Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns

Daniel Kaiser*1, Lukas Strnad*1, Katharina N. Seidl2, Sabine Kastner2, Marius V. Peelen1

1Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto (TN), Italy
2Department of Psychology, Princeton University, Princeton, USA

*D.K. and L.S. contributed equally to this work

Abbreviated title: Person-evoked fMRI responses in fusiform gyrus

Correspondence to M.V.P.:

Marius V. Peelen, Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto(TN), Italy. E-mail: marius.peelen@unitn.it.

Author Contributions:

K.N.S., S.K., and M.V.P. designed research; K.N.S. and M.V.P. performed research; D.K., L.S., and M.V.P. analyzed data; D.K., L.K., and M.V.P. wrote the paper.
Abstract

Visual cues from the face and the body provide information about another’s identity, emotional state, and intentions. Previous neuroimaging studies that investigated neural responses to (bodiless) faces and (headless) bodies have reported overlapping face- and body-selective brain regions in right fusiform gyrus (FG). In daily life, however, faces and bodies are typically perceived together and are effortlessly integrated into the percept of a whole person, raising the possibility that neural responses to whole persons are qualitatively different than responses to isolated faces and bodies. The present study used fMRI to examine how FG activity in response to a whole person relates to activity in response to the same face and body but presented in isolation. Using multivoxel pattern analysis, we modeled person-evoked response patterns in right FG through a linear combination of face- and body-evoked response patterns. We found that these synthetic patterns were able to accurately approximate the response patterns to whole persons, with face and body patterns each adding unique information to the response patterns evoked by whole-person stimuli. These results suggest that whole-person responses in FG primarily arise from the co-activation of independent face- and body-selective neural populations.

Keywords

FFA, FBA, EBA, object perception, category selectivity
INTRODUCTION

The Gestalt psychologists first pointed towards the interactive nature of the grouping of distinct, basic forms into wholes by formulating rules for these integration processes (Wertheimer 1932/1928). Gestalt grouping not only has drastic effects on perception (as for example seen in visual illusions), but also influences neural processing and attentional competition (Spillmann 2006; McMains and Kastner 2010). A large body of literature focuses on low-level effects that are explained via the laws of Gestalt processing (for a recent review, see Wagemans et al. 2012).

More recently, studies have started to investigate the integration of object parts into neural representations of whole objects. Typically, complex objects are composed of parts, which on their own might also be considered objects (e.g., the tires and frame of a car). Nevertheless, at a conceptual level objects are different from the combination of their parts. This raises the question of whether neural representations of complex objects are qualitatively different from the combined neural representations of their constituent parts. Single-cell recording studies of monkey inferotemporal cortex (IT) have provided evidence for nonlinear neural response summation to feature conjunctions (Perrett et al. 1982; Tanaka et al. 1991), but have also provided evidence for linear response summation of features of shapes and faces (Sripati and Olsen 2010; Freiwald et al. 2009). It is not yet clear to what extent these neural underpinnings of feature grouping in IT apply to the grouping of larger parts into complex objects in the human brain.

Here, we investigate a special case of the grouping of components into a whole: the composition of the human body. A person consists of two main parts: a body and a face. These parts are, however, grouped together into the percept of a single person (McArthur and Baron...
Despite the natural and seamless integration of the body and face into a person, imaging studies have typically focused on investigating the neural correlates of either face or body perception (Kanwisher et al., 1997; Downing et al., 2001). It is undoubtedly the case that faces and bodies convey partly different kinds of information. For example, faces are more informative about an individual’s identity, whereas bodies carry more information about actions that the individual performs. Nevertheless, information extracted from faces and bodies is rapidly integrated for judgments about expressed emotions, intentions, and attractiveness (Aviezer et al. 2012; de Gelder 2006; Meeren et al. 2007; van den Stock et al. 2007). One explanation for this integration is that face and body representations are combined into a larger perceptual unit (Aviezer et al. 2012). An alternative explanation is that the integration at the behavioral level is the result of a weighted sum of body and face information, each of which is processed more or less independently in sensory systems (Ekman et al. 1982; Walbott 1998).

At the neural level, the perception of faces and bodies has been associated with selective responses in high-level visual cortex. Interestingly, both faces and bodies evoke selective responses in the right fusiform gyrus (FG) at approximately the same anatomical location (Kanwisher et al. 1997; Peelen and Downing 2005). It has been suggested that the body-selective FG response reflects a post-perceptual recruitment of face-selective mechanisms (Meeren et al. 2013), possibly caused by mental imagery of a face when seeing a blurred face on top of a body (see Cox et al. 2004). However, several studies using headless bodies have provided evidence for separable face and body selectivity in FG (Schwarzlose et al., 2005; Peelen et al., 2006), suggesting that body-evoked responses cannot be fully explained by activity in face-selective neurons. The overlap between face- and body-evoked FG responses partly persists at high spatial resolution (Schwarzlose et al. 2005), raising the possibility that FG contains holistic person
representations. Face- and body-selective responses are also found posterior to FG, with faces activating the “occipital face area” (OFA; Gauthier et al. 2000) and bodies activating the “extrastriate body area” (EBA; Downing et al. 2001). Unlike the face- and body-selective regions in FG, these more posterior regions are anatomically separate from each other, and are thus less likely candidates for housing integrated person representations.

In the present study we asked how FG responses to whole persons relate to FG responses to faces and bodies presented in isolation. Our analysis approach was based on previous studies using intracranial recordings in monkeys and humans (Zoccolan et al. 2005; Zoccolan et al. 2007; Agam et al. 2010), and fMRI in humans (MacEvoy and Epstein 2009; Reddy et al. 2009). These studies have shown that visual cortex activity patterns in response to displays of multiple unrelated objects can be very accurately modeled by a linear combination of response patterns evoked by the individual objects (but see e.g. Heuer and Britten 2002, for a nonlinear relationship). The properties of this linear relationship are not yet clear. While a number of studies suggest the mean as the best operator for combining single-object responses (Zoccolan et al. 2005; MacEvoy and Epstein 2009), others provided evidence for linear models with different weights depending on the strength of the response to the individual objects (Baeck et al. 2013), or even a pure MAX operator, only taking the response of the strongest stimulus into account (Gawne and Martin 2002). By contrast to the present study, previous fMRI studies investigating multiple object responses have used unrelated (pairs of e.g. cars and chairs; MacEvoy and Epstein 2009) and always clearly separate objects, even in the case of action grouping (Baeck et al. 2013).

In this study, we tested how FG responses to whole persons relate to responses to a person’s main parts (face, body) presented in isolation. To do so, we identified a person-selective
FG region in each participant and tested how person-evoked responses in this region relate to responses evoked by faces and bodies shown in isolation. We considered two possible outcomes (Figure 1). First, FG response patterns to whole persons might be well approximated by a linear combination of response patterns to isolated face- and body-evoked response patterns, similar to response averaging previously observed for unrelated object pairs presented simultaneously (MacEvoy and Epstein 2009). This outcome would suggest that FG encodes persons in a part-based manner. Alternatively, FG response patterns to whole persons might be distinct from a linear combination of face- and body-evoked response patterns, reflecting an integrated (holistic) person representation (Figure 1B).
METHODS

Participants. Thirteen participants participated in the study (6 female; mean age: 24.6 years, range: 19-32). All participants had no history of neurological or psychiatric diseases and normal or corrected-to-normal vision. The study was approved by the Institutional Review Panel of Princeton University. Participants gave written consent to their participation.

Stimuli and Paradigm. Stimuli consisted of gray-scale exemplars from six object categories (trees, cars, headless bodies, faces, whole persons, outdoor scenes) depicted on a white background. Blocks of tree and scene stimuli were related to another experiment (Seidl et al. 2012) and are modeled, but not analyzed here. See Figure 2 for stimulus examples. Stimuli subtended approximately 12°x12° visual angle. Stimuli were presented centrally, with a random jitter of 1°. Stimuli were presented on a translucent screen located at the end of the scanner bore using a projector outside the scanner room. Participants could see the screen via a tilted mirror mounted to the head coil.

Participants performed 4 runs. One run consisted of 29 blocks of 14 s each. Blocks 1, 8, 15, 22, and 29 were fixation-only blocks, in which a central fixation cross was presented on a white background. One block of each condition was presented between two fixation epochs. For the first half of the experiment, the order of blocks was determined randomly. The block order for the second half of the run was the mirrored-order of the first half of the run to equate the mean serial position of each condition within the run. Each block contained 10 sequential presentations of exemplars from one category. Stimuli were shown for 400ms and followed by a
white fixation-only screen of equal duration. Subjects performed a one-back repetition detection task. Stimulus repetitions occurred twice during each block.

**Data Acquisition.** Data was acquired using a Siemens Allegra 3T MRI Scanner (Erlangen, Germany) equipped with a standard head coil (Nova Medical). Functional images were obtained in T2*-weighted echo-planar-imaging sequences (TR=2000ms, TE=30ms, 90° flip angle, 34 slices, axial orientation, 3x3x3mm voxel size, 1mm gap, field-of-view: 192, 64x64 in-plane matrix). In addition, a high-resolution T1-weighted scan was acquired using a MPRAGE sequence (TR=2500ms, TE=4.3ms, 8° flip angle, 1mm³ voxel size, 256x256 in-plane matrix).

**fMRI data preprocessing.** Data analysis was performed using the AFNI software package (Cox 1996) and custom-written software in Python, MATLAB and R. For each subject, a high-resolution anatomical image was aligned with the first volume of the first functional run. Functional data were corrected for slice timing and head motion, and transformed into Talairach space using parameters derived from the warping of the high-resolution anatomical image. The time series in each voxel was scaled to a common mean. No spatial smoothing was applied to the data used for MVPA. However, we also created a copy of the data that was spatially smoothed with a Gaussian kernel (4mm full-width half-maximum). This version of the data was used for ROI definition and for univariate analyses.

**Univariate analyses.** In order to ensure independence of data used to define ROIs from data used to compare BOLD responses to the different stimulus categories in these ROIs, the four runs of every participant were separated into two groups. Two runs were used for ROI definition, and
BOLD responses associated with the stimulus categories were then estimated from the other two runs. The results were then averaged over all six possible splits of the four runs into two groups. BOLD responses were estimated using the general linear model (GLM). The GLM included regressors for each stimulus category (created by convolving the boxcar function indicating when each stimulus type was presented with the double-gamma canonical hemodynamic response function), as well as 6 nuisance regressors modeling head motion. For each run, 4 additional regressors were included to account for mean, linear, quadratic, and cubic trends in the data.

Because our hypotheses (Figure 1) were about how person-evoked responses relate to face- and body-evoked responses, we investigated responses in ROIs that were most strongly responsive to the person stimuli and that could thus potentially house person-only representations (Figure 1B). We did not localize face- or body-selective regions, as these regions are less likely to house person-only neurons. Two right-hemisphere person-selective ROIs, one in FG and one in lateral occipitotemporal cortex (LOTC), were functionally defined using an objective (automated) method, selecting all voxels that showed significant person selectivity (person > cars; p<0.001, uncorrected) within box-shaped FG (30 < x < 50; 55 < y < 30; 25 < z < 10; Peelen et al. 2009) and LOTC regions (35 < x < 60; 80 < y < 55; 10 < z < 15; Peelen et al. 2009). The pFG ROI contained on average 38 voxels (mean Talairach coordinates: x = 41.0, y = -45.4, z = -15.5), and the pLOTC ROI contained on average 67 voxels (mean Talairach coordinates: x = 46.5, y = -68.6, z = 3.6). Univariate ROI analyses were based on the mean value of the regressor coefficients (betas) for each stimulus category, estimated from the second half of the data in the split and extracted from the regions defined above. In addition to the functionally defined regions, we also anatomically defined an occipital cortex (OC) region corresponding to primary
visual cortex (V1) using the Talairach atlas (Brodman area 17). For congruency with the functional ROIs, only right-hemisphere OC was analyzed.

Multivoxel pattern analysis (MVPA). The procedure for ROI definition for MVPA was similar to that used for the univariate analysis. However, for MVPA, we used all four runs to define ROIs. Participants whose FG ROI contained less than 15 voxels were excluded from the analyses. Two participants were excluded on the basis of this criterion. For MVPA, each condition was modeled by a separate regressor for each run. All other parameters were identical to those of the GLMs used for univariate analyses. Multi-voxel patterns for subsequent analyses were created from the beta values of the category regressors. A single pattern would consist of the beta values of a single regressor extracted from all voxels of a given ROI.

MVPA was performed in a split-half manner. All analyses were performed for every possible two-way split of the four runs (six possible splits altogether) and these results were combined by averaging. In each half, a synthetic person pattern was computed as the average of the face pattern and the body pattern. The average was taken following previous work on object pairs (MacEvoy and Epstein 2009). To assess the similarity of response patterns, we computed correlations between all possible pairs of patterns A - B such that A comes from one of the halves, and B from the other. Correlations between pairs that included the same categories stemming from different runs were averaged (e.g., A_face - B_body and A_body - B_face correlations were averaged). For present purposes, only the correlations involving face, body, person, and synthetic conditions were relevant. Before group-level statistical analysis, all correlations calculated on the single-subject level were Fisher transformed.
To test for the optimal weights for combining face and body patterns to approximate the person pattern, we obtained the values of the coefficients maximizing the correlation between a linear combination of the face and body patterns with the person pattern through a simple optimization procedure. Since for the calculation of correlation only the relative magnitude of the coefficients associated with the face and body patterns matters, we imposed a further constraint that the two coefficients must add up to one. Note that, therefore, this analysis only reveals information about the relative contribution of face and body patterns, but not about their absolute contributions. We denote the face coefficient alpha, and body coefficient beta, but note that beta is determined by alpha as beta = (1 - alpha). The ‘synthetic’ pattern approximating the person pattern was thus equal to alpha * face pattern + beta * body pattern. To compute the coefficients, data from all four runs were first combined by averaging. We then tested, separately for each participant, for the value of alpha yielding the highest correlation between the synthetic face-body pattern and the person pattern. Note that this analysis was done separately from the analyses involving the split-half correlations between the conditions. In those analyses, the alpha was always set to 0.5.
RESULTS

Univariate analyses. We first compared the mean BOLD signal elicited by faces, bodies, and persons within person-selective FG (pFG), person-selective LOTC (pLOTC), and anatomically-defined OC (Figure 3). In each of the three ROIs, there was a main effect of stimulus category (all $F[3,36] > 8.04, p < 0.001$). In pFG, persons elicited higher responses than faces, bodies, and cars (all $t[12] > 5.00, p < 0.001$). Faces and bodies each elicited higher responses than cars (both $t[12] > 6.70, p < 0.001$), while the difference between face- and body-evoked responses did not reach significance ($t[12] = 2.09, p = 0.058$). pLOTC responses were higher for faces, bodies and persons as compared to cars (all $t[12] > 3.00, p < 0.011$). However, in contrast to pFG, responses in pLOTC were higher for persons and bodies than for faces (both $t[12] > 8.60, p < 0.001$) and the region did not differentiate between bodies and persons ($t[12] = 0.60, p = 0.56$). The lack of difference between the body and person categories suggests that pLOTC is only sensitive to body information present in both body and person stimuli, and that it is not influenced by the additional face information present only in person stimuli (see Morris et al. (2006) for evidence that the presence of a face may even reduce activity in body-selective cortex). In OC, the face condition led to lower responses than all other conditions (all $t[12] > 6.73, p < 0.001$), with no difference between bodies, persons and cars (all $t[12] < 0.92, p > 0.38$). The low response to faces is likely due to the smaller retinal size of the face condition relative to the other conditions (Figure 2).

MVPA. In the multivoxel pattern analysis, we aimed to test hypotheses about how person-evoked response patterns in pFG relate to face- and body-evoked response patterns (Figure 1). In order
to assess the degree of specificity of the effects found in pFG, we also analyzed results in pLOTC and OC.

Previous studies investigating neural responses in the ventral stream to displays consisting of multiple distinct objects report that these responses (measures both by single cell recordings and by multi-voxel fMRI patterns) can be very accurately approximated as a simple average of responses to the individual constituent objects (Zoccolan et al. 2005; MacEvoy and Epstein 2009). Since the face and the body could be thought of as constituent objects of a person image, we hypothesized that it may be possible to approximate the person pattern in FG as a linear combination (such as the average) of the face and body patterns. To test directly for the optimal coefficients for approximating person patterns, we computed the correlation between synthetic and person patterns as a function of the coefficients (Figure 4). For every subject and all 3 ROIs, we used an optimization procedure to estimate the values of the linear coefficients for which this correlation is the highest (see Methods for details). We then tested whether, at the group level, these coefficients are different from 0.5. In pFG, this difference was not significant ($t[10] = 0.034, p = 0.74$), indicating that the linear combination of the face and body patterns that approximates the person pattern most successfully was comprised of equal weights for faces and bodies. By contrast, in both pLOTC and OC the body coefficient was significantly greater than 0.5 (pLOTC: $t[10] = 6.60, p < 0.001$; OC: $t[10] = 4.35, p = 0.0014$).

By combining the face and body patterns through a linear combination with the coefficients equal to 0.5, we next created synthetic multivoxel patterns approximating the person patterns in FG for every participant. We then compared multivoxel pattern correlations between these synthetic patterns and the patterns associated with faces, bodies and whole persons (Figure 5). We found that the synthetic pattern was better in approximating the person pattern than were
the individual face or body patterns. In particular, we tested whether the person pattern was reliably more similar to the synthetic pattern than to the face or body pattern alone. A one-way ANOVA including the person-synthetic, person-face, and person-body correlations yielded a main effect of category pair (F[2, 20] = 5.11, p = 0.016). In follow-up tests, we found a positive difference between person-synthetic and person-face correlations (t[10] = 4.42, p = 0.0013), and also between person-synthetic and person-body correlations (t[10] = 3.22, p = 0.0091). Therefore, the high degree of similarity between the synthetic pattern and the person pattern can be ascribed to the combination of distinct pieces of information contained in the individual face and body patterns.

Additionally, we tested whether the synthetic pattern could at all be distinguished from the person pattern in pFG. We computed a one-way ANOVA including the person-person, person-synthetic, and synthetic-synthetic correlations. This ANOVA did not yield a significant main effect of category pair (F[2, 20] = 1.36, p = 0.28), suggesting that the synthetic pattern is statistically indistinguishable from the person pattern. To underline this point, we also directly tested for a difference between the person-person and person-synthetic correlations (t[10] = 1.14, p = 0.28), and between the synthetic-synthetic and person-synthetic correlations (t[10] = 1.04, p = 0.32), neither of which yielded a significant result. Taken together, these analyses indicate that the person pattern in pFG can be accurately modeled as a linear combination of the face and body patterns, with equal weights for both categories.

These results were specific to pFG. Unlike in pFG, the person patterns in pLOTC and OC were equally correlated with the body pattern as with the synthetic pattern (LOTC: t[10] = 1.25, p = 0.24; V1: t[10] = 0.95, p = 0.37), suggesting that the synthetic pattern did not contain any meaningful information but the one already included in the body pattern. Furthermore, unlike in
pFG, the synthetic patterns were distinguishable from the person patterns: While the synthetic-
synthetic correlation was equal to the synthetic-person correlation (pLOTC: t[10] = 0.083, p =
0.93; OC: t[10] = 0.36, p = 0.73), the person-person correlation was significantly higher than the
person-synthetic correlation (pLOTC: t[10] = 2.71, p = 0.022; OC: t[10] = 5.64, p < 0.001).
DISCUSSION

In the present study, we tested how fusiform gyrus (FG) responses to whole persons relate to responses to a person’s main parts (face, body) presented in isolation. In a univariate analysis, we found that person-selective FG was more responsive to pictures of whole persons than to pictures of faces and bodies shown in isolation. In contrast to FG, we observed comparable responses to whole persons and isolated bodies in person-selective LOTC and anatomically-defined occipital cortex (OC). Interestingly, multivoxel pattern analysis showed that FG response patterns to whole persons could be accurately modeled by a linear combination of response patterns evoked by isolated faces and bodies, with equal weights for both categories. Taken together, these results suggest that FG encodes persons in a part-based rather than an integrated (holistic) manner.

Distinct representations of faces and bodies in FG

The finding that response patterns to faces and bodies contributed unique information to whole person-evoked response patterns provide further evidence that face and body representations in FG are separable, thus arguing against the possibility that body-selective FG responses are fully accounted for by late activation of face-selective neural populations (Meeren et al. 2013). These conclusions are in line with several previous findings. For example, high-resolution fMRI allows for isolating FG voxels that are only selective for faces or for bodies (Schwarzlose et al. 2005, 2008), although an overlap between face-and body-selective responses persists even at high spatial resolution. Other studies have used MVPA in FG to show that body-selective response patterns, but not face-selective response patterns, correlate with response patterns evoked by point-light biological motion displays of whole-body actions (Atkinson et al.
2012; Peelen et al. 2006) and with activity evoked by emotional (versus neutral) body expressions (Peelen et al. 2007). Developmental work has shown that face- and body-selective fusiform regions develop along different trajectories (Peelen et al. 2009). In the monkey, adjacent and partly overlapping face- and body-selective regions have also been identified using fMRI (Pinsk et al. 2005, 2009; Tsao et al. 2003). Single-cell recording studies in monkey IT have found neural populations that respond selectively to either faces (Desimone et al. 1984; Tsao et al. 2006) or body parts (e.g., hands; Gross et al. 1972). Furthermore, the pattern of responses across hundreds of IT neurons showed distinct clusters for faces, hands, and bodies (Kiani et al. 2007). Together with the present findings, these results suggest that representations of faces and bodies in FG are likely to be largely distinct at the neuronal level. This raises the question of why faces and bodies are represented so close together. One possibility is that the consistent co-occurrence and mutual relevance of faces and bodies results in these categories being represented nearby as this minimizes axon length and thus optimizes efficient interactions between these representations (e.g. Aflalo and Graziano 2011; Van Essen 1997). Another possibility is that innate anatomical connectivity patterns between visual cortex and downstream areas involved in social cognition are strongest in the part of FG where representations of faces and bodies are located, thus optimizing this part of visual cortex for representing socially-relevant stimuli (Simmons and Martin 2012; Mahon and Caramazza 2011). Finally, it has been proposed that neurons in FG may be well suited to perform particular computations such as subordinate-level discrimination (Tarr and Gauthier 2000), which is highly relevant for both face and body processing.
No evidence for holistic person representations in FG

The present results showed that response patterns to whole persons could be accurately approximated by a linear combination of response patterns to faces and bodies shown in isolation. This result is most in line with scenario A in Figure 1, in which person-evoked responses are accounted for by concurrent activity in face- and body-selective neural populations, thus without assuming additional populations tuned preferentially or exclusively to whole persons. It should be noted, however, that the absence of evidence for holistic person representations in FG does not conclusively proof that such populations do not exist in this region. First, it may be that such populations are largely outnumbered by populations of face- and body-selective neurons and are therefore hard to detect. For example, a study recording from cells in monkey STS (Wachsmuth et al. 1994) found that, of the cells responding to whole persons, only 17% responded exclusively to whole persons (the other cells responded to either the face or the body, or to both). Second, person-selective neurons could be evenly distributed across voxels, unlike face- and body-selective neurons. In this case, these neurons would not form a pattern across voxels and therefore would not be captured by MVPA.

Our conclusions may appear to contradict those of a recent paper that reported person-selective fMRI adaptation effects in right fusiform gyrus (Schmalzl et al. 2012). In this study, participants were shown two subsequently presented person pictures. The second picture could either be an exact repetition of the first picture, it could be a picture of a different person, or it could be a partial repetition of the first picture, with either the face or the body being that of a different individual (with the other part repeating). Results showed that fMRI adaptation in right FG to exact-picture repetitions was larger than the sum of the adaptation effects to face-only and body-only repetitions. The interpretation of these results was that FG contains representations of
whole persons in addition to representations of faces and bodies; if the whole person repeats, this would adapt neurons coding for the whole person in addition to neurons coding for the face and body, thereby explaining the superadditive adaptation effects. Alternatively, however, the partial change conditions (in which only the body or the face changed) may have given higher activity (i.e., less adaptation) partly because it attracted more attention than the condition in which the exact same picture was repeated. Differences in attention and expectation across conditions may complicate the interpretation of fMRI adaptation studies (e.g., Henson et al., 2003; Summerfield et al., 2008; Mur et al., 2010), and such effects may partly account for the results of the Schmalzl et al. study: First, a change of either a face or a body (with the other part unchanged) is highly unnatural and may thus increase overall attention and/or may direct attention to the part that changed thereby drawing attention away from the repeated part. Second, the task of the participants was to detect specific target faces and bodies occurring on 10% of trials. When the exact same picture repeated, participants could quickly establish, based on low-level cues, that the response (target or no target) to the second picture was the same as the response to the first picture, thus no longer requiring the identification of the face or body. More generally, the different conclusions reached by the Schmalzl et al. (2012) study and our study could be due to differences in methodological approaches (for differences between MVPA and fMRI adaptation results see e.g. Drucker and Aguirre 2009; Epstein and Morgan 2012). Another important difference between our study and the study of Schmalzl et al. (2012) is that their experiment was focused on finding adaptation in neurons coding for specific identities of faces, bodies, and whole persons (whole persons were presented on every trial). By contrast, our study looked at the distribution of category-level selectivity to faces, bodies, and their combined presentation.
The absence of evidence for holistic person representations in FG in our study raises the question of whether information from these stimuli are integrated elsewhere in the brain, and if so where. Candidate regions for person representations that are abstracted away from visual cues are the medial prefrontal cortex and posterior superior temporal cortex. These regions have been implicated in social cognition, mentalizing, and theory-of-mind (Adolphs 2009; Gallagher and Frith 2003; Mitchell et al. 2002; Saxe and Kanwisher 2003), and have been shown to contain representations of perceived mental states that generalize across face and body cues (Peelen et al. 2010). Future work is needed to test whether these regions may underlie the body-face integration effects observed in behavioral studies (Aviezer et al. 2012; de Gelder 2006; Meeren et al. 2007; van den Stock et al. 2007; Ghuman et al. 2010), or whether these effects instead arise as a result of rapid interactions between distinct but nearby representations of faces and bodies in FG.
REFERENCES


Morris JP, Pelphrey KA, McCarthy G. Occipitotemporal activation evoked by the perception of human bodies is modulated by the presence or absence of the face. *Neuropsychologia* 44: 1919–1927, 2006.


FIGURE LEGENDS

Figure 1: Two hypothetical scenarios for the nature of whole-person responses in pFG. A) Person perception activates distinct but intermixed face- and body-selective neural populations. B) Person perception activates, in addition to face- and body-selective neurons, neurons that are exclusively responsive to whole persons. These scenarios make similar predictions for overall BOLD responses, but different predictions for multivoxel pattern correlations. Specifically, in scenario A, person-evoked responses are accurately modeled by a linear combination of face- and body-evoked responses, whereas in scenario B this would not be the case.

Figure 2: Example stimuli. Whole persons were always a compound of one of the faces and one of the bodies presented. Stimuli were presented for 400ms each in blocks of 14s.

Figure 3: Univariate results. BOLD signal changes in response to stimuli of different categories, averaged over all voxels in each ROI. In the pFG ROI, the person stimuli elicited the highest response, while faces and bodies evoked comparable responses. In pLOTC, by contrast, the responses to the person stimuli did not differ significantly from the responses to the body stimuli, which were both higher than responses to faces. OC showed weaker responses for faces, because of their overall smaller retinal sizes. Error bars indicate standard error of the mean.

Figure 4: Correlations between the person pattern as a function of different linear combinations of face and body patterns in pFG, pLOTC and OC. Single subject estimates for the optimal coefficients are represented by the dots in the three left panels, with their mean
and standard error of the optimal estimate represented on the top; the right panel shows the mean results across participants for all regions. For pFG, the highest correlation was obtained when the weights for face and body patterns were equal. By contrast, the person patterns in pLOTC and OC were best approximated by a synthetic pattern that consisted almost exclusively of the body pattern. All correlations shown were Fisher transformed.

**Figure 5: Results of multivoxel pattern analysis.** Correlations in pFG, pLOTC and OC between the multivoxel patterns associated with the face, body, and person stimuli, and the synthetic patterns created by averaging face- and body-related patterns. The pFG results best match scenario A in Figure 1. All correlations shown were Fisher transformed.
Possible distributions of neurons in one example voxel in FG

A
Independent Face- and Body-Populations

B
Person-only Neurons

Predicted BOLD-responses from the example voxel

Predicted multi-voxel pattern correlations