Representation of Contextually Related Multiple Objects in the Human Ventral Visual Pathway

Yiying Song[1], Yu L. L. Luo[1], Xueting Li[1], Miao Xu[1], and Jia Liu[1,2]

Abstract

■ Real-world scenes usually contain a set of cluttered and yet contextually related objects. Here we used fMRI to investigate where and how contextually related multiple objects were represented in the human ventral visual pathway. Specifically, we measured the responses in face-selective and body-selective regions along the ventral pathway when faces and bodies were presented either simultaneously or in isolation. We found that, in the posterior regions, the response for the face and body pair was the weighted average response for faces and bodies presented in isolation. In contrast, the anterior regions encoded the face and body pair in a mutually facilitative fashion, with the response for the pair significantly higher than that for its constituent objects. Furthermore, in the right fusiform face area, the face and body pair was represented as one inseparable object, possibly to reduce perceptual load and increase representation efficiency. Therefore, our study suggests that the visual system uses a hierarchical representation scheme to process multiple objects in natural scenes: the average mechanism in posterior regions helps retaining information of individual objects in clutter, whereas the nonaverage mechanism in the anterior regions uses the contextual information to optimize the representation for multiple objects. ■

INTRODUCTION

In a natural environment, objects rarely appear in isolation. Instead, objects usually appear within a visual context with typically associated objects. Previous behavioral studies have reported that processing of objects is facilitated by other contextually related objects in object detection and recognition (Auckland, Cave, & Donnelly, 2007; Davenport, 2007; Bar & Ullman, 1996; Biederman, Mezzanotte, & Rabinowitz, 1982), object categorization (Chaigneau, Barsalou, & Zamani, 2009), and visual search (Hollingworth, 2009; Chun, 2000). Here we used fMRI to investigate where and how contextually related multiple objects were represented in the human ventral visual pathway.

Previous neurophysiology and fMRI studies mainly focus on the representation of multiple objects when they are contextually unrelated. These studies have consistently demonstrated that responses for preferred objects are usually reduced with the presence of unrelated nonpreferred objects in primate ventral visual pathway (Kastner, De Weerd, Desimone, & Ungerleider, 1998; Rolls & Tovee, 1995; Miller, Gochin, & Gross, 1993), and responses for object pairs are approximately the weighted average response for objects presented in isolation (MacEvoy & Epstein, 2009, 2011; Reddy, Kanwisher, & VanRullen, 2009; Zoccolan, Cox, & DiCarlo, 2005). However, because object processing is facilitated by contextually related objects, we expect that the response for contextually related multiple objects would deviate from the weighted average rule, with its response higher than the response for the preferred objects presented in isolation. Here, we chose face and body stimuli in particular because (1) they always appear together in a natural environment and (2) they are either preferred or nonpreferred stimuli for the face-selective or body-selective regions, respectively. The contextual influence on the representation of the face and body pair (i.e., the context effect) was calculated by comparing neural responses when faces and bodies were presented together (the FB condition) with when the preferred object was presented in isolation (the face condition or the body condition; Figure 1).

Previous fMRI studies have identified several face-selective regions and body-selective regions in the ventral visual pathway (Liu, Harris, & Kanwisher, 2010; Peelen & Downing, 2005; Downing, Jiang, Shuman, & Kanwisher, 2001; Kanwisher, McDermott, & Chun, 1997). Here instead of examining one particular region, we examined the representation for the face and body pair in both the anterior and posterior face-selective and body-selective regions. Because representations in the anterior portion of the pathway tend to be more abstract and global than those in the posterior portion (Grill-Spector & Malach, 2004), our first prediction is that the context effect is more likely to be observed in the anterior regions than in the posterior ones. Second, recent behavioral studies have shown that faces and bodies facilitate the perception of each other because the face and body pair is perceived as

[1]Beijing Normal University, [2]Chinese Academy of Sciences

Journal of Cognitive Neuroscience 25:8, pp. 1261–1269
doi:10.1162/jocn_a_00406

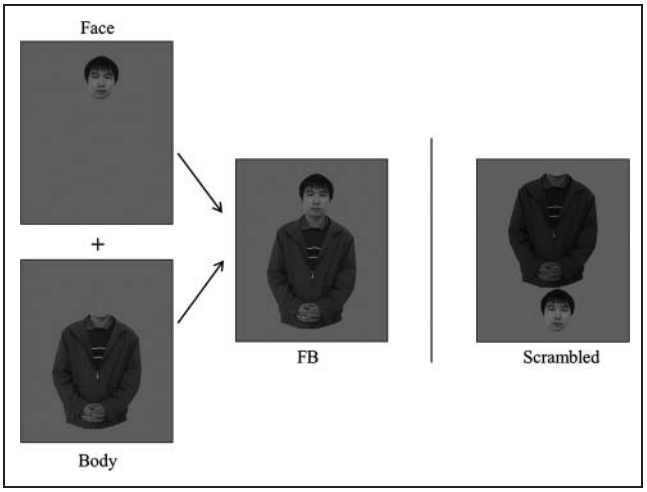

Figure 1. Stimulus exemplars.

an inseparable whole (Aviezer, Trope, & Todorov, 2012; Ghuman, McDaniel, & Martin, 2010). In addition, fMRI studies have shown that the fusiform face area (FFA) is responsive to body stimuli even when blurred blobs, not faces, are presented on the top of bodies (Cox, Meyers, & Sinha, 2004; see also Andrews, Davies-Thompson, Kingstone, & Young, 2010; Brandman & Yovel, 2010), suggesting that the FFA encodes contextual body cues. Thus, our second prediction is that the face and object pair in the FB condition might be represented holistically in the FFA, no longer similar to the representations of either its constituent objects (faces or bodies) or the face and object pair with the contextual relation being disrupted (the scrambled condition; Figure 1).

METHODS

Participants

Fourteen volunteers (age = 22–32 years; seven women) participated in the experiment. All participants were right-handed and had normal or corrected-to-normal visual acuity. The fMRI protocol was approved by the institutional review board of Beijing Normal University, Beijing, China. Written informed consent was obtained from all participants before their participation.

Stimuli

Four types of stimuli created from grayscale human frontal-view photos were used in the experiment to examine cortical representations of contextually related multiple objects (Figure 1). In the face condition, faces were presented in the upper location of the visual field, whereas in the body condition, bodies were presented in the lower location. In the FB (abbreviated for faces and bodies) condition, faces and bodies appeared in the upper and lower locations simultaneously. Evidence from previous fMRI studies has shown that object-selective regions in ventral visual pathway are sensitive to the location of objects presented in the visual field (e.g., Chan, Kravitz, Truong, Arizpe, & Baker, 2010; Kravitz, Kriegeskorte, & Baker, 2010; Malach, Levy, & Hasson, 2002); therefore, the location of faces (or bodies) in the face (or body) condition was strictly matched to the location of faces (or bodies) in the FB condition. In addition, we also included a scrambled condition in which the locations of faces and bodies in the FB condition were reversed, with faces appearing under bodies. This condition was used to examine whether faces and bodies in the FB condition were integrated as an inseparable whole.

Procedure

Each volunteer participated in a single session consisting of (1) two blocked-design functional localizer runs and (2) four blocked-design experimental runs. The localizer scan consisted of four conditions, that is, human frontal-view faces, human body parts, scenes, and line-drawing objects. Each localizer run lasted 5 min and 36 sec, consisting of sixteen 16-sec blocks with five 16-sec blocks of fixation periods interleaved. In each nonfixation block, 20 different exemplars of a given condition were shown on the screen, each of which was displayed for 300 msec followed by a blank interval of 500 msec. During the scan, participants pressed a button whenever they saw two identical stimuli in a row (i.e., identity 1-back task).

The experimental scan consisted of four conditions: face, body, FB (i.e., face above body), and scrambled (i.e., face under body). Each experimental run lasted 5 min and 20 sec, consisting of four 60-sec blocks with five 16-sec blocks of fixation periods interleaved. Each nonfixation block consisted of 40 trials of a given condition. In a trial, an exemplar of a given condition was presented for 700 msec, followed by a blank interval varying from 300 to 1300 msec. Participants were instructed to perform the identity 1-back task. Specifically, in the face and body blocks, participants reported the repetition of faces and bodies, respectively. In the FB and scrambled blocks, either faces or bodies in the composite images would be repeated, and participants were instructed to detect the repetition of either faces or bodies in a row. This task was designated to encourage participants to divide attention among objects in the composite images. There were four repetitions (i.e., targets) in a block.

MRI Data Acquisition

Scanning was conducted on a 3T Siemens Trio scanner (Erlangen, Germany) with a 12-channel phase-arrayed coil at BNU Imaging Center for Brain Research, Beijing, China. Twenty-five 4-mm-thick (20% skip) axial slices were collected (in-plane resolution = 3 × 3 mm). T2*-weighted gradient-echo, EPI procedures were used (repetition time = 2 sec, echo time = 32 msec, flip angle = 90°). In

addition, MPRAGE, an inversion prepared gradient-echo sequence (repetition time/echo time/inversion time = 2.73 sec/3.44 msec/1 sec, flip angle = 7°, voxel size 1.1 × 1.1 × 1.9 mm), was used to acquire 3-D structural images.

fMRI Data Analysis

Functional data were analyzed with the Freesurfer functional analysis stream (CorTechs, Inc., Charlestown, MA; Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999), fROI (froi.sourceforge.net), and in-house Matlab code. Data preprocessing included head motion correction (by aligning each volume to the first volume of the first run), intensity normalization (by scaling the intensity at each voxel at each time point with the global mean intensity averaged across all voxels and time points of the entire functional volume for each run), and spatial smoothing (with a Gaussian kernel of 6-mm FWHM). After data preprocessing, each condition was modeled by a boxcar regressor matching its time course, which was then convolved with a gamma function (delta = 2.25, tau = 1.25).

The functional ROI approach was used. First, the localizer scan was used to localize regions selective for faces, bodies, scenes, and objects separately for each participant. The face-selective regions were defined as the set of contiguous voxels by the contrast of faces versus the remaining conditions (i.e., objects, scenes, and bodies; $p < .01$, uncorrected) in the fusiform gyrus (FFA; Kanwisher et al., 1997) and in the inferior occipital cortex (occipital face area, OFA; Gauthier et al., 2000) respectively. Similarly, the body-selective regions were defined by the contrast of bodies versus the remaining conditions in the fusiform gyrus (fusiform body area, FBA; Peelen & Downing, 2005) and in the extrastriate cortex (extrastriate body area, EBA; Downing et al., 2001), respectively. Also, the object-selective regions were defined by the contrast of objects versus the remaining conditions in posterior fusiform gyrus (pFs) and in the lateral occipital cortex (LO), respectively (Grill-Spector et al., 1999). Finally, the parahippocampal place area (PPA; Epstein & Kanwisher, 1998) was defined by the contrast of scenes versus the remaining conditions in the parahippocampal gyrus. The object-selective and place-selective regions were used as control regions. All ROIs were successfully localized in both hemispheres in every participant (Table 1), possibly because we adopted a liberal threshold. To facilitate comparisons with previous studies, approximate Talairach coordinates of the ROIs (Table 1) were derived by registering the structural volume for each participant to the Montreal Neurological Institute (MNI305) atlas and then transforming the resulting coordinates to standard Talairach space using an algorithm developed by Matthew Brett (imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach).

After the definition of the ROIs, percent signal changes were extracted and then averaged by conditions across all experimental runs and all voxels within each predefined ROI for each participant. The magnitude of response in an ROI was measured as the mean percent signal changes in a block as compared with the baseline fixation. Percent signal changes, one per experimental condition per ROI per participant, were submitted to repeated-measures ANOVA, followed by post hoc pairwise two-tailed t tests. Previous studies have shown that FFA response for an object pair consisting of a face and an object from contextually unrelated categories (e.g., houses, cars, and shoes) was approximately equal to the response for faces appearing in isolation, corresponding to a weighted average response for faces and objects with the weight for faces being 1 (Reddy et al., 2009), whereas LO response for an object pair was approximately the weighted averaged response with a weight of 0.5 for each object (MacEvoy & Epstein, 2009, 2011). Therefore, in our study, the context effect was defined as the response difference between the FB condition and the face condition in the FFA and OFA, the response difference between the FB condition and the body condition in the FBA and EBA, and the response difference between the FB condition and the average of the face and body conditions in the pFs and LO.

To further examine whether faces and bodies in the FB condition were represented holistically, multivoxel pattern analysis (MVPA) was used to examine which spatial pattern of neural activation, the FB condition or the scrambled condition, was more similar to the "synthetic" pattern created by averaging the spatial pattern for the face condition and that for body condition (MacEvoy & Epstein, 2009, 2011). In particular, activation patterns

Table 1. Talairach Coordinates and Voxel Numbers of ROIs Averaged across Participants (Mean ± *SD*)

		Talairach Coordinates			*Voxel*
ROI	*Hemisphere*	*x*	*y*	*z*	*Numbers*
FFA	Right	40 ± 5	−53 ± 8	−13 ± 3	52 ± 15
	Left	−43 ± 5	−60 ± 8	−13 ± 4	29 ± 5
OFA	Right	38 ± 5	−80 ± 4	−4 ± 4	41 ± 6
	Left	−37 ± 7	−83 ± 8	−3 ± 4	31 ± 8
FBA	Right	43 ± 6	−49 ± 7	−13 ± 4	52 ± 10
	Left	−45 ± 4	−51 ± 8	−12 ± 5	49 ± 7
EBA	Right	48 ± 7	−71 ± 7	7 ± 5	241 ± 29
	Left	−46 ± 4	−76 ± 8	8 ± 8	195 ± 21
LO	Right	33 ± 6	−85 ± 8	3 ± 6	51 ± 9
	Left	−33 ± 6	−88 ± 7	−1 ± 6	62 ± 13
pFs	Right	31 ± 3	−50 ± 8	−10 ± 7	62 ± 6
	Left	−32 ± 6	−55 ± 11	−12 ± 5	94 ± 9
PPA	Right	30 ± 3	−47 ± 6	−7 ± 5	142 ± 11
	Left	−28 ± 3	−50 ± 7	−7 ± 4	158 ± 19

were extracted for each condition and each ROI from each of the four experimental runs. Data were then divided into halves for every possible binary split of the four scans, totaling six splits. For each split, patterns for each condition were averaged within each half of the data. A "cocktail" mean pattern (i.e., the average pattern across all conditions) was calculated separately for each half of the data and was then subtracted from each of the individual patterns before classification. To examine whether faces and bodies in the FB condition were represented holistically, a 1 nearest neighbor classifier with a Euclidean distance metric was used, and classification based on the similarity in spatial patterns was conducted through a series of pairwise comparisons between the synthetic pattern in one half of the data and the pattern for the FB or the scrambled condition in the other half (MacEvoy & Epstein, 2011; Haxby et al., 2001). If faces and bodies in the FB condition were integrated as an inseparable whole, they were no longer able to be decomposed into two components (i.e., face and body). Therefore, the synthetic pattern would be more similar to the spatial pattern for the scrambled condition than the FB condition. In contrast, if the representation of faces and bodies in the FB condition was additive, the synthetic pattern would be equally classified as either the FB condition or the scrambled condition (i.e., 50%).

RESULTS

Context Effect Increases along the Ventral Pathway

In our study, the context effect in the predefined face-selective and body-selective regions was quantified as the difference between the responses for the face and body pairs and that for preferred stimuli appearing alone. First, we examined the context effect in the posterior (the OFA) and anterior (the FFA) face-selective regions by comparing the response magnitude for the FB condition with that for the face condition. The context effect was observed in the FFA, with the response for the FB condition significantly higher than that for the face condition in both the left ($t(13) = 6.13$, $p < .001$) and right FFA ($t(13) = 6.58$, $p < .001$; Figure 2A, left). In contrast, the response for the FB condition was not significantly different from that for the face condition in either the left or right OFA ($ps > .05$). In other words, the OFA responded to the FB stimuli as if only the faces were presented, meaning that the response for the FB condition

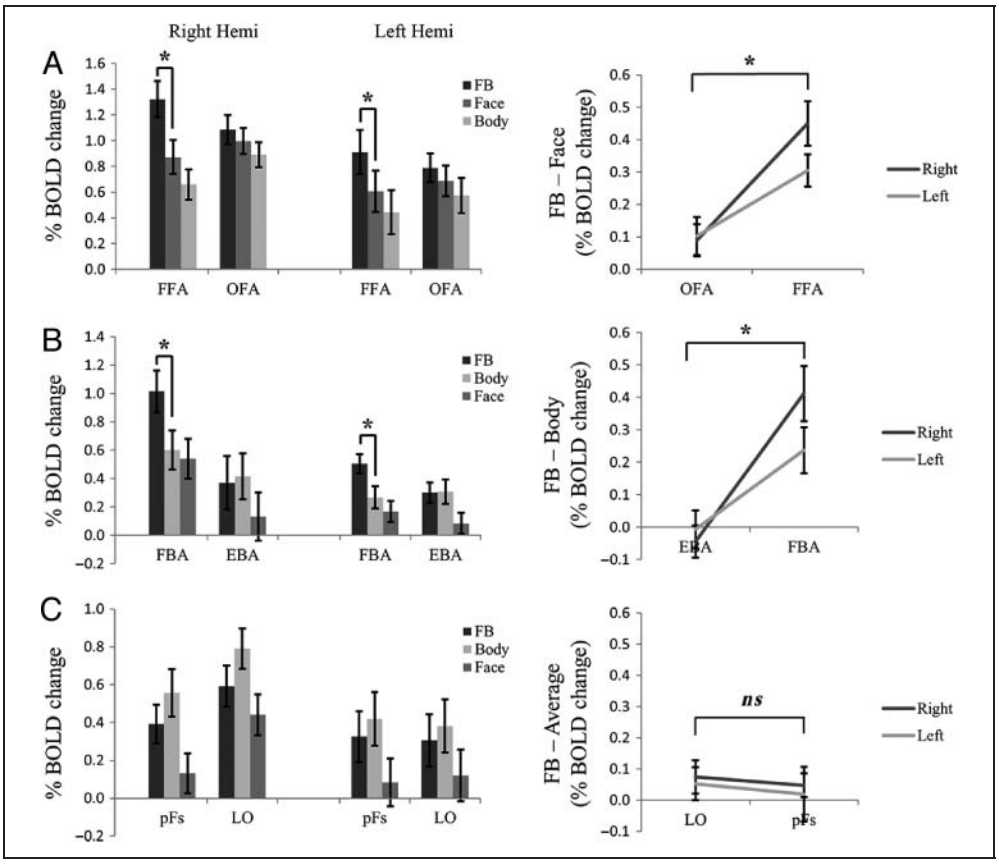

Figure 2. Context effect along the ventral pathway. (A) Left: The response magnitudes for the FB, face, and body conditions in the face-selective areas. The *y* axis indicates the percent BOLD signal change, and error bars indicate ±1 standard error of mean (*SEM*). *$p < .05$. Right: The context effect in the face-selective areas. Dark gray line shows the data from the right hemisphere, and the light gray line indicates the data from the left hemisphere. The *y* axis denotes the context effect, which is indexed by the difference in response magnitude between the FB condition and the face condition. Error bars indicate ±1 *SEM*. Asterisks indicate that the main effect of anatomical location (anterior versus posterior) was significant ($p < .05$). (B) Left: The response magnitudes for the FB, body, and face conditions in the body-selective areas. Right: The context effect in the body-selective areas, indexed by the difference in response magnitude between the FB condition and the body condition. (C) Left: The response magnitudes for the FB, body, and face conditions in the object-selective areas. Right: The context effect in object-selective areas, indexed by the difference between response for the FB condition and average response for the face and body conditions.

was approximately the weighted average response for faces and bodies presented alone, with the weight biased toward faces being 1 (Reddy et al., 2009). The finding that the context effect was larger in the anterior than posterior face-selective regions was further confirmed by a two-way ANOVA on the context effect (i.e., the difference between the FB condition and the face condition), with Anatomical Location (posterior vs. anterior) and Hemisphere (left vs. right) being within-subject factors. The main effect of Anatomical Location was significant, $F(1, 13) = 32.57$, $p < .001$, with the context effect being larger in the FFA than the OFA (Figure 2A, right). The main effect of Hemisphere did not reach significance ($p > .05$). In addition, there was a significant two-way interaction of Location × Hemisphere, $F(1, 13) = 4.87$, $p = .046$. Post hoc pairwise t tests showed that the context effect in the right FFA was larger than that in the left FFA ($t(13) = 2.63$, $p = .02$).

A similar pattern was observed in the posterior (i.e., the EBA) and anterior (i.e., the FBA) body-selective regions along the ventral pathway. We found that response for the FB condition was higher than that for the body condition in both the left ($t(13) = 4.83$, $p < .001$) and right FBA ($t(13) = 3.34$, $p = .005$), whereas there was no significant difference between the FB and body condition in either the left or right EBA (ps $> .05$; Figure 2B, left). A two-way ANOVA of Anatomical Location (FBA versus EBA) × Hemisphere on the context effect (i.e., the difference between the FB condition and the body condition) revealed a significant main effect of Anatomical Location, $F(1, 13) = 39.06$, $p < .001$, with the context effect being larger in the FBA than the EBA (Figure 2B, right). The main effect of Hemisphere did not reach significance, $F(1, 13) = 3.24$, $p = .10$. In addition, there was a significant two-way interaction of Location × Hemisphere, $F(1, 13) = 9.76$, $p = .01$. Post hoc pairwise t tests showed that the context effect in the right FBA was larger than that in the left FBA ($t(13) = 3.30$, $p = .01$). Together, these results suggest that the context effect increases from posterior to anterior along the ventral pathway in both face-selective and body-selective regions, with the largest context effect observed in the right FFA and the right FBA.

However, in our study, the right FFA and right FBA were partially overlapped, with 29% of voxels on average in the right FFA also being body sensitive (Spiridon, Fischl, & Kanwisher, 2006; Peelen & Downing, 2005). Therefore, the context effect observed in the FFA and FBA might result from the overlapping voxels that were sensitive to both faces and bodies. To rule out this possibility, we excluded all overlapping voxels from the right FFA and right FBA, respectively. The results showed the same pattern. That is, the response for the FB condition was higher than that for the face condition in the FFA ($t(12) = 5.35$, $p < .001$) and that for the body condition in the FBA ($t(12) = 3.67$, $p = .003$). More importantly, the context effect was larger in the FFA than the OFA ($t(12) = 5.79$, $p < .001$) and larger in the FBA than the EBA ($t(12) = 4.44$, $p = .001$). Therefore, the context effect observed in the FFA and FBA was likely from neurons exclusively sensitive to faces in the FFA and neurons exclusively sensitive to bodies in the FBA.

In addition, we asked whether the context effect was specific to the regions that are selective to the objects tested. To address this question, we examined the response for the face and body pairs in the posterior (i.e., the LO) and anterior (i.e., the pFs) object-selective regions. In contrast to the face-selective and body-selective regions, the response magnitude for the FB condition lay in the middle of the response magnitude for the face condition and that for the body condition (Figure 2C, left). Indeed, the response magnitude for the FB condition was equal to the average response for faces and bodies presented alone in all object-selective regions (ps $> .05$, Figure 2C, right), in line with a weighted average response with a weight of 0.5 for face and body, respectively. Finally, we examined the PPA, the hypothesized locus for encoding individual objects with strong contextual associations (Bar, 2004), in response to the face and body pairs. We found that the response for the FB condition in both left and right PPA was equal to the average response for faces and bodies presented alone (ps $> .05$). Taken together, the context effect on the representation for the face and body pair in the ventral pathway was apparently specific to the regions where the stimuli were selectively processed.

Faces and Bodies Are Integrated Holistically in the FFA

Our results showed that the face-selective response in the FFA and the body-selective response in the FBA were enhanced by contextually related objects presented simultaneously. However, the mechanism underlying this context effect remained unclear. That is, were the responses for faces and bodies combined linearly or holistically when they were presented simultaneously? To address this question, we compared the response for the FB condition with that for the scrambled condition where faces were presented under bodies in the anterior face-selective and body-selective regions where the context effect were observed. We found that the response for the FB condition was significantly higher than that for the scrambled condition only in the right FFA ($t(13) = 2.29$, $p = .04$), but not in the left FFA or in the FBA (all ts < 1), implying that faces and bodies in the FB condition may be combined holistically in the right FFA and additively in the left FFA and FBA. However, stimuli in the FB condition differed in many aspects from those in the scrambled condition, such as locations in the visual fields and familiarity. To rule out these confounding factors and to directly examine the representation in the right FFA for the face and object pair in the FB condition, we used MVPA to examine whether the synthetic spatial pattern of the face condition and the body condition (see Methods) was more similar to the spatial pattern for the FB condition (i.e., face above body)

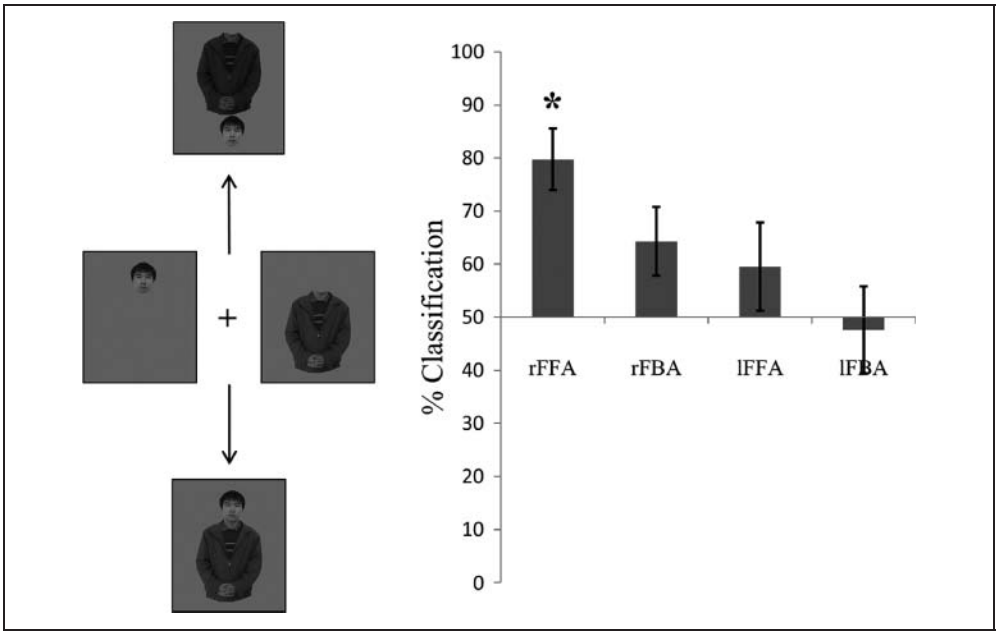

Figure 3. Holistic representation of faces and bodies in the right FFA. The *y* axis indicates the likelihood of the synthetic pattern (the linear combination of the spatial activation patterns for the face condition and the body condition) being classified as the scrambled condition. The value 50% means that the synthetic pattern is equally classified as either the FB condition or the scrambled condition. Values larger than 50% mean that the synthetic pattern is more likely to be classified as the scrambled condition. In other words, the spatial pattern for the FB condition is less likely to be decomposed into spatial patterns for constituent faces and bodies. Error bars indicate ±1 *SEM*. Asterisks indicate the likelihood being significantly different from 50% ($p < .05$).

or to that for the scrambled condition (i.e., face under body).

We reasoned that if faces and bodies in the FB condition were combined holistically, the spatial pattern was not able to be decomposed into the spatial patterns for its constituent faces and bodies; therefore, the synthetic pattern would be more similar to the spatial pattern for the scrambled condition than that for the FB condition. On the other hand, if faces and bodies in the FB condition were linearly combined, the synthetic pattern would be equally classified as either the FB condition or the scrambled condition (i.e., 50%). The MVPA showed that in the right FFA the synthetic pattern was more similar to the spatial pattern for the scrambled condition than the FB condition, making the synthetic pattern more likely being classified as the scrambled condition than the FB condition ($t(13) = 5.10, p < .001$; Figure 3). This finding cannot be explained in terms of locations in the visual field, because locations of faces in the face condition and bodies in the body condition were more similar to those in the FB condition than those in the scrambled condition. Neither can this finding be accounted for by the number of spatially separated entities contained in stimuli, because both the face condition and the body condition contained one entity, but their spatial pattern was more similar to that of stimuli consisting of two spatially separated entities (i.e., the Scrambled condition), rather than stimuli presented as one entity (i.e., the FB condition). Instead, this result suggests that the right FFA may integrate faces and bodies into an inseparable whole. In contrast, in the left FFA and the FBA, the synthetic pattern was equally classified as either the FB condition or the scrambled condition (all ps $> .05$), suggesting that responses for faces and bodies were combined linearly in these regions. In addition, the likelihood of the synthetic pattern being classified as the scrambled condition was significantly larger in the right FFA than in the left FFA ($t(13) = 3.47, p = .004$), the right FBA ($t(13) = 2.06, p = .06$), and the left FBA ($t(13) = 3.88, p = .002$). Finally, in the posterior ROIs (i.e., OFA and EBA) and object-selective ROIs (i.e., LO and pFs) where no context effect was observed, the synthetic pattern was equally classified either as the FB condition or as the scrambled condition (all ps $> .05$).

In addition, the holistic representation of the face and body pair in the right FFA was not accounted for by the overlap between the right FBA and the right FFA. After removing the overlapping voxels between the two regions, the result remained the same. With the overlapping voxels, the likelihood of the synthetic pattern being classified as the scrambled condition in the right FFA was 80%; after removing the overlapping voxels, the likelihood was 81%, and the difference in classification between the right FFA and right FBA was even larger ($t(11) = 3.00, p = .01$). Therefore, it is the voxels sensitive to faces, not those sensitive to bodies, in the right FFA that were engaged in holistic representation of faces and bodies when they were presented simultaneously within a context.

DISCUSSION

In this study, we examined the representation for contextually related multiple objects in the human ventral visual pathway by comparing neural responses when faces and bodies were presented together versus when they were presented in isolation. Two novel results were

found. First, the influence of contextual information on the representation for multiple objects increased along the ventral visual pathway. In particular, in the posterior regions, the response for the face and body pair corresponded to the weighted average responses for faces and bodies presented in isolation. In contrast, the anterior regions encoded the face and body pair in a mutually facilitative fashion, with the response for the pair significantly higher than that for its constituent objects. Second, the spatial activation pattern of the face and object pair in the right FFA was no longer able to be decoupled into the spatial patterns of its constituent objects, suggesting that the contextual information helps integrating the face and body pair into an inseparable whole.

Neurophysiological and fMRI studies on the representation of multiple objects in primate ventral pathway have consistently shown that the response for contextually unrelated multiple objects is approximately the weighted average response for its constituent objects (MacEvoy & Epstein, 2009; Reddy et al., 2009; Zoccolan et al., 2005; Reynolds, Chelazzi, & Desimone, 1999). This finding has often been interpreted in the framework of the response competition model that multiple objects compete for limited processing capacity of the visual system in a mutually suppressive fashion (Desimone & Duncan, 1995). Although neural responses for contextually unrelated objects may suppress each other, our study demonstrated that contextually related multiple objects could interact with each other in a facilitative fashion in the ventral pathway. That is, the neural response for the face and body pair was significantly larger than that for either faces or bodies presented in isolation, which may be the neural basis for the facilitated processing of contextually related objects observed in behavioral studies (e.g., Auckland et al., 2007; Davenport, 2007; Bar & Ullman, 1996; Biederman et al., 1982).

Interestingly, the context effect on the representation of multiple objects was observed in the anterior regions, not in the posterior regions, of the ventral visual pathway. This result fits nicely with a previous study where the lateral fusiform gyrus (possibly the FFA and FBA) responds higher to face and body pairs than either faces alone or bodies alone, whereas the middle occipital gyrus (possibly the EBA and LO) shows an intermediate response to the pairs lying between responses to faces alone and bodies alone (Morris, Pelphrey, & McCarthy, 2006). The insensitivity to the contextual information in the posterior regions is also in line with recent studies showing that the response for multiple objects in the LO corresponded to the averaged response of individual objects regardless of whether they were contextually related or not (MacEvoy & Epstein, 2009, 2011). In a broader picture, our finding supports the idea that object recognition is achieved through a series of hierarchical processing stages, transforming the sensitivity to low-level and local stimulus properties in the posterior portion of the ventral pathway to the sensitivity to more abstract and global properties in the anterior portion (DiCarlo & Cox, 2007; Grill-Spector & Malach, 2004). For example, the OFA and EBA analyze faces and bodies at the level of parts, whereas the FFA and FBA are engaged in analyzing faces and bodies at global level (Zhang, Li, Song, & Liu, 2012; Liu et al., 2010; Taylor, Wiggett, & Downing, 2007; Schiltz & Rossion, 2006). Our result extends this hierarchical structure from the representation of individual objects to the representation of multiple objects, with the anterior regions being more sensitive to the contextual information that binds multiple objects together.

Why are the anterior regions more sensitive to the contextual information among multiple objects than the posterior ones? In daily life, visual scenes typically consist of cluttered and yet contextually related multiple objects. On the one hand, the cluttered objects pose challenges for the visual system to differentiate each individual object. On the other hand, the simultaneously presented objects provide opportunities for the visual system to facilitate the recognition of individual objects with the assistance of the contextual information. The hierarchical encoding of the contextual information in the ventral pathway provides a solution for these two apparently contradictory computational demands in two consecutive steps. The first step is the transformation of representations from two mutually suppressive objects to two mutually facilitative objects. In the posterior regions (the OFA and EBA), the representation of the face and body pair followed the weighted average rule, although the face and body were contextually related and appeared in a single entity (see also MacEvoy & Epstein, 2011). That is, the face and body pair was processed in a mutually suppressive fashion with the response biased toward the preferred object, similar to contextually unrelated objects (Reddy & Kanwisher, 2007; Reddy et al., 2009). The average mechanism for multiple objects has been shown to be helpful for preserving information of individual objects in clutter (MacEvoy & Epstein, 2009, 2011; Reddy et al., 2009; Reddy & Kanwisher, 2007). In contrast, in the anterior regions (the FBA and left FFA), the face and body pair was processed as two mutually facilitative objects in an additive fashion.

The second step is the transformation of representations from two mutually facilitative objects into one integrated object. In the right FFA, the MVPA showed that representation for the face and object pair was resulted from a nonlinear combination of the representations for its constituent faces and bodies, indicating a holistic representation for the face and body pair. This finding is consistent with previous studies showing that the FFA responds to bodies shown under blurred blobs (Cox et al., 2004), and releases from adaptation to identical faces but with different bodies (Andrews et al., 2010) or even to whole persons with varied postures in which faces were fully occluded (Brandman & Yovel, 2010). Furthermore, our MVPA result extends these findings by showing that the context effect observed in the FFA results from a nonadditive combination of responses to faces and bodies, and that

the right hemisphere is dominant in holistic processing (see also Schiltz, Dricot, Goebel, & Rossion, 2010; Jacques & Rossion, 2009; Schiltz & Rossion, 2006). The advantage of multiple objects being integrated into one inseparable object is likely to reduce the perceptual load from processing multiple objects to processing one single object, leading the representation of cluttered objects more efficient. Taken together, our study suggests that the visual system uses a hierarchical representation scheme to process multiple objects in natural scenes: The average mechanism in posterior regions helps retaining information of the constituent objects in clutter, whereas the nonaverage mechanism in the anterior regions uses the contextual information to optimize the representation for multiple objects.

In summary, our study provides the first evidence for a hierarchical scheme to represent multiple objects in the ventral visual pathway. However, several issues remain unresolved that are important topics for future studies. First, it is unclear whether the context effect observed with the face and body pair can be generalized to contextually related objects other than faces and bodies. In our study, we failed to find the context effect in either object-selective pFs or place-selective PPA, suggesting that the contextual information is encoded at the locus where its constituent objects are selectively processed. In other words, place-related multiple objects are likely represented in the PPA (Gronau, Neta, & Bar, 2008) and contextually related general objects are likely represented in the pFs (MacEvoy & Epstein, 2011). Second, the neural source of the contextual effect is unclear. It is possible that our long-term experiences with visual environments provide top–down predictions about object associations and their spatial relations (Song, Bu, Hu, Luo, & Liu, 2010; Song, Bu, & Liu, 2010; Song, Hu, Li, Li, & Liu, 2010; Bar, 2004). Future studies are needed to investigate the interaction of the top–down predictions representing context-based associations and the bottom–up processing of visual stimulus properties in the visual system to represent natural scenes efficiently (Song, Tian, & Liu, 2012; Price & Devlin, 2011).

Acknowledgments

This study was funded by the National Natural Science Foundation of China (91132703, 31100808) and the National Basic Research Program of China (2010CB833903).

Reprint requests should be sent to Jia Liu, Room 405, Yingdong Building, 19 Xinjiekouwai St., Haidian District, Beijing 100875, China, or via e-mail: liujia@bnu.edu.cn.

REFERENCES

Andrews, T. J., Davies-Thompson, J., Kingstone, A., & Young, A. W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience, 30,* 3544–3552.

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review, 14,* 332–337.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Holistic person processing: Faces with bodies tell the whole story. *Journal of Personality and Social Psychology, 103,* 20–37.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617–629.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception, 25,* 343–352.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Brandman, T., & Yovel, G. (2010). The body inversion effect is mediated by face-selective, not body-selective, mechanisms. *Journal of Neuroscience, 30,* 10534–10540.

Chaigneau, S. E., Barsalou, L. W., & Zamani, M. (2009). Situational information contributes to object categorization and inference. *Acta Psychologica (Amst), 130,* 81–94.

Chan, A. W., Kravitz, D. J., Truong, S., Arizpe, J., & Baker, C. I. (2010). Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nature Neuroscience, 13,* 417–418.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4,* 170–178.

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science, 304,* 115–117.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage, 9,* 179–194.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35,* 393–401.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9,* 195–207.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12,* 495–504.

Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face adaptation without a face. *Current Biology, 20,* 32–36.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24,* 187–203.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27,* 649–677.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience, 20,* 371–388.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17,* 273–291.
Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision, 9,* 8.1–8.16.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.
Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282,* 108–111.
Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-level visual object representations are constrained by position. *Cerebral Cortex, 20,* 2916–2925.
Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience, 22,* 203–211.
MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology, 19,* 943–947.
MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience, 14,* 1323–1329.
Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences, 6,* 176–184.
Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research, 616,* 25–29.
Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2006). Occipitotemporal activation evoked by the perception of human bodies is modulated by the presence or absence of the face. *Neuropsychologia, 44,* 1919–1927.
Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93,* 603–608.
Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences, 15,* 246–253.
Reddy, L., & Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Current Biology, 17,* 2067–2072.
Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 21447–21452.
Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19,* 1736–1753.
Rolls, E. T., & Tovee, M. J. (1995). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Experimental Brain Research, 103,* 409–420.
Schiltz, C., Dricot, L., Goebel, R., & Rossion, B. (2010). Holistic perception of individual faces in the right middle fusiform gyrus as evidenced by the composite face illusion. *Journal of Vision, 10,* 25.1–25.16.
Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage, 32,* 1385–1394.
Song, Y., Bu, Y., Hu, S., Luo, Y., & Liu, J. (2010). Short-term language experience shapes the plasticity of the visual word form area. *Brain Research, 1316,* 83–91.
Song, Y., Bu, Y., & Liu, J. (2010). General associative learning shapes the plasticity of the visual word form area. *NeuroReport, 21,* 333–337.
Song, Y., Hu, S., Li, X., Li, W., & Liu, J. (2010). The role of top–down task context in learning to perceive objects. *Journal of Neuroscience, 30,* 9869–9876.
Song, Y., Tian, M., & Liu, J. (2012). Top–down processing of symbolic meanings modulates the visual word form area. *Journal of Neuroscience, 32,* 12277–12283.
Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human Brain Mapping, 27,* 77–89.
Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology, 98,* 1626–1633.
Zhang, J., Li, X., Song, Y., & Liu, J. (2012). The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PLoS One, 7,* e40390.
Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience, 25,* 8150–8164.