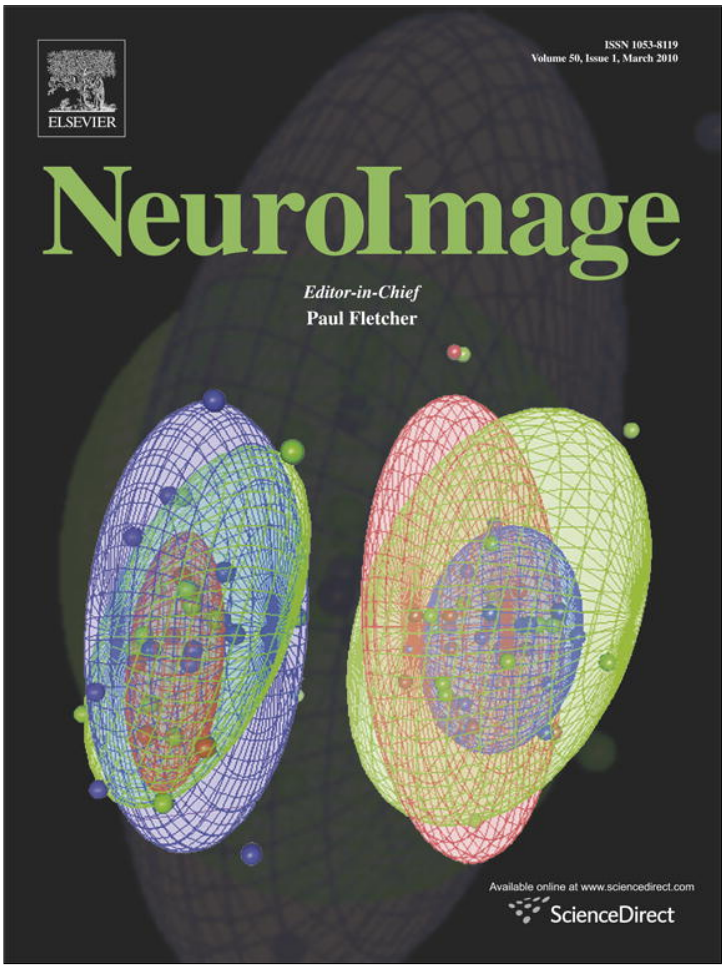
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.




This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

NeuroImage 50 (2010) 56–71

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Evaluating functional localizers: The case of the FFA

Marc G. Berman [a,b,*], Joonkoo Park [a], Richard Gonzalez [a], Thad A. Polk [a,c], Amanda Gehrke [a], Scott Knaffla [a], John Jonides [a]

[a] Department of Psychology, University of Michigan at Ann Arbor, Ann Arbor, MI, USA
[b] Department of Industrial and Operations Engineering, University of Michigan at Ann Arbor, Ann Arbor, MI, USA
[c] Department of Electrical Engineering and Computer Science, University of Michigan at Ann Arbor, Ann Arbor, MI, USA

ARTICLE INFO

*Article history:*
Received 4 August 2009
Revised 27 October 2009
Accepted 4 December 2009
Available online 16 December 2009

ABSTRACT

Functional localizers are routinely used in neuroimaging studies to test hypotheses about the function of specific brain areas. The specific tasks and stimuli used to localize particular regions vary widely from study to study even when the same cortical region is targeted. Thus, it is important to ask whether task and stimulus changes lead to differences in localization or whether localization procedures are largely immune to differences in tasks and contrasting stimuli. We present two experiments and a literature review that explore whether face localizer tasks yield differential localization in the fusiform gyrus as a function of task and contrasting stimuli. We tested standard localization tasks–passive viewing, 1-back, and 2-back memory tests–and did not find differences in localization based on task. We did, however, find differences in the extent, strength and patterns/reliabilities of the activation in the fusiform gyrus based on comparison stimuli (faces vs. houses compared to faces vs. scrambled stimuli).

© 2009 Elsevier Inc. All rights reserved.

## Introduction

The brain is organized in a distributed fashion in which networks of brain regions interact, representing the underlying machinery of thought and action. Yet even with such distributed functioning, there are modular features to brain organization. An outstanding example is the Fusiform Face Area (FFA), an area in the fusiform gyrus that is heavily involved in the processing of faces (Sergent et al., 1992; Kanwisher et al., 1997). The FFA has typically been defined as the area in the fusiform gyrus that responds more to faces than to comparison stimuli, such as houses or other objects (Kanwisher et al., 1997; Puce et al., 1995). While the exact role that the FFA plays in face processing is not entirely clear (Kung et al., 2007; Gauthier et al., 2000a–c), this area activates very robustly for any task involving faces as stimuli though it has been known to activate for other stimuli as well (see Gauthier et al., 2000a,b; Haxby et al., 2001).

One common method used to define any functional region such as the FFA is to conduct an independent experiment to localize this region for each individual participant (Saxe et al., 2006; Friston et al., 2006; Friston and Henson, 2006). Once defined, this region can then be used to analyze a task of interest in a constrained way. For instance, a researcher may be interested in a short-term memory task involving faces, wanting to explore how FFA activation changes with increased memory load. In order to define the FFA, this researcher may first have participants perform a simple task on faces (such as passive viewing) to define the FFA for each participant in order to use as this Region-of-Interest (ROI) to analyze the memory task of interest (Druzgal and D'Esposito, 2001 employed this very procedure).

Some researchers have questioned whether defining functional ROIs in this way is appropriate. For example, Friston and colleagues (2006) criticized this procedure claiming that the FFA defined by a localizer task may differ based on the functional context of the task (e.g., the task conditions and comparison stimuli). Therefore, Friston et al. (2006) argued that it is better to embed the localizer task in a factorial design. Similarly, Poldrack (2007) pointed out that the distribution of activations over many voxels in a ROI may differ based on the task demands (Haxby et al., 2001; Poldrack, 2006). Related to this point, the independent localizer approach will miss potential interactions between the localizer task and the main experimental task of interest. For example, if the localizer task has participants passively view images of faces vs. scrambled images of faces, and the main task of interest involves an *n*-back memory test of faces and houses, the independent localizer approach will not provide the ability to look at the interaction of task (passive viewing and *n*-back) and comparison stimulus (scrambled face and house). As such, important results could be missed and definitions of the ROIs may be inappropriate if localization changes based on these experimental manipulations.

While having a factorial design seems ideal, such a design requires more trial combinations than might be available in limited scanning time. Additionally, the functional localizer approach can expedite tailored definitions of specific functional areas, and provide entirely

* Corresponding author. Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043, USA.
*E-mail address:* bermanm@umich.edu (M.G. Berman).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.12.024

Author's personal copy



independent data to define ROIs in a task of interest (Saxe et al., 2006). Much of the critique of functional localizers hinges on the idea that the task that participants perform or the stimuli against which a stimulus of interest is compared may alter the apparent localization in FFA (Friston et al., 2006). This is an empirical question, which we test in this paper. We ask this question with reference to a frequently studied brain module, the FFA. Thus, we ask: To what extent do variations in typical tasks and stimuli lead to changes in the localization of FFA?

Duncan et al. (2009) have explored a similar issue for word and object processing. The authors had participants perform a 1-back memory test on words and objects in order to localize word-sensitive and object-sensitive areas in Inferior Temporal Cortex (IT). The authors found a surprising lack of consistency in the voxels activated in IT for word and object perception for the 1-back memory task. When individual participants performed the memory test in run 1 of the experiment, fewer than 20% of the same voxels involved in word or object perception were activated in run 2 of the same task with the same stimuli at conservative statistical thresholds. Duncan et al. (2009) drew the conclusion that the use of functional localizers may not be appropriate based on the lack of consistency in activated voxels from run 1 to run 2 of the same task. Kung et al. (2007) have data and results similar to Duncan et al. (2009) for an *n*-back memory task on faces and greebles. While very informative, these findings speak more to the reliability of fMRI data over time, and less to whether functional

**Table 1**
Here we list the studies that were used for our literature review. We provide the left and right localized coordinates for each study (in Talairach space) as well as the comparison stimuli used to contrast the faces and the localizer task.

| Study | Right | | | Left | | | Task | Comparison stimuli |
|---|---|---|---|---|---|---|---|---|
| | *x* | *y* | *z* | *x* | *y* | *z* | | |
| Aguirre et al., 1999 | 37 | −62 | −17 | n/a | n/a | n/a | passive | Multiple objects |
| Andrews and Ewbank, 2004 | 44 | −58 | −22 | −46 | −61 | −27 | 1-back | Multiple objects |
| Andrews and Schluppeck, 2004 | 44 | −59 | −15 | n/a | n/a | n/a | passive | Multiple objects |
| Avidan et al., 2003 | 34 | −44 | −15 | −38 | −48 | −15 | 1-back | Multiple objects |
| Caldara et al., 2006 | 40 | −50 | −17 | −38 | −52 | −17 | 1-back | Multiple objects |
| Carlson et al., 2006 | 41 | −51 | −6 | −43 | −58 | −9 | passive | Multiple objects |
| Chen et al., 2007 | 28 | −57 | −14 | n/a | n/a | n/a | passive | Scramble |
| Clark et al., 1997 | 37 | −55 | 10 | n/a | n/a | n/a | 1-back | Scramble |
| Courtney et al., 1997 | 31 | −60 | −20 | −32 | −60 | −16 | 1-back | Scramble |
| Dove et al., 2008 | 42 | −51 | −14 | n/a | n/a | n/a | 1-back | Multiple objects |
| Downing et al., 2006 | 37 | −46 | −15 | n/a | n/a | n/a | passive | Multiple objects |
| Druzgal and D'Esposito, 2001 | 41 | −56 | −25 | n/a | n/a | n/a | *n*-back | Multiple objects |
| Eger et al., 2005[a] | 45 | −45 | −24 | −45 | −45 | −23 | symmetry judgment | Scramble |
| Eger et al., 2004 | 40 | −54 | −24 | −41 | −57 | −23 | passive | Scramble |
| Epstein et al., 2006 | 42 | −52 | −15 | −44 | −54 | −17 | 1-back | Multiple objects |
| Fang et al., 2007 | 33 | −40 | −14 | n/a | n/a | n/a | passive | Multiple objects |
| Ganel et al., 2005 | 36 | −51 | −16 | n/a | n/a | n/a | passive | Multiple objects |
| Gathers et al., 2004 | 42 | −56 | −22 | n/a | n/a | n/a | passive | Multiple objects |
| Gauthier et al., 2000a | 35 | −49 | −8 | −35 | −56 | −6 | 1-back | Single object |
| Gauthier et al., 2000b | 38 | −50 | −7 | −38 | −56 | −6 | passive | Multiple objects |
| Gauthier et al., 1999 | 41 | −55 | −10 | −40 | −46 | −12 | passive | Multiple objects |
| George et al., 1999 | 52 | −60 | −24 | −42 | −68 | −22 | recognition | Single object |
| Gilaie-Dotan et al., 2009 | 29 | −47 | −14 | −36 | −54 | −18 | 1-back | Multiple objects |
| Grill-Spector et. al., 2004 | 39 | −40 | −16 | −37 | −42 | −16 | passive | Multiple objects |
| Hadjikhani and de Gelder, 2003 | 35 | −55 | −14 | −34 | −55 | −13 | passive | Scramble |
| Henson and Mouchlianitis, 2007 | 42 | −44 | −15 | −42 | −50 | −13 | name/attend | Single object |
| Horovitz et al., 2004 | 35 | −53 | −13 | n/a | n/a | n/a | passive | Single object |
| Iacoboni et al., 2004[b] | 36 | −82 | −22 | −20 | −72 | −12 | passive | Single object |
| Iidaka et al., 2004 | 60 | −40 | 2 | −46 | −34 | −6 | 1-back | Single object |
| Ishai et al., 2006 | 40 | −57 | −17 | −42 | −58 | −18 | passive | Scramble |
| Kanwisher, McDermott and Chun, 1997 | 40 | −55 | 10 | n/a | n/a | n/a | passive | Single object |
| Kesler-West et al., 2001[c] | 39 | −47 | −12 | −39 | −53 | −12 | passive | Scramble |
| Lehmann et al., 2004 | 32 | −46 | −18 | −43 | −56 | −16 | passive | Single object |
| Leibenluft et al., 2004 | 39 | −59 | −15 | −39 | −61 | −17 | 1-back | Single object |
| Loffler et al., 2005 | 37 | −44 | −15 | n/a | n/a | n/a | passive | Single object |
| Maurer et al., 2007 | 44 | −60 | −24 | −48 | −60 | −28 | passive | Multiple objects |
| O'Craven and Kanwisher, 2000 | 37 | −36 | −18 | −37 | −39 | −15 | passive | Single object |
| Peelen and Downing, 2005 | 41 | −47 | −17 | −38 | −46 | −16 | passive | Multiple objects |
| Pourtois et al., 2005a[d] | 48 | −57 | −33 | −45 | −51 | −27 | 1-back | Single object |
| Pourtois et al., 2005b | 48 | −57 | −33 | −45 | −51 | −27 | 1-back | Single object |
| Puce et al., 1995 | 31 | −60 | −20 | −32 | −60 | −16 | passive | Scramble |
| Puce et al., 1996[e] | 30 | −54 | −20 | −38 | −59 | −21 | passive | Single object |
| Puce et al., 1996 | 31 | −54 | −21 | −39 | −54 | −23 | passive | Single object |
| Reddy et al., 2007 | 38 | −50 | −13 | n/a | n/a | n/a | passive | Single object |
| Reinholz and Pollmann, 2005[f] | 37 | −50 | −12 | −41 | −50 | −14 | passive | Single object |
| Rotshtein et al., 2007 | 43 | −45 | −20 | n/a | n/a | n/a | 1-back | Multiple objects |
| Serences et al., 2004[g] | 33 | −57 | −14 | n/a | n/a | n/a | passive | Single object |
| Von Kriegstein et al., 2005 | 42 | −45 | −24 | −36 | −45 | −30 | passive | Multiple objects |
| Yi et al., 2006 | 39 | −48 | −15 | −39 | −48 | −15 | 1-back | Single object |
| Zhang et al., 2008 | n/a | n/a | n/a | −48 | −42 | −2 | 1-back | Multiple objects |

[a] Averaged the coordinates of the two left peaks.
[b] The stimulus was a movie clip.
[c] Selected the larger FFA clusters.
[d] We believe the same subjects were used in these tasks, therefore we only used the localizer points once in our meta-analysis.
[e] For Puce et al., 1996, the same study was used, but there were two different localizers for different contrasting stimuli (letters and textures).
[f] Averaged the coordinates of the two left peaks.
[g] Averaged the coordinates of the two right peaks.

Author's personal copy



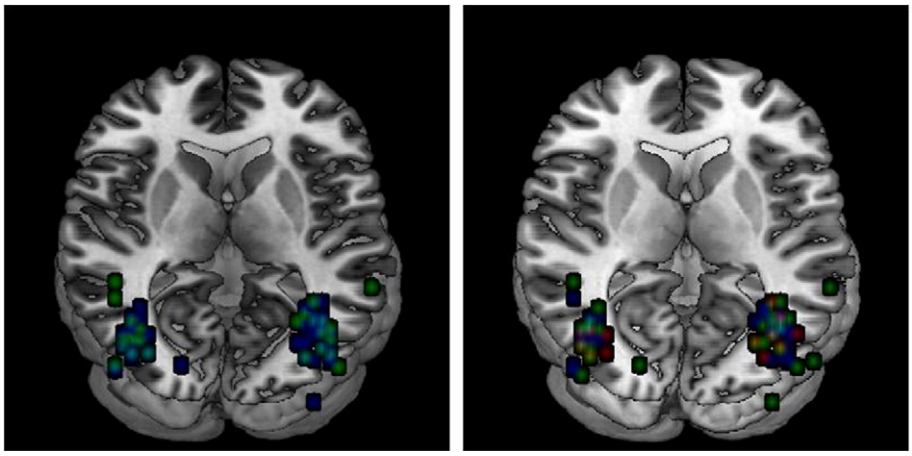


**Fig. 1.** Peaks for different localizer task types (left). 1-back are in green, passive-viewing in blue. On the right are the peaks for the different contrasting stimuli. Peaks for different contrasting stimuli (right). Multi-objects are in blue, scrambles in red, and single-objects in green.

localizers are appropriate overall because we do not know how within-task and between-task consistencies compare. To define the appropriateness of functional localizers, one would like to compare variance in activated voxels both within- and between-tasks.

We address this issue with data from two empirical studies and a literature review. In our literature review, we explore reports that used various tasks to localize the FFA, and we composed 3-dimensional confidence intervals (which formed ellipsoids) to compare localization based on different localizer tasks and contrasting stimuli. In the two empirical studies, we manipulated the task and comparison stimuli and compare consistency (in the sense of Duncan et al., 2009, and Kung et al., 2007), reliability of activation patterns (as measured with Pearson correlations), extent and magnitude of activation both within- and between-tasks, thereby testing how task context affects activation in the region of interest that we chose, the FFA. Our empirical and literature review findings suggest that while there is more reliability for an ROI within a task, between-task reliabilities are impressively high, independent of task and comparison stimuli. Additionally, we address Friston et al.'s (2006) concerns with the functional localizer approach and indicate where we believe functional context matters and how task and comparison stimuli may interact to produce differences in FFA measurements. We focus on the FFA to constrain our analysis, though similar analyses could be performed on other functional ROIs as well.

## Literature review

To examine the effect of task and comparison stimuli on localization in the FFA, we reviewed 339 papers that studied the FFA, coding how those studies localized the FFA. Our search included all dates up until April, 2008 using Web of Science (copyright Thomson Reuters, 2009). Our searches included the keywords: *fMRI, Face, Fusiform, FFA, and Localizer*, in different combinations leading to a total of 339 papers that were reviewed. Of these 339 studies, we excluded papers that did not list the coordinates of the localized FFA, or did not specify clearly what task was used to localize the FFA. Additionally, we included only experiments that were performed by healthy adults. After narrowing our search space, we were left with 49 papers which are listed in Table 1. Of the 49 studies, 16 used passive viewing, 30 used a 1-back memory task, and 3 used other localizer tasks. In addition, of the 49 studies, 22 used multiple objects as contrasting stimuli, 19 used single objects (houses, chairs, scenes, etc.) and 9 used scrambled stimuli[1]. Table 1 also includes the localized coordinates in the left and right fusiform gyrus for the peak activated voxel.

### *Results and analysis*

Fig. 1 shows the plots of the peak activated voxel across these different studies, colored by the different task conditions and the different comparison stimuli. The overall pattern shows substantial overlap among the different task and stimulus conditions. These qualitative findings were validated with quantitative tests. We performed a multivariate analysis to compute 3-dimensional confidence intervals for each task and comparison stimulus to allow us to compare the location of the different peaks more holistically rather than comparing *x*-, *y*-, and *z*-coordinates separately.

We composed 3-dimensional confidence intervals as can be seen in Fig. 2 for different task conditions and in Fig. 3 for different comparison stimuli. We performed a singular value decomposition on the covariance matrix of the *x*-, *y*- and *z*-coordinates giving us eigenvectors and eigenvalues that represent the lengths in the *x*-, *y*- and *z*-directions for the 3-dimensional confidence intervals. The eigenvalue length describes how varied the data are in the eigenvector direction. We then assumed a multivariate normal distribution to form the ellipses to represent 3-dimensional 68% confidence intervals (C.I.), which correspond to $\pm 1$ standard error (s.e.) away from each ellipse's mean (this is analogous to a two sample *t*-test that essentially examines $+1$ s.e. from one mean and $-1$ s.e. from the other mean). Unfortunately, many studies did not list other factors, such as the extent or magnitude of the localized clusters' activations, which could have further informed our construction of these confidence intervals. Off-center (diagonal) ellipse orientations indicate correlation among coordinates. The computed ellipses are relatively parallel (i.e., the ellipses are not tilted) indicating relatively independent measurements of the *x*-, *y*- and *z*-coordinates (i.e., low between-coordinate correlation). Additionally, there appears to be more variance in the *z*-coordinate in all cases as the ellipses have greater length in the *z*-direction. Future analyses could incorporate additional information such as cluster extent, sample size of each study, and variance estimates from each study to calculate more accurate confidence intervals.

This literature review shows that the type of localizer task and type of comparison stimuli do not appear to matter much in localizing the FFA in that there is a great deal of overlap in the confidence intervals for different tasks and comparison stimuli. Additionally, all the C.I.'s were centered to the same region. There appear to be some trends in the extent of variance in the activation peaks based on comparison stimuli, but no consistent trends were seen. For example, there

[1] Puce et al., 2006 used 2 comparison stimuli, which is why our total number of studies for comparison stimuli adds up to 50.

Author's personal copy



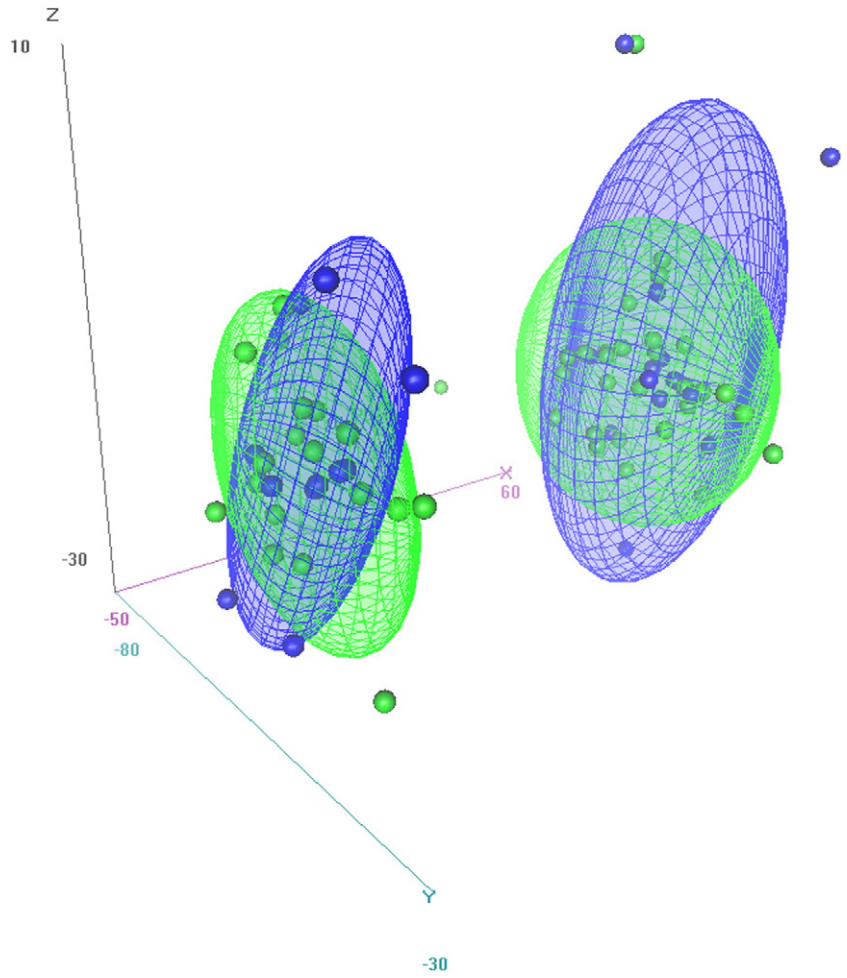



**Fig. 2.** 3-Dimensional 68% confidence intervals for passive viewing (blue) and 1-back (green) tasks. The coordinates represent the dimensions in Talairach Space in millimeter units.

appears to be more variance when comparing faces vs. scrambled stimuli, compared to faces vs. multiple objects in the right ROI (see Fig. 3). This might be due to more overlapping features of faces and multiple objects and more differential features between faces and scrambled stimuli. However, one must be careful in drawing conclusions based on these data, as there were more data-points for faces vs. multiple objects compared to faces vs. scrambles, which could also contribute to the increased variance for scrambles. Additionally, a different variance pattern was seen in the left ROI, which would complicate a feature interpretation.

In sum, the results from our literature review indicate that the type of task and comparison stimuli do not significantly affect localization of the FFA. However, this analysis was not only a between-subject analysis, it was a between-study analysis, which may not afford us the sensitivity to uncover differences due to these experimental manipulations. Sensitivity becomes important when results support the null hypothesis. Therefore, Experiments 1 and 2 aimed to explore these issues within individual participants performing the same tasks in the same experimental settings.

## Experiment 1

Experiment 1 was designed to explore localization of the FFA within subjects for different experimental manipulations. Such an experiment would provide greater sensitivity to uncover differences in FFA localization due to various experimental manipulations. Therefore, we could explore how task context might affect localization of the FFA, thereby exploring claims made by Friston et al. (2006) that the localizer task and experimental task contexts may alter localization.

### *Materials and methods*

#### *Participants*

Eighteen University of Michigan undergraduates participated in the study (9 males, 9 females, mean age: 21.45, s.d. = 2.84). Three participants were excluded for poor performance (at or near chance levels for all trial-types, including passive viewing where a button response was required every time a face or house was shown) and one participant was removed due to scanner artifact, leaving 14 participants. All participants gave informed consent as overseen by the University of Michigan's Institutional Review Board.

#### *Experimental procedure*

Participants performed 6 runs of an *n*-back memory task on faces and houses that varied in memory load. There were two runs of passive viewing in which participants pressed a key every time they saw a face or a house, two runs of a 1-back task in which participants responded whether the current face or house matched or did not match the previous face or house, and two runs of a 2-back task in which participants responded whether the current face or house matched or did not match the face or house that occurred two trials back.

Runs were randomly assigned, and a participant performed only one task per run: passive-viewing (0-back), 1-back, or 2-back. Each run began with an instruction about which task the participant should

Author's personal copy



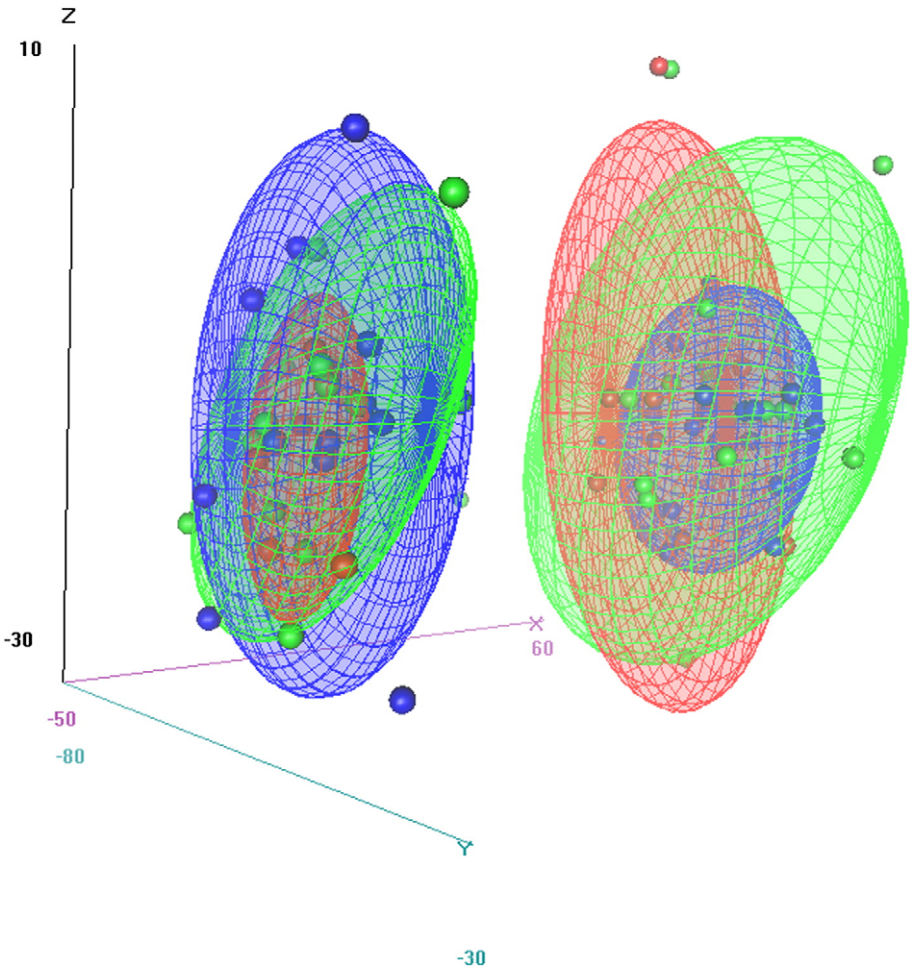



**Fig. 3.** 3-Dimensional 68% confidence interval for multiple object (blue), single object (green), and scrambled images (red) as comparison stimuli. The coordinates represent the dimensions in Talairach Space in millimeter units.

perform. Runs were then separated into blocks of faces or houses that alternated in a predictable fashion, either face-house or house-face. Within each block a participant had 15 trials. On average there were seven 0-back trials (trials where a novel face or house was shown), four 1-back trials and four 2-back trials. So, if the participant was in a 1-back run, the participant would respond "yes" for the 1-back trials and "no" to the 0-back and 2-back trials. This meant that no matter what the task, the stimuli displayed per block and run were the same and only the task demands/instructions differed. There were 36 face stimuli and 36 house stimuli, and each face or house was shown once, unless that face or house was a 1-back or 2-back trial. All faces had neutral expressions and all the face and house images were in black and white. Each block took 30 s, followed by a 15-s fixation cross. There were 10 blocks per run, and each run was 7 min and 40 s in length (there was 10 s of discarded acquisitions at the beginning of each run). Each stimulus was displayed for 1750 ms, with a 250 ms fixation cross separating each stimulus. In each passive-viewing block, participants were instructed to press either the button under their left or right index finger whenever they saw a face or a house (this was under their own discretion). Experimental tasks were presented using E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA). Participants practiced 3 runs of the task (a run at each *n*-back level) the day before they were scanned.

*Functional neuroimaging parameters*

Whole brain imaging was acquired on a GE Signa 3-T scanner equipped with a standard quadrature head coil. Head movement was minimized using foam padding and a cloth restraint strapped across participants' foreheads. Functional T2* weighted images were acquired using a spiral sequence with 40 contiguous slices with $3.44 \times 3.44 \times 3$ mm voxels (repetition time, or TR = 2000 ms; echo time, or TE = 30 ms; flip angle = 90°; field of view, or FOV = 22 $mm^2$). A T1-weighted gradient echo anatomical overlay was acquired using the same FOV and slices (TR = 250 ms, TE = 5.7 ms, flip angle = 90°). Additionally, a 124-slice high-resolution T1-weighted anatomical image was collected using spoiled-gradient-recalled acquisition (SPGR) in steady-state imaging (TR = 9 ms, TE = 1.8 ms, flip angle = 15°, FOV = 25–26 $mm^2$, slice thickness = 1.2 mm).

Each SPGR anatomical image was corrected for signal inhomogeneity and skull-stripped using FSL's Brain Extraction Tool (Smith et al., 2004). These images were then segmented using SPM5 (Wellcome Department of Cognitive Neurology, London) to separate gray and white matter voxels using the International Consortium of Brain Mapping (ICBM) tissue probability maps, and affine normalization parameters were calculated from those maps which were in standard MNI space. These parameters were then applied to the functional images to normalize all participants' brains into the same space with a resolution of $3 \times 3 \times 3$ mm and spatially smoothed[2] with a Gaussian kernel of $8 \times 8 \times 8$ mm. Functional images were

[2] We spatially smoothed our data as most fMRI studies do. However, it could be that spatially smoothing the data could mask differences in localization, by making the data more spatially correlated than they already are. However, most researchers, when performing a localizer task, smooth both their localizer data and the data of their task of interest as spatially smoothing improves power. Therefore, we believe that spatially smoothing our data was appropriate in this context. Additionally, we performed some analyses on unsmoothed data and obtained similar results.

Author's personal copy



corrected for differences in slice timing using 4-point sinc interpolation (Oppenheim et al., 1999) and were corrected for head movement using MCFLIRT (Jenkinson et al., 2002). To reduce the impact of spike artifacts we winsorized functional images on a voxel-by-voxel basis so that no voxel had a signal greater than 3.5 standard deviations from the mean of the run (Lazar et al., 2001). All analyses included a temporal high-pass filter (128 s).

The resulting images were entered into a general linear model with 3 conditions of interest: fixation, face and house, for each run. Additionally, 0-back, 1-back and 2-back conditions were modeled separately. Blocks were convolved with the canonical hemodynamic response function of duration 30 s for face and house, and 15 s for fixation. Additionally, we added 24 motion regressors into our model which included the linear, squared, derivative and squared derivative of the six rigid body movement parameters (Lund et al., 2005). In total, there were 168 regressors (27 for each run, with 6 run constants).

### Results and discussion

#### Behavioral results

As expected, participants had more difficulty with the task as the *n*-back level increased from 0-back (passive viewing) to 2-back. Four 1-way ANOVAs were composed with *n*-back level as the single predictor. ANOVAs were conducted for face and house separately and also separately for accuracy and reaction time data. Mean accuracy decreased linearly with increased *n*-back levels for both face $F(1,13)=24.00$ $p<0.001$, and house $F(1,13)=56.37$ $p<0.001$ blocks, though accuracy was above 92% for all blocks. Additionally, mean Reaction Times increased linearly with increased *n*-back level for both face $F(1,13)=95.60$ $p<0.001$, and house $F(1,13)=110.67$ $p<0.001$ blocks. Additionally, when we ran 2 (stimulus: face, house)×3 (*n*-back level: 0, 1, 2) ANOVAs separately for accuracy and reaction, we found no interactions between stimulus and *n*-back level for accuracy, $F(2,26)=0.58$, n.s., or reaction time, $F(2,26)=1.45$, n.s. These results are summarized in Table 2.

#### fMRI results

Our main goal in this experiment was to assess the degree to which FFA activation was altered by different experimental manipulations such as *n*-back memory load. To perform these analyses, we composed an ROI in temporal cortex that spanned the FFA. This ROI was created anatomically with the Wake Forest PickAtlas (Casanova et al., 2007) for the fusiform gyrus, which we truncated to restrict it to the left and right mid-fusiform gyri. The standard space coordinates (Montreal Neurological Institute; MNI) for this ROI were $X=-55$ to $-18$ for the left ROI, and $X=55$ to 18 for the right ROI and $Y=-88$ to $-30$ and $Z=-28$ to $-5$, and contained approximately 750 voxels for each ROI (some participants had a few more voxels based on their anatomy). This anatomical ROI was used as a bounded and uniform search space with which to look for peak voxels for each individual participant. Here we present results for the contrasts of Faces–Houses. Additionally, when we separated our results for different tasks, we always used the corresponding house condition as the comparison (i.e., if we explored face activation for the 1-back task, we used the house 1-back runs as the contrast).

**Table 2**
Behavioral results for Experiment 1.

| | Accuracy | | Reaction time | |
|---|---|---|---|---|
| | Mean | Std. deviation | Mean | Std. deviation |
| Face 0-back | 0.99 | 0.02 | 403.79 | 111.07 |
| Face 1-back | 0.96 | 0.03 | 649.03 | 112.38 |
| Face 2-back | 0.93 | 0.06 | 730.54 | 126.86 |
| House 0-back | 0.99 | 0.03 | 420.10 | 143.37 |
| House 1-back | 0.96 | 0.03 | 674.59 | 123.31 |
| House2 2-back | 0.92 | 0.05 | 771.58 | 120.48 |

#### Peak distances

The first analysis that we performed for Experiment 1 was to assess whether the location of the peak voxel differed depending on task conditions. Participants performed two runs of passive viewing (0-back), 1-back, and 2-back tasks. Therefore, we could compare differences in peak voxel location both within- and between-tasks to interrogate whether task demands changed the location of the peak activated voxel.

In Fig. 4, we plot the location of the peak activated voxel for each participant performing either the first or second run of passive viewing, 1-back, or 2-back tasks. There is no discernable difference between the locations of the peak voxel for different task conditions. To test this, we calculated the Euclidean distance for the location of the peak activated voxel both within- and between-tasks for each individual subject. In other words, we calculated the distance between the locations of the peak voxel for the first run of performing passive viewing with the second run of passive viewing in order to get a within-task distance measure. We did the same calculation for the other tasks. In our paradigm, we had 3 pairs of within-task differences, which we averaged together. To assess between-task distance for the peaks, we calculated the average of all 12 pairwise Euclidean distance measures. We performed these analyses separately for both the left and right hemispheres. For the right-hemisphere ROI, we found that the average within-task distance in peak activation across all subjects was 8.9 mm[3] , s.d. = 8.74, and the average between-task difference across all subjects was 7.8 mm, s.d. = 6.06. With a paired *t*-test we found that these differences were not reliably different from one another, $t(13)=1.04$ n.s. We found similar results in the Left hemisphere ROI: The average within-task distance in peak activation was 9.3 mm, s.d. = 6.96, and the average between-task difference was 9.5 mm, s.d. = 7.33. With a paired *t*-test we found that these differences were not reliably different from one another, $t(13)=0.26$ n.s. In addition, we also performed the same analyses taking a random sample of 3 between-task distances to equate the number of distance calculations both between- and within-tasks. When doing so, we again found no differences in the variance of peak locations within- and between-tasks, $t(13)=1.38$ n.s. for the right ROI and $t(13)=0.12$ n.s. for the left ROI.

In summary, when calculating the differences in the location of the peak activated voxel we found no differences in the variability of peak locations based on task. In addition, our paradigm allowed us to assess the reliability of the peak location (within-task distances) compared to between-task distances all within each participant, which is an added advantage over past studies that compared peaks to grand mean distances rather than to individual participant peaks.

#### Consistency

While there may not be differences in the location of the peak voxel as a function of task, there could be differences in the spatial overlap of the activated voxels for different task conditions. Duncan et al. (2009) computed a consistency metric to assess this with the following equation:

$$R_{ij} = 2 \times V_{ij} / \left(V_i + V_j\right), \quad (1)$$

where $V_{ij}$ are the number of voxels that were active in both tasks *i* and *j*, $V_i$ are the number of voxels activated in task *i*, and $V_j$ are the number of voxels activated in task *j*. A value of '1' indicates that a set of active voxels is active in both tasks, and a value of '0' indicates that the tasks

[3] The two passive viewing runs had more variance in the peak location, which increased the average within-task distance above the average between-task distances.

Author's personal copy



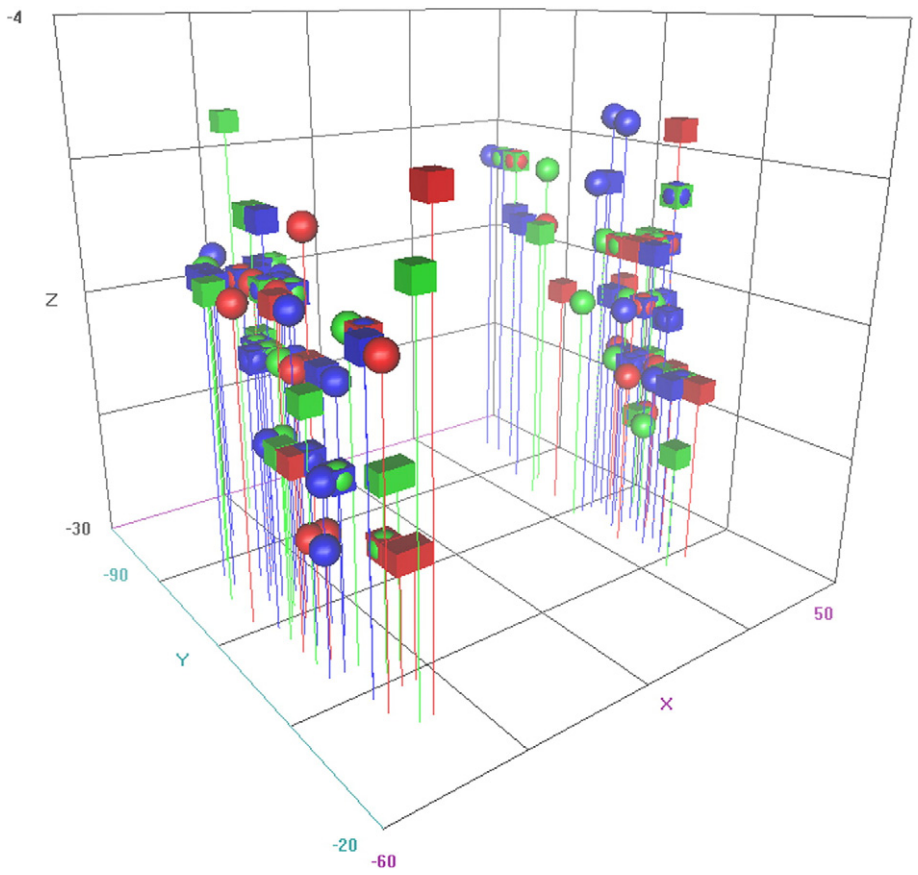



**Fig. 4.** Plotted Peaks for each participant for each run of the passive viewing, 1-back and 2-back memory tests for Faces>Houses. Blue denotes passive viewing, green denotes 1-back, and red denotes 2-back. Spheres denote the first iteration of the task, and cubes denote the second iteration. From the figure one can see no discernable pattern of changes in peak location based on task, or the first or second iteration of performing the same task.

activate disjoint sets (Duncan et al., 2009). Duncan et al. (2009) found that when the activation threshold was increased, the number of shared or consistent voxels decreased. We performed the same analysis on our data and found a similar decrease in consistency with increasing thresholds which can be seen in Figs. 5 and 6. We do not find this decrease surprising. As we have shown in our previous analysis of peak distances, there is noise in the location of the peak activated voxel. Therefore, for one task condition the peak will be in one spot, and for another task condition (or the same task just performed a second time) the peak will be in another different, but nearby spot. These peaks are also surrounded by a distribution of activated voxels that are likely shaped as a 3-dimensional Gaussian. At lower statistical thresholds there will be more overlapping voxels between these two peaks than at higher statistical thresholds.

It might be more informative to compare the decrease in consistency as a function of whether the calculation is performed within- or between-tasks. For example, if there is less consistency between-tasks compared to within-tasks, then we might conclude that using functional localizers may not be ideal because it suggests that between-task consistency is less than within-task consistency. We performed this calculation on our data. Fig. 7 shows both within- and between-task consistency values for the right and left Fusiform ROI. We randomly selected 3 between-task combinations to attain our average between-task consistency score, and then averaged the 3 within-task combinations to attain the average within-task consistency score. As shown in Fig. 7, there are no apparent differences in consistency as a function of whether consistency was calculated within or between tasks. Additionally, we tested the slope of the linear trend in decreasing consistency as a function of increasing activation thresholds, and this slope does not differ whether the calculation is performed within- or between-tasks, $t(13) = -0.72$, n.s. for the right ROI and $t(13) = -1.16$, n.s. for the left ROI. Therefore, we do not believe this consistency measure distinguishes localizer tasks as being different from one another.

Moreover, Duncan et al.'s (2009) data and our data show roughly 60-70% consistency at the lowest statistical thresholds, which indicates that nearly 1/3 of the data cannot be replicated from one iteration of the task to the next iteration of the same task. The question is, however, if these inconsistent voxels are noise, or if they do carry some non-noise information even if they are not consistent according to the above definition. It has been shown that even when the peak activations with the most reliable activation scores are removed, multivoxel pattern classifier algorithms can still robustly classify object categories by brain activity patterns without these peak activations (Haxby et al., 2001). Therefore, reliable and important information may be conveyed by less active voxels as now shown in many independent datasets (see for example Dinstein et al., 2008; Downing et al., 2007; Kamitani and Tong, 2005). These less active voxels are likely less consistent as well, but may still contain useful information. We appreciate Duncan et al.'s (2009) concern that even at the most liberal thresholds, 1/3 of the voxels are not being consistently activated, but more work should be done exploring how consistency relates to the amount of information that can be drawn from less consistent voxels, especially if those voxels are analyzed in a multivariate manner.

*Multivoxel Pearson correlations*

Even though our previous consistency calculations take into account multiple voxels, the consistency calculation (i.e., Eq. (1)) may not be an ideal metric to evaluate the similarity or reliability of activation in an ROI for different tasks. First, the consistency metric is calculated in a univariate manner comparing one voxel to itself for

Author's personal copy



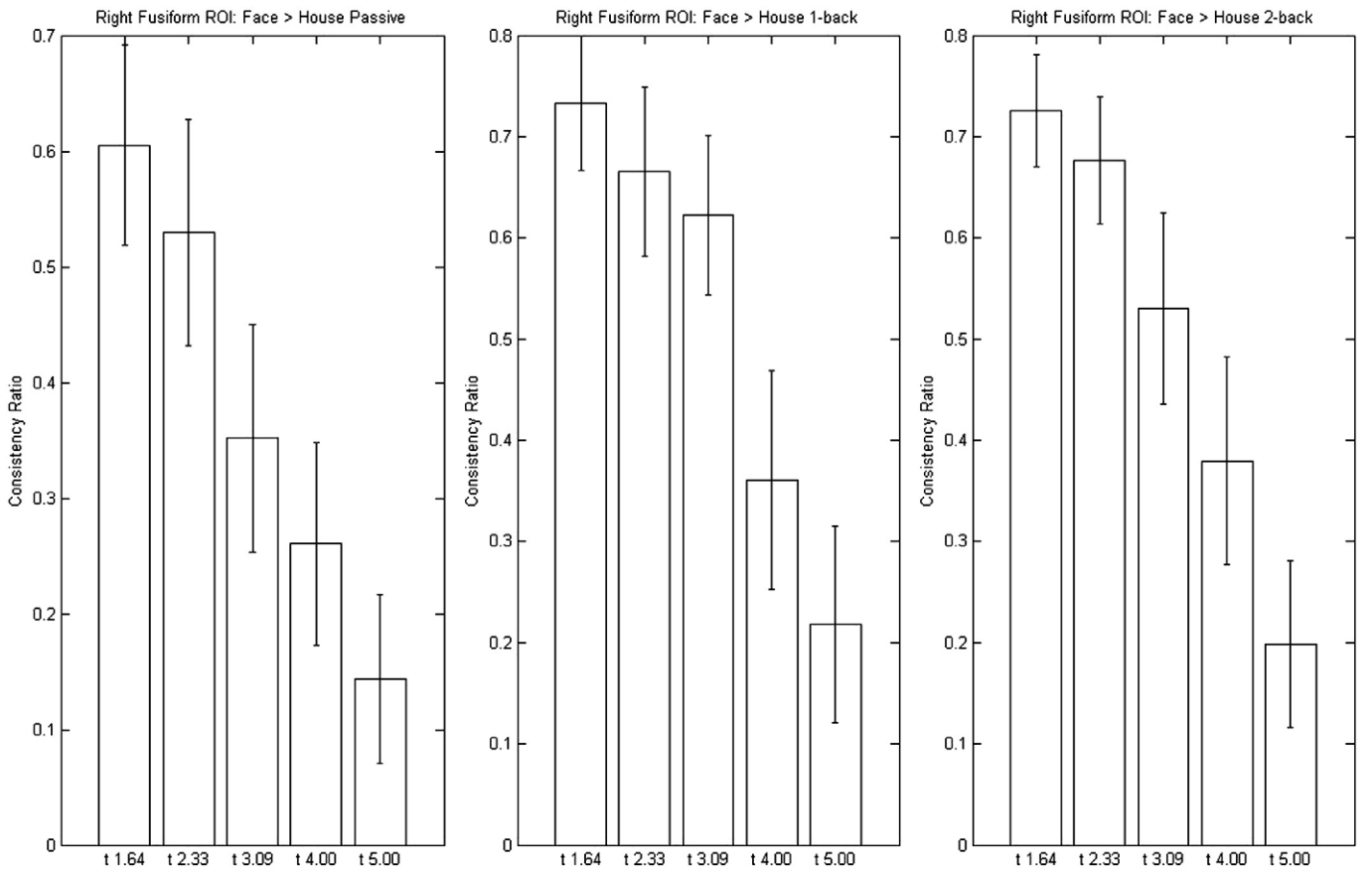



**Fig. 5.** Mean (±1 standard error) within-task Consistency Ratios across all subjects, as a function of increasing the threshold for what constitutes an active voxel in our right fusiform ROI.

different task conditions independently from the other voxels in the ROI. In other words, the response magnitudes of voxel i from task 1 and from task 2 are compared to each other, but the other voxels in the ROI are not considered when calculating the relationship between those two voxels. This indicates that voxels are compared in isolation, which may mean that the overall pattern of activity in the ROI could be missed. Second, the consistency equation requires the selection of a statistical threshold to assess similarity, which has its own problems (i.e., what is an appropriate threshold?). This also creates an additional quandary as consistency is defined in binary terms (either both on or not), which is problematic because the data are continuous. All of these problems may cause this consistency calculation to miss the overall pattern of activations within the ROI, and therefore underestimate the reliability of a functional localizer task. As we stated earlier, if the ROI has the same pattern of activation for two tasks, but task 1 activates much more robustly than task 2 (i.e., double the magnitude), this consistency calculation could potentially find zero consistency, even if the pattern of activity is identical for the two tasks. Therefore, the consistency measure may actually underestimate the reliability or similarity both within- and between-tasks. Although thresholding and binary decisions (i.e., the voxel is either on or off) are typically employed to define functionally localized areas in practice, evaluating of the reliability of activation patterns can be done more accurately by calculating Pearson correlations for patterns of activation across multiple voxels, which overcomes these outlined problems.

Here is an example to make this procedure more concrete. If we are interested in finding the similarity between passive viewing and 1-back tasks we would take the activation of all 750 voxels in the right ROI for passive viewing and correlate the activation of those 750 voxels with the activation of those same 750 voxels in the 1-back task to get a single correlation coefficient. This would give us the similarity of passive viewing with 1-back (i.e., a between-task similarity measure). This differs from the consistency equation where 750 binary comparisons would be made for each individual voxel to see if each voxel was above some statistical activation threshold for passive viewing and 1-back.

With Pearson correlations we assessed the similarity/reliability in the overall pattern of activation within-task by averaging the Pearson correlations for our 3 within-task pairs. In other words, we correlated all 750 voxels in the right ROI for the first run of passive viewing, with all 750 voxels in the right ROI for the second run of passive viewing and obtained a single Pearson correlation coefficient ($r$). We performed the same calculation for the two runs of the 1-back and 2-back tasks, to get a total of 3 within-task $r$-values. We then averaged these 3 $r$-values together to get a single within-task correlation coefficient. We implemented a similar procedure as before to calculate the between-task Pearson correlations, in which we randomly selected 3 between-task pairs and averaged their correlation scores together to get a single between-task correlation score. This analysis speaks to Poldrack's (2007) concern that the distribution or pattern of activation in an ROI may change as a function of task demands. In other words, while the peak activated voxel's location may not change as a function of task, it may be that the pattern of activation could change as a function of task. For example, it could be that the distribution of activation for voxels for the 0-back task would be more diffuse or irregular compared to the pattern of activation for the 2-back task.

In Experiment 1, we did not find any differences in $r$-values for within- and between-task correlations. The mean correlation coefficients across participants for both between- and within-task conditions were high (all above $r = 0.78$). With paired $t$-tests we compared the correlation scores both within- and between-tasks for the left and

Author's personal copy



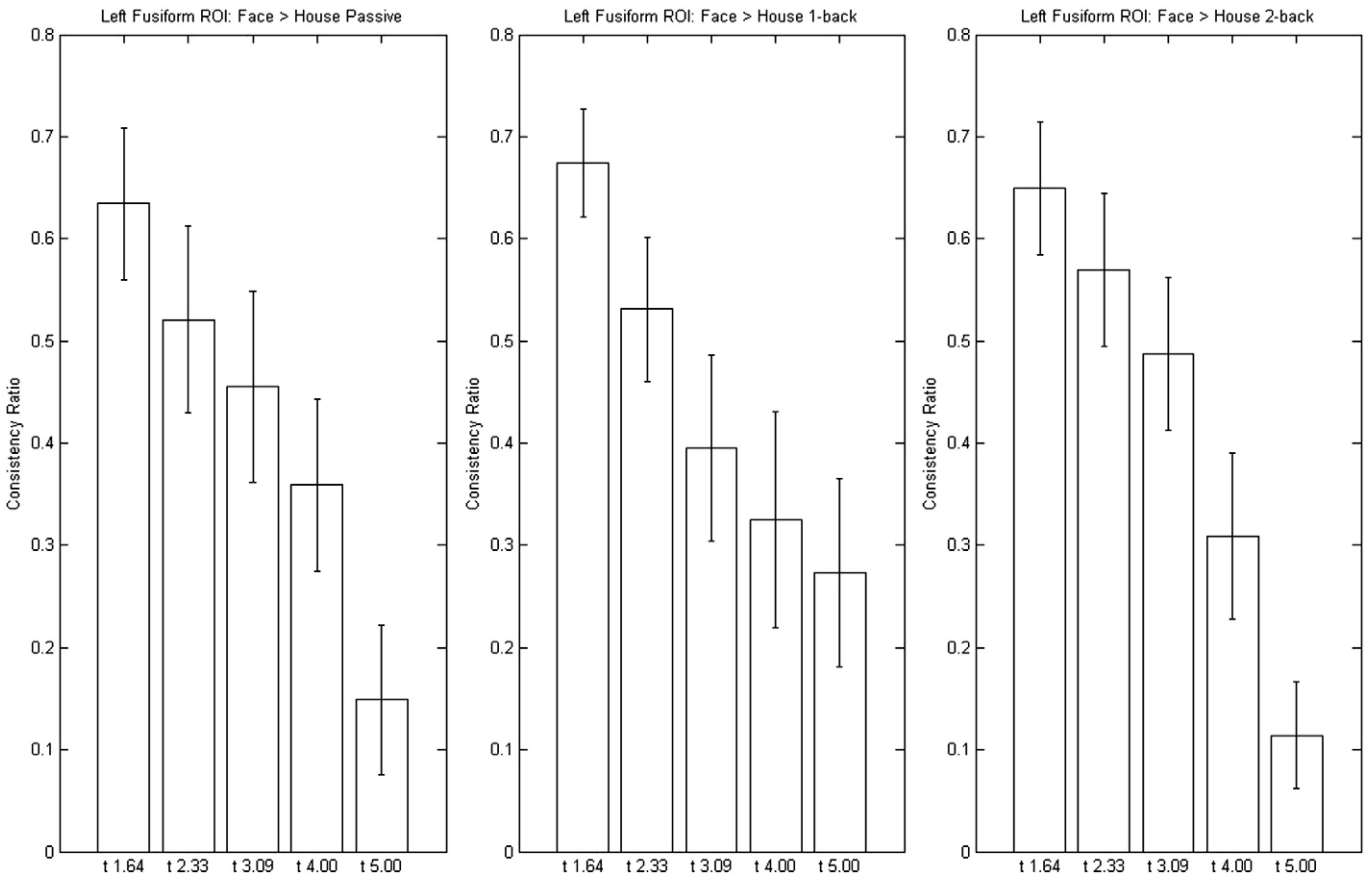



**Fig. 6.** Mean (±1 standard error) within-task Consistency Ratios across all subjects, as a function of increasing the threshold for what constitutes an active voxel in our left fusiform ROI.

right fusiform ROIs, and we found no reliable differences, $t(13) = -0.47$, n.s. for the right ROI and $t(13) = 0.95$, n.s. for the left ROI. Because some of the same data were used for the between- and within-task correlations we performed an additional analysis in which we calculated within-task correlations with one random within-task pair (i.e., 1-back with 1-back) and then randomly picked one of the between-task correlations that did not involve that task (i.e., 1-back in this example). When we performed this type of calculation we still found no reliable differences for within- and between-task correlations; each within- and between-task correlation was high (always 0.78 or above for different random combinations).

These analyses also address another issue related to averaging the activation of all voxels in an ROI. For example, a researcher could find similar activations when averaging activations across all voxels in an ROI for different task conditions (to obtain a single activation value), and potentially miss the distribution of how those voxels are activated (i.e., the activation pattern). This was a concern raised by Friston et al. (2006) and Poldrack (2007). Our analyses here indicate that for the FFA, different task demands still maintain high similarity/reliability of activation patterns in this region as measured in a multivoxel manner.

In sum, we found no reliable differences for within- and between-task correlations based on task demands. In addition, this procedure examined the overall pattern of activation in the ROI for different tasks demands, and so it overcomes the other problems associated with the calculation of consistency and therefore provides a good measure of reliability of the localization.

*Comparison stimuli, extent and magnitude*

With three different analyses we found the same null result. It seems that our different localizer tasks do not produce differences in localization in FFA. Thus far, however, all of our analyses were exploring the location and pattern of activity in the FFA, and did not assess differences in magnitude or extent of the FFA activations. Additionally, all of our analyses thus far used houses as the contrasting stimuli. As such, we can perform similar analyses using fixation as the control condition. We note that using fixation is not ideal as there is not any experimental control over what participants are doing when presented with fixation alone; nevertheless we found it informative to explore the impact of using fixation as a contrast to our face activation. With fixation as the contrasting stimulation, we again found no differences in the location of the peak in the right or left fusiform ROI when comparing distances both between- and within-task, $t(13) = 0.69$, n.s. in the right ROI and $t(13) = 1.57$, n.s. in the left ROI.

The most pronounced effect when using fixation as the contrast stimulus was larger activations in the FFA (greater $t$-values) and also greater spatial extent. We performed a repeated measures ($2 \times 3$) ANOVA separately for the left and right Fusiform ROIs for two dependent variables; the mean activation in the ROI and also the spatial extent of the activation, which was defined as the number of voxels activated with a $t > 2.5$. Our model had two predictor variables: stimulus, which had two levels (house or fixation), and $n$-back level, which had 3 levels (0-, 1-, and 2-back). This analysis speaks directly to Friston et al.'s (2006) concern that many functional localizer approaches do not allow for the testing of interactive effects, which may turn out to be significant.

In our analysis we found significant main effects of stimulus type on both mean activation levels in the ROI and spatial extent for both the right and left ROIs. For the right ROI, $F(1,13) = 86.07$, $p < 0.001$ for differences in mean activation by stimulus type and $F(1,13) = 191.12$, $p < 0.001$ for differences in extent by stimulus type. The same was true in the left ROI , $F(1,13) = 57.60$, $p < 0.001$ for differences in mean

Author's personal copy



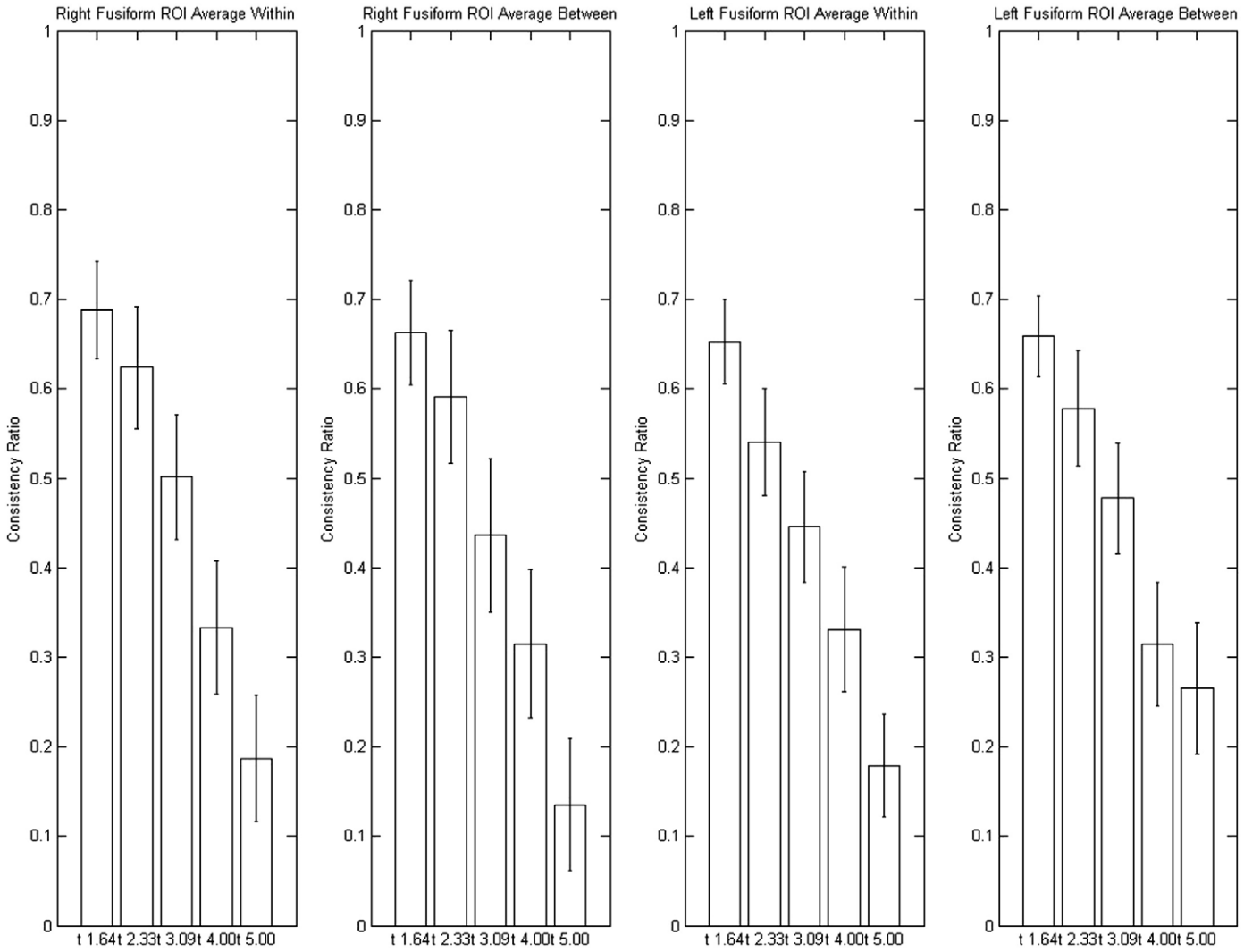



**Fig. 7.** Mean (±1 standard error) Consistency Calculations for both within- and between-task for the left and right Fusiform ROIs as a function of increasing the threshold for what constitutes an active voxel.

activation by stimulus type and $F(1,13) = 78.80$, $p < 0.001$ for differences in extent by stimulus type.

Additionally, the interaction of *n*-back level and stimulus type was also reliable for both mean activation levels and spatial extent for both the left and right ROIs ($F(2,26) = 7.63$, $p < 0.005$ for differences in mean activation in the right ROI; $F(2,26) = 4.40$, $p < 0.05$ for differences in extent level in the right ROI; $F(2,26) = 10.12$, $p < 0.005$ for differences in mean activation level in the left ROI; $F(2,26) = 8.17$, $p < 0.005$ for differences in extent level in the left ROI). This meant that the load predictor (*n*-back level) was only a reliable factor when using fixation as the contrasting stimulus. Therefore, we replicated Druzgal and D'Esposito's (2001) finding of load effects in FFA only when fixation is used as the contrasting stimulus. A summary of the results from these analyses is shown in Table 3.

Additionally, we computed the similarity/reliability of activation patterns across comparison stimuli and task conditions using Pearson correlations. We found that the correlations between-stimuli (house and fixation) were significantly lower ($r = 0.50$) compared to within-stimuli ($r = 0.80$; $p < 0.001$ for left and right ROIs), but were still high and did not depend on task conditions.

Therefore, it appears that the foremost effects of different comparison stimuli are on the extent and magnitude of activation within the Fusiform ROI. While significantly reduced, the pattern or reliability of activations within the ROI still maintain relatively high levels (as calculated with Pearson correlations). In other words, the correlation patterns are higher for the same comparison stimuli (e.g., face vs. house 1-back correlated with face vs. house 2-back), than for two different comparison stimuli (e.g., face vs. house 1-back correlated with face vs. fixation 1-back). In some ways, this matches an earlier finding from Kanwisher et al. (1997) who found more face selective voxels for face vs. house contrasts compared to face vs. scrambled face contrasts (i.e., for face vs. scrambles some voxels may be active that are not necessarily involved in face processing). In sum, even with different comparison stimuli, the reliability of the localized region remains strong compared to localizing the same area with the same task conditions.

**Table 3**
Localizer results for Experiment 1. Mean activations are negative for Faces vs. Houses because less voxels were active for that contrast as exhibited by the smaller spatial extents compared to Faces vs. Fixation. Mean activations are mean *t*-values across all 750 voxels.

| | Faces vs. Houses | | | Faces vs. Fixation | | |
|---|---|---|---|---|---|---|
| *Left ROI* | 0-back | 1-back | 2-back | 0-back | 1-back | 2-back |
| Mean activation | −1.11 | −1.91 | −2.00 | 1.56 | 2.41 | 2.11 |
| Extent | 98.57 | 69.43 | 70.86 | 245.14 | 302.36 | 284.07 |
| *Right ROI* | | | | | | |
| Mean activation | −1.92 | −2.44 | −2.50 | 2.19 | 3.04 | 3.32 |
| Extent | 87.93 | 78.79 | 71.07 | 317.86 | 369.50 | 373.93 |

Author's personal copy



*Summary of Experiment 1*

From Experiment 1 we found that the type of task did not alter activation location, magnitude, consistency, or activation reliabilities within left and right Fusiform ROIs. However, when comparing across different comparison stimuli, differences did emerge in terms of magnitude, spatial extent, and reliability of activation patterns. This validates some of Friston et al.'s (2006) concerns that task context may matter, and different experimental manipulations (such as task and comparison stimuli) may interact. Importantly, these interactions affected the strength of the activations in the FFA, but not the locations where the activations were occurring.

Yet even with different comparison stimuli, we still found high correlations (>0.50) in the reproducibility of activation patterns in the ROIs. The reliable differences observed across comparison stimuli suggest that the present study was sufficiently powered to find effects, so sample size is not an issue. Additionally, the sample size in Experiment 1 is within the range of a typical study looking at differences in FFA activation as a function of experimental manipulations. However, sample size and the type of comparison stimulus used could affect the localizer procedure. Therefore, we attempted to replicate the findings of Experiment 1 with an experiment containing a larger sample size and utilizing a different comparison stimulus.

## Experiment 2

In an independent dataset we performed the same calculations and drew similar conclusions. The relative advantages of this second dataset are (1) that there were more participants which results in larger power to detect differences, (2) that different comparison stimuli were used which allowed us to investigate the effect of comparison stimuli in more constrained ways compared to Experiment 1 and (3) there were more replications for each experimental manipulation which allowed us to compute between- and within-task effects more independently.

### *Materials and methods*

#### *Participants*

Forty-eight University of Michigan undergraduates (32 females and 16 males with mean age of 20.8) participated in this study. All participants gave informed consent as overseen by the University of Michigan's Institutional Review Board. Portions of these data had been previously published in another study (Polk et al., 2007).

#### *Experimental procedure*

Participants first performed three runs of a passive viewing task on grey-scale pictures from five categories: faces, houses, pseudowords, chairs, and phase-scrambled control stimuli (versions of all the experimental images in which the phase information was scrambled, but with spatial frequency preserved). Each run consisted of fifteen 20-s blocks (three blocks of each of the five categories of stimuli in pseudorandom order) with a 20-s rest period preceding each run. Each block consisted of 10 items from the same category presented for 1500 ms each, followed by a 500 -ms inter-trial interval. Passive-viewing blocks did not require any response from the participants. Participants then performed three runs of a one-back matching task (pressing a button with the right, middle finger if the current stimulus matched the last stimulus, pressing a button with the right, index finger if they did not match) on the same set of stimuli in a similar fashion.

#### *fMRI parameters*

High-resolution T1-weighted anatomical images were collected using spoiled-gradient-recalled acquisition (SPGR) in axial slices parallel to the AC/PC line with a resolution of $0.9375 \times 0.9375 \times 5.0$ mm and with the following parameters (TR = 9 ms, TE = 1.8 ms, flip angle = 15°, FOV = 25–26 $mm^2$, slice thickness = 1.2 mm). Neural activity was estimated based on the blood–oxygen level dependent (BOLD) signal measured by a GE 3T scanner using a spiral acquisition sequence (TR = 2000 ms, 30 5 mm thick axial slices with an in-plane resolution of $3.75 \times 3.75$ mm, FOV = 24 $mm^2$, TE = 30 ms, flip angle = 90°). A T1-weighted gradient echo anatomical overlay was acquired using the same FOV and slices (TR = 250 ms, TE = 5.7 ms, flip angle = 90°) as the functional images (spirals).

The functional images (T2*-weighted echo planar images) for each participant underwent reconstruction, slice timing correction, and realignment as part of preprocessing. High-resolution anatomical image for each participant, corrected for signal inhomogeneity, was coregistered to the functional images. Then, the anatomical image was segmented using SPM5 (Wellcome Department of Cognitive Neurology, London) to separate gray and white matter voxels using the International Consortium of Brain Mapping (ICBM) tissue probability maps, and affine normalization parameters were calculated from those maps which were in standard MNI space. The functional images were then normalized to the template space with a resolution of $3 \times 3 \times 3$ mm and spatially smoothed with a Gaussian kernel of $8 \times 8 \times 8$ mm. All analyses included a temporal high-pass filter (128 s). In our analyses we chose to use faces vs. houses and faces vs. scrambled stimuli for some parallelism to our first experiment.

### *Results and discussion*

#### *fMRI results*

In addition to performing many of the same analyses as in Experiment 1, we could also perform many of the same tests comparing not only between-task differences, but also differences produced by different contrasting stimuli. In addition, by having 3 replications of each task, more independent tests could be performed compared to Experiment 1.

#### *Peak distances*

In Experiment 2 there were no differences in the variance of the location of the peak activated voxel for different task conditions

**Table 4**
Mean peak distances for different task conditions and contrasting stimuli. Two-tailed pair-wise *t*-test was performed between the within-task and between-task peak distances and between the within-stimulus and between-stimulus peak distances in the left and the right ROI.

| | | Contrasting stimulus | | | Task | |
|---|---|---|---|---|---|---|
| Left ROI | | House | Scrambled | | Passive | 1-Back |
| | Within-task | 13.12 | 11.39 | Within-stimulus | 11.364 | 14.121 |
| | Between-task | 14.64 | 14.03 | Between-stimulus | 12.602 | 13.193 |
| | *t*(47) | 0.822 | 1.236 | *t*(47) | 0.548 | −0.432 |
| | *p* | 0.416 | 0.223 | *p* | 0.586 | 0.668 |
| Right ROI | Within-task | 13.85 | 9.65 | Within-stimulus | 12.949 | 11.626 |
| | Between-task | 17.26 | 9.60 | Between-stimulus | 13.442 | 14.941 |
| | *t*(47) | 1.274 | −0.035 | *t*(47) | 0.191 | 1.507 |
| | *p* | 0.209 | 0.992 | *p* | 0.849 | 0.139 |

Author's personal copy



(passive viewing or 1-back) or contrasting stimuli (houses or scrambled images). Table 4 summarizes the results. In most cases, within-task or within-stimulus peak distances were smaller than between-task or between-stimulus peak distances. This trend was expected since performing the same task or viewing the same visual category should yield identical results in theory. However, the pairwise *t*-tests on these differences did *not* provide evidence that the variability of peak distances within-task and between-task or within-stimulus and between-stimulus were significantly different (Table 4). In short, we did not find any indications that task demands changed the variability of the location of the peak (a replication of Experiment 1), but additionally, we did not find differences utilizing a different comparison stimulus (i.e., scrambled images) which extended our findings from Experiment 1.

*Consistency*

Using the same consistency equation as in Experiment 1 (Eq. (1)), we found a similar pattern of decreases in consistency as a function of

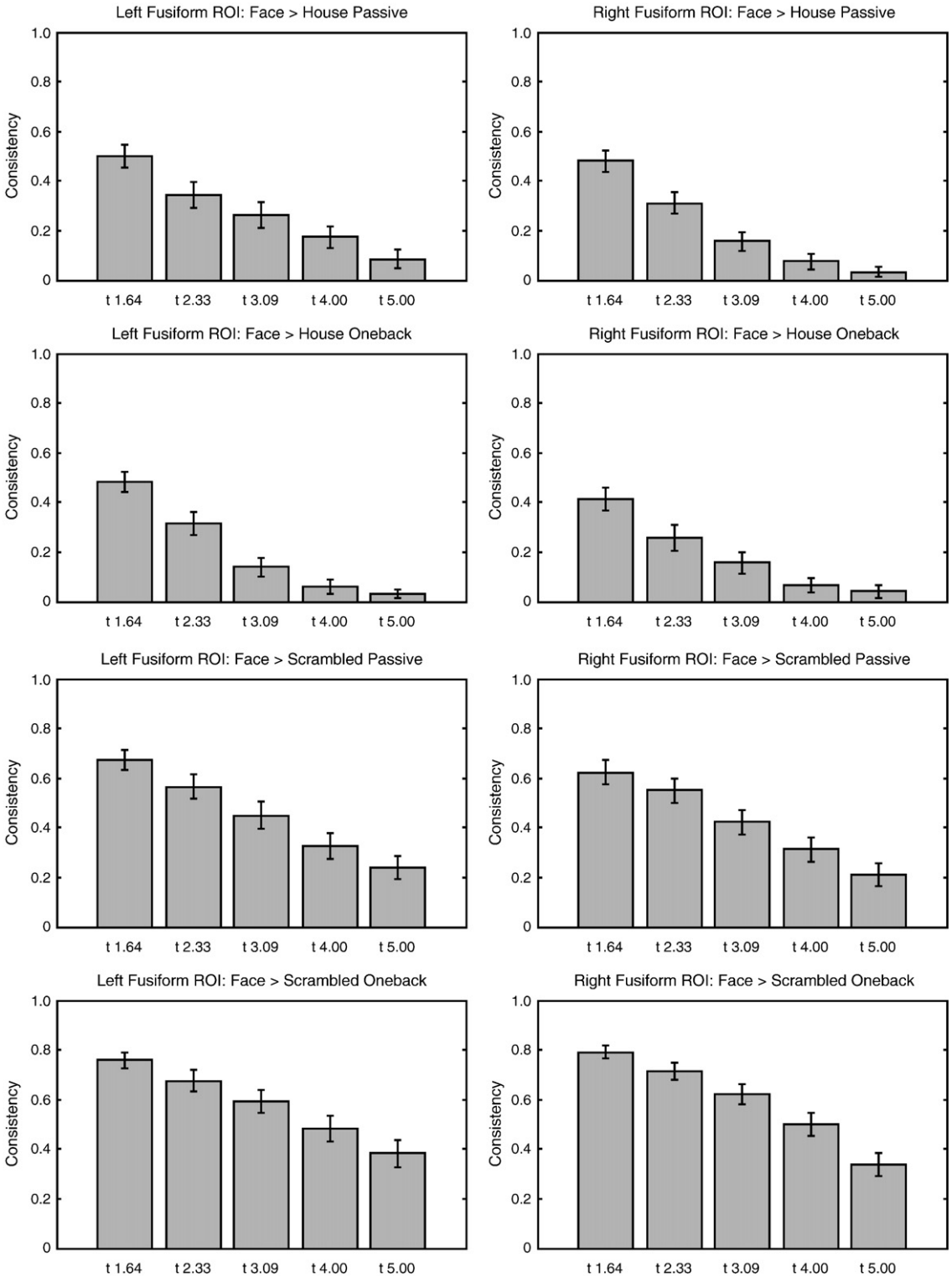



**Fig. 8.** Consistency results for Experiment 2 for the left and right ROIs for passive viewing and one-back for different comparison stimuli (houses and scrambles).

Author's personal copy



increasing statistical thresholds and did not find differences in consistency as a function of task. Fig. 8 shows the consistency patterns across varying statistical thresholds for different task demands and comparison stimuli. One will note that consistencies are on average higher when the contrasting stimuli are scrambled images rather than houses. This seems to be driven by greater magnitudes of activation for face vs. scramble compared to face vs. house, and matches our findings from Experiment 1 where we found greater activation for face vs. fixation compared to face vs. house.

*Multivoxel Pearson correlations*

Using Pearson correlations, we compared face related activation both within- and between-tasks for different comparison stimuli. For each participant, two of the three runs in each task condition (i.e. either passive-viewing or 1-back) were selected randomly to provide within-task correlation measures. We were then left with one independent run in the passive viewing and 1-back task conditions to provide more independent between-task measurements. Fig. 9 shows the between- and within-task correlation patterns for the different comparison stimuli. A few interesting patterns emerged. First, note that in all conditions, the correlations were relatively high (>0.60). Between-task correlations were always less than within-task correlations although not reliably so using pair-wise *t*-tests ($t(47) = 0.423$, $p = 0.674$ in the left ROI and $t(47) = 1.222$, $p = 0.228$ in the right ROI). This is not too unexpected because the reliability of a measure will always be higher than the correlation of that measure with some other measure. Of more interest is the trend of increased reliability when the task demands increased, specifically in the left ROI ($t(47) = 2.373$, $p = 0.028$; $t(47) = 0.908$, $p = 0.368$ in the right ROI). In other words, the within-task correlations for 1-back were reliably higher than the correlations for passive viewing (at least in the left ROI), which suggests that the more difficult task may have led to more reliable activation in the ROI. A second interesting finding is that the face vs. house contrast produced more reliable activation than the face vs. scramble contrast ($t(47) = 5.315$, $p < 0.001$ collapsing all bars in Fig. 9). This may be due to the face vs. house contrast having more constrained activation patterns in the fusiform gyrus compared to face vs. scramble, which also matches a related finding from Kanwisher et al. (1997). Lastly, we found relatively high cross-stimulus correlations (>0.60), but these correlations were reliably smaller than the within-stimulus correlations $t(47) = 30.2$ $p < 0.0001$ for the left ROI and $t(47) = 33.2$, $p < 0.0001$ for the right ROI. Therefore, different task demands and contrasting stimuli may provide more or less reliable/reproducible localization data.

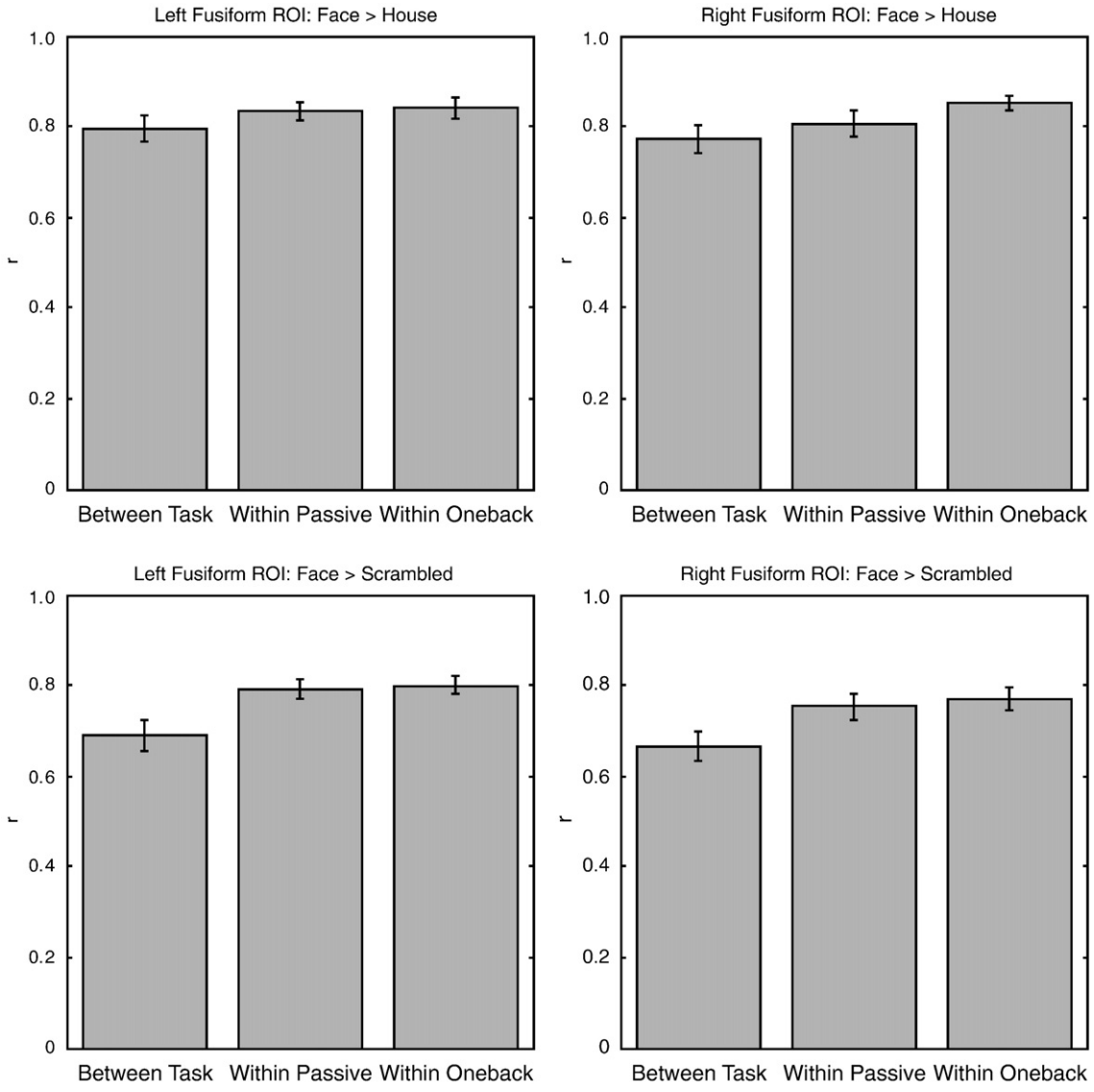



**Fig. 9.** Pearson correlations between- and within-task for Experiment 2.

Author's personal copy



*Comparison stimuli, extent and magnitude*

We found greater activation magnitudes and extents for faces vs. scrambled stimuli compared to faces vs. houses, but not for different task demands. This replicated our findings from Experiment 1. As one might expect, we found that faces vs. scrambled stimuli produced more robust activation in the FFA compared to faces vs. houses and also produced a larger swath of activation in the FFA (all $p<0.001$, $t(47)=14.95$ and $t(47)=13.12$ for magnitude difference in the left ROI and right ROI, respectively, collapsing across task; $t(47)=16.39$ and $t(47)=13.89$ for extent differences in the left ROI and right ROI, respectively, collapsing across task). These results are summarized in Table 5. Additionally, we did not find any consistent differences in the location of the peak activated voxel for different tasks, comparison stimuli, or different combinations of tasks and comparison stimuli. Lastly, the overall reliability of activation patterns in our ROIs all had high Pearson correlations (>0.5) even for different tasks and comparison stimuli, but the reliability of activations remains more stable for the same comparison stimuli compared to between-stimuli.

*Summary of Experiment 2*

The results of Experiment 2 replicated many of the same findings from Experiment 1, but extended them in important ways. While we found that different task demands and comparison stimuli leave the localization of FFA relatively invariant in terms of the variability of peak location, we found some important differences. First, with increasing task demands, we found more reliable activation within our ROIs. Second, the reliability of activation produced by different comparison stimuli, was also an important factor, as faces vs. houses produced more reliable activation compared to faces vs. scrambles. Lastly, with less constrained comparison stimuli (such as scrambles and fixation), we observed greater activation magnitudes and more extent (a replication of Experiment 1), but less reliability, which are important considerations when one chooses a localizer task. As in Experiment 1, changing comparison stimuli had the largest effects by altering the magnitudes, extents and reliabilities of activation patterns in the ROI.

## General discussion

The present research shows that different localizer task conditions do not appear to alter the variability of the locations of activations in the fusiform gyrus when responding to face stimuli. This is important because it indicates that experimenters have some flexibility in terms of what tasks they employ to define face responsive areas to form ROIs to analyze their tasks of interest. With that said, there are a few caveats to note.

First, we may not be finding differences in localization based on task demands because our experimental tasks may not be sufficiently dissimilar. That is, passive viewing and *n*-back tasks share many of the same task demands. It is not clear whether performing other more varied tasks on faces would lead to different localization. While we grant that our tasks are not radically dissimilar, they are not particularly similar either. Further, it would be difficult to test the whole task space. We chose these tasks because researchers in the field typically use these tasks to localize face-responsive areas in cortex as our literature review shows (almost all researchers use either a passive viewing or 1-back task). With that said, we find that task demands do not matter much in terms of localizing face responsive areas in FFA as evidenced by our literature review and two experiments. However, comparison stimuli do matter in terms of magnitude, extent and the reliability of activity over the ROI (as found with Pearson correlations), so a researcher should take care to use the same or similar contrasting stimuli for localizer tasks and tasks of interest.

Second, we are not claiming that no differences exist for different localizer tasks. For example, one can obtain a more reliable, robust, and larger activation cluster depending on task and contrasting stimuli. From our data it seems that faces vs. scrambles provide a larger and more robust activation cluster in the FFA compared to faces vs. houses. As such, faces vs. scrambles also produced more variance than faces vs. houses so a researcher must determine which variable is more important (reliability or activation magnitude). Additionally, if the task is more attentionally demanding, like the 1-back task, the activation patterns are more reliable as measured by Pearson correlations. These are important considerations when choosing a localizer task or choosing any task to explore in the magnet. These results also speak to Friston et al.'s (2006) concerns with using functional localizers. While the same or similar areas of cortex were active for different task demands, magnitudes, extents, and reliabilities of activation patterns may vary based on task context and may interact with various experimental manipulations (e.g. task and comparison stimuli interacted in Experiment 1 for magnitude and extent).

Third, it is important to consider what constitutes a significant difference. For example, if one finds a statistically significant difference in the location of the peak activated voxel for two different tasks, does that mean that the two tasks activated different regions? This is why we believe that using multivoxel pattern/correlation analyses (using Pearson correlation coefficients in this experiment) will become increasingly important, especially given the noise in fMRI data. It seems that when we consider whole patterns, we may find more similarities in our datasets than we might have imagined. Therefore, potentially important differences could emerge, such as differences in activation patterns as a function of different contrasting stimuli. Additionally and importantly, these correlation analyses provide a good measure of the reliability of our localizer tasks. In practice, functional localizers are typically employed by finding peaks (and defining a spherical region of interest around the peak) or thresholding activation maps. So the use of multivoxel correlation analysis at this time may be restricted to evaluating the reliability/reproducibility of the functional localizers rather than defining functionally localized areas. In the future, however, researchers may move to multivariate approaches when defining localized areas.

Lastly, it is not clear whether these results would generalize to other cortical regions, e.g., the visual word form area (McCandliss et al., 2003), or the extrastriate body area (Downing et al., 2001) though we do not see any compelling reason for many of these same results not to be found for other ROIs as well. In addition, while we focused on the fusiform gyrus, other brain areas were also active for the experimental conditions. Therefore, we are not advocating that researchers look only within ROIs when performing their fMRI experiments. For example, we found more PFC activations with increasing load in the *n*-back task, and we also found amygdala activations for face vs. house contrasts. The aim of this paper was to test whether various experimental manipulations alter the locale of FFA activations, not to restrict analyses to ROIs.

In sum, a researcher interested in using the functional localizer approach to localize the FFA should consider matching contrasting stimuli to the experiment of interest, but different localizer tasks seem to activate the same or similar regions of the FFA, which may alleviate many concerns with using the localizer approach (Saxe et al., 2006;

**Table 5**
Magnitude and extent results for Experiment 2. Magnitudes are expressed in *t*-stats and extent values are the number of voxels above a $t=2.5$.

| | Faces vs. Houses | | Faces vs. Scrambles | |
|---|---|---|---|---|
| | Magnitude | Extent | Magnitude | Extent |
| Left ROI | 5.42 (Passive) | 51.68 (Passive) | 9.61 (Passive) | 201.13 (Passive) |
| | 5.24 (1-back) | 41.74 (1-back) | 11.50 (1-back) | 286.31 (1-back) |
| Right ROI | 4.93 (Passive) | 50.46 (Passive) | 8.61 (Passive) | 181.15 (Passive) |
| | 4.84 (1-back) | 35.19 (1-back) | 11.34 (1-back) | 287.00 (1-back) |

Author's personal copy



Friston et al., 2006). If the researcher, however, is using the localizer approach not to define an area from a location standpoint, but to make additional claims, then the concerns of Friston et al. (2006) come into play, because task demands and contrasting stimuli did interact in terms of the magnitude, extent and the patterns/reliabilities of activations in FFA.

## Acknowledgments

We would like to thank Kripa Thummala and Keith Newnham for helping to collect our fMRI data. This research was supported in part by a grant from the National Science Foundation (NSF; #0520992), in part by a grant from NIMH (MH 60655) and in part by an NSF Graduate Research Fellowship to the first author.

## References

Aguirre, G.K., Singh, R., D'Esposito, M., 1999. Stimulus inversion and the responses of face and object-sensitive cortical areas. Neuroreport 10 (1), 189–194.
Andrews, T.J., Ewbank, M.P., 2004. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. NeuroImage 23 (3), 905–913.
Andrews, T.J., Schluppeck, D., 2004. Neural responses to money images reveal a modular representation of faces in human visual cortex. NeuroImage 21 (1), 91–98.
Avidan, G., Levy, I., Hendler, T., Zohary, E., Malach, R., 2003. Spatial vs. object specific attention in high-order visual areas. NeuroImage 19 (2), 308–318.
Caldara, R., Seghier, M.L., Rossion, B., Lazeyras, F., Michel, C., Hauert, C.A., 2006. The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. NeuroImage 31 (1), 313–319.
Carlson, T., Grol, M.J., Verstraten, F.A.J., 2006. Dynamics of visual recognition revealed by fMRI. NeuroImage 32 (2), 892–905.
Casanova, R., Srikanth, R., Baer, A., Laurienti, P.J., Burdette, J.H., Hayasaka, S., Flowers, L., Wood, F., Maldjian, J.A., 2007. Biological parametric mapping: a statistical toolbox for multimodality brain image analysis. NeuroImage 34 (1), 137–143.
Chen, C.C., Kao, K.L.C., Tyler, C.W., 2007. Face configuration processing in the human brain: the role of symmetry. Cerebral Cortex 17 (6), 1423–1432.
Clark, V.P., Parasuraman, R., Keil, K., Kulansky, R., Fannon, S., Maisog, J.M., et al., 1997. Selective attention to face identity and color studied with fMRI. Human Brain Mapping 5 (4), 293–297.
Courtney, S.M., Ungerleider, B.G., Keil, K., Haxby, J.V., 1997. Transient and sustained activity in a distributed neural system for human working memory. Nature 386 (6625), 608–611.
Dinstein, I., Gardner, J.L., Jazayeri, M., Heeger, D.J., 2008. Executed and observed movements have different distributed representations in human aIPS. The Journal of Neuroscience 28, 11231–11239.
Dove, A., Manly, T., Epstein, R., Owen, A.M., 2008. The engagement of mid-ventrolateral prefrontal cortex and posterior brain regions in intentional cognitive activity. Human Brain Mapping 29 (1), 107–119.
Downing, P.E., Jiang, Y.H., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293 (5539), 2470–2473.
Downing, P.E., Chan, A.W.Y., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. Cerebral Cortex 16 (10), 1453–1461.
Downing, P.E., Wiggett, A.J., Peelen, M.V., 2007. Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. Journal of Neuroscience 27, 226–233.
Druzgal, T.J., D'Esposito, M., 2001. Activity in fusiform face area modulated as a function of working memory load. Cognitive Brain Research 10 (3), 355–364.
Duncan, K.J., Pattamadilok, C., Knierim, I., Devlin, J.T., 2009. Consistency and variability in functional localisers. NEUROIMAGE 46 (4), 1018–1026.
Eger, E., Schyns, P.G., Kleinschmidt, A., 2004. Scale invariant adaptation in fusiform face-responsive regions. NeuroImage 22 (1), 232–242.
Eger, E., Schweinberger, S.R., Dolan, R.J., Henson, R.N., 2005. Familiarity enhances invariance of face representations in human ventral visual cortex: FMRI evidence. NeuroImage 26 (4), 1128–1139.
Epstein, R.A., Higgins, J.S., Parker, W., Aguirre, G.K., Cooperman, S., 2006. Cortical correlates of face and scene inversion: a comparison. Neuropsychologia 44 (7), 1145–1158.
Fang, F., Murray, S.O., He, S., 2007. Duration-dependent fMRI adaptation and distributed viewer-centered face representation in human visual cortex. Cerebral Cortex 17 (6), 1402–1411.
Friston, K.J., Henson, R.N., 2006. Commentary on: divide and conquer; a defence of functional localisers. NeuroImage 30 (4), 1097–1099.
Friston, K.J., Rotshtein, P., Geng, J.J., Sterzer, P., Henson, R.N., 2006. A critique of functional localisers. NeuroImage 30 (4), 1077–1087.
Ganel, T., Valyear, K.F., Goshen-Gottstein, Y., Goodale, M.A., 2005. The involvement of the "fusiform face area" in processing facial expression. Neuropsychologia 43 (11), 1645–1654.
Gathers, A.D., Bhatt, R., Corbly, C.R., Farley, A.B., Joseph, J.E., 2004. Developmental shifts in cortical loci for face and object recognition. Neuroreport 15 (10), 1549–1553.
Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 1999. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nature Neuroscience 2 (6), 568–573.
Gauthier, I., Anderson, A.W., Skudlarski, P., Gore, J.C., 2000a. Expertise for cars and birds recruits right hemisphere face areas. Journal of Cognitive Neuroscience 19C.
Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000b. Expertise for cars and birds recruits brain areas involved in face recognition. Nature Neuroscience 3 (2), 191–197.
Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000c. The fusiform "face area" is part of a network that processes faces at the individual level. Journal of Cognitive Neuroscience 12 (3), 495–504.
George, N., Dolan, R.J., Fink, G.R., Baylis, G.C., Russell, C., Driver, J., 1999. Contrast polarity and face recognition in the human fusiform gyrus. Nature Neuroscience 2 (6), 574–580.
Gilaie-Dotan, S., Perry, A., Bonneh, Y., Malach, R., Bentin, S., 2009. Seeing with profoundly deactivated mid-level visual areas: non-hierarchical functioning in the human visual cortex. Cerebral Cortex 19 (7), 1687–1703.
Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. Nature Neuroscience 7 (5), 555–562.
Hadjikhani, N., de Gelder, B., 2003. Seeing fearful body expressions activates the fusiform cortex and amygdala. Current Biology 13 (24), 2201–2205.
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 (5539), 2425–2430.
Henson, R.N., Mouchlianitis, E., 2007. Effect of spatial attention on stimulus-specific haemodynamic repetition effects. NeuroImage 35 (3), 1317–1329.
Horovitz, S.G., Rossion, B., Skudlarski, P., Gore, J.C., 2004. Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. NeuroImage 22 (4), 1587–1595.
Iacoboni, M., Lieberman, M.D., Knowlton, B.J., Molnar-Szakacs, I., Moritz, M., Throop, C.J., et al., 2004. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. NeuroImage 21 (3), 1167–1173.
Iidaka, T., Yamashita, K., Kashikura, K., Yonekura, Y., 2004. Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe: an investigation with event-related fMRI. Cognitive Brain Research 18 (2), 196–204.
Ishai, A., Bikle, P.C., Ungerleider, L.G., 2006. Temporal dynamics of face repetition suppression. Brain Research Bulletin 70 (4-6), 289–295.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17 (2), 825–841.
Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nature Neuroscience 8, 679–685.
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 (11), 4302–4311.
Kesler-West, M.L., Andersen, A.H., Smith, C.D., Avison, M.J., Davis, C.E., Kryscio, R.J., et al., 2001. Neural substrates of facial emotion processing using fMRI. Cognitive Brain Research 11 (2), 213–226.
Kung, C.C., Peissig, J.J., Tarr, M.J., 2007. Is region-of-interest overlap comparison a reliable measure of category specificity. Journal of Cognitive Neuroscience 19 (12), 2019–2034.
Lazar, N.A., Eddy, W.F., Genovese, C.R., Welling, J., 2001. Statistical issues in fMRI for brain imaging. International Statistical Review 69 (1), 105–127.
Lehmann, C., Mueller, T., Federspiel, A., Hubl, D., Schroth, G., Huber, O., et al., 2004. Dissociation between overt and unconscious face processing in fusiform face area. NeuroImage 21 (1), 75–83.
Leibenluft, E., Gobbini, M.I., Harrison, T., Haxby, J.V., 2004. Mothers' neural activation in response to pictures of their children and other children. Biological Psychiatry 56 (4), 225–232.
Loffler, G., Yourganov, G., Wilkinson, F., Wilson, H.R., 2005. fMRI evidence for the neural representation of faces. Nature Neuroscience 8 (10), 1386–1390.
Lund, T.E., Norgaard, M.D., Rostrup, E., Rowe, J.B., Paulson, O.B., 2005. Motion or activity: their role in intra- and inter-subject variation in fMRI. NeuroImage 26 (3), 960–964.
Maurer, D., O'Craven, K.M., Le Grand, R., Mondloch, C.J., Springer, M.V., Lewis, T.L., et al., 2007. Neural correlates of processing facial identity based on features versus their spacing. Neuropsychologia 45 (7), 1438–1451.
McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends in Cognitive Sciences 7 (7), 293–299.
O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. Journal of Cognitive Neuroscience 12 (6), 1013–1023.
Oppenheim, A.V., Schafer, R.W., Buck, J.R., 1999. Discrete-Time Signal Processing 2nd ed. Prentice-Hall, Upper Saddle River, NJ.
Peelen, M.V., Downing, P.E., 2005. Within-subject reproducibility of category-specific visual activation with functional MRI. Human Brain Mapping 25 (4), 402–408.
Poldrack, R.A., 2006. Can cognitive processes be inferred from neuroimaging data. Trends in Cognitive Sciences 10 (2), 59–63.
Poldrack, R.A., 2007. Region of interest analysis for fMRI. Social Cognitive and Affective Neuroscience 2 (1), 67–70.
Polk, T.A., Park, J., Smith, M.R., Park, D.C., 2007. Nature versus nurture in ventral visual cortex: a functional magnetic resonance imaging study of twins. Journal of Neuroscience 27 (51), 13921–13925.

Author's personal copy



Pourtois, G., Schwartz, S., Seghier, M.L., Lazeyras, F.O., Vuilleumier, P., 2005a. View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMI study. NeuroImage 24 (4), 1214–1224.
Pourtois, G., Schwartz, S., Seghier, M.L., Lazeyras, F., Vuilleumier, P., 2005b. Portraits or people? Distinct representations of face identity in the human visual cortex. Journal of Cognitive Neuroscience 17 (7), 1043–1057.
Puce, A., Allison, T., Gore, J.C., Mccarthy, G., 1995. Face-sensitive regions in human extrastriate cortex studied by functional mri. Journal of Neurophysiology 74 (3), 1192–1199.
Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. Journal of Neuroscience 16 (16), 5205–5215.
Reddy, L., Moradi, F., Koch, C., 2007. Top–down biases win against focal attention in the fusiform face area. NeuroImage 38, 730–739.
Reinholz, J., Pollmann, S., 2005. Differential activation of object-selective visual areas by passive viewing of pictures and words. Cognitive Brain Research 24 (3), 702–714.
Rotshtein, P., Vuilleumier, P., Winston, J., Driver, J., Dolan, R., 2007. Distinct and convergent visual processing of high and low spatial frequency information in faces. Cerebral Cortex 17 (11), 2713–2724.
Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. NeuroImage 30 (4), 1088–1096.
Serences, J.T., Schwarzbach, J., Courtney, S.M., Golay, X., Yantis, S., 2004. Control of object-based attention in human cortex. Cerebral Cortex 14 (12), 1346–1357.
Sergent, J., Ohta, S., Macdonald, B., 1992. Functional neuroanatomy of face and object processing—a positron emission tomography study. Brain 115, 15–36 Part 1.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208–S219.
von Kriegstein, K., Kleinschmidt, A., Sterzer, P., Giraud, A.L., 2005. Interaction of face and voice areas during speaker recognition. Journal of Cognitive Neuroscience 17 (3), 367–376.
Yi, D.J., Kelley, T.A., Marois, R., Chun, M.M., 2006. Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. Brain Research 1080, 53–62.
Zhang, H., Liu, J., Huber, D.E., Rieth, C.A., Tian, J., Lee, K., 2008. Detecting faces in pure noise images: a functional MRI study on top-down perception. Neuroreport 19 (2), 229–233.