faces (black), hands (magenta), and the ventral scene area (white) for each hemisphere in each monkey are overlaid on each monkey's own eccentricity map. The B9 eccentricity map is less clear than those of the older monkeys because younger monkeys do not fixate as well or as long as older monkeys. (c) Histograms of eccentricity representations (1° bins) within each category-selective ROIs averaged across hemispheres for control and deprived monkeys. Shading indicates sem. Dots above graphs indicate the mapped eccentricity of the most category-selective voxel in each face-, hand-, or scene- ROI for each hemisphere for each monkey for CIT patches and AIT patches; color indicates ROI category.
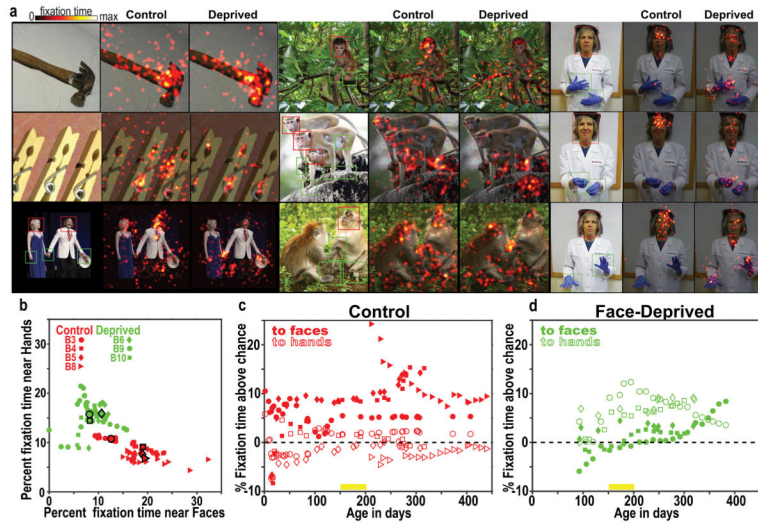
Figure 5.
Looking behavior of control and face-deprived monkeys. (a) Normalized heat maps of fixations for 9 of the images the monkeys freely viewed while gaze was monitored. In the leftmost panel for each of the face&hand images, red boxes outline the face regions and green boxes the hand regions used for analysis in panels b–d. Heat maps combine data from 4 control and 3 face-deprived monkeys overlaid onto a darkened version of the image viewed. (Left bottom photograph by Alan Stubbs, courtesy of the Neural Correlate Society. Middle middle photograph courtesy of David M. Barron/oxygengroup. Lower middle photograph by Pawel Opaska, used with permission. Top middle image is a substitute image for a configurally identical image that we do not have permission to reproduce.) (b) Quantification of looking behavior for 4 control (red) and 3 face-deprived (green) monkeys from 90 days to 1 year of age. Large black-outlined symbols show mean for each monkey. (c) Longitudinal looking behavior for 4 control monkeys from 2 days to >1 year of age at regions containing faces (filled symbols) or at regions containing hands (open symbols). Each symbol represents the % average looking time at hand or face regions, minus the % of the entire image covered by the hand or face region, over one viewing session. Yellow bar indicates the range of ages at which control monkeys developed strong stable face patches[20]. (d) Same as c for 3 face-deprived monkeys from 90 to 383 days of age.