Is Region-of-Interest Overlap Comparison a Reliable Measure of Category Specificity?

Chun-Chia Kung[1,2,*], Jessie J. Peissig[1,3], and Michael J. Tarr[1]

Abstract

■ Analysis of the degree of overlap between functional magnetic resonance imaging–derived regions of interest (ROIs) has been used to assess the functional convergence and/or segregation of category-selective brain areas. An examination of the extant literature reveals no consistent usage for how such overlap is calculated, nor any systematic comparison between different methods. We argue that how ROI overlap is computed, especially the choice of the denominator in the formula, can profoundly affect the results and interpretation of such an analysis. To do this, we compared the overlap of the FFA–FFA (fusiform face area) and FFA–FGA (fusiform Greeble-selective area) in a localizer study testing both Greeble novices and experts. When using a single ROI as the denominator, we found a significant difference in FFA–FFA versus FFA–FGA overlap, consistent with the result of a previous study arguing for face specificity of the FFA [Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. Is the fusiform face area specialized for faces, individuation, or expert individuation? *J Cogn Neurosci, 16,* 189–203, 2004]. However, these ROI overlap differences disappeared when the denominator combined both of the involved ROIs, and the patterns of such overlap comparisons were dependent on given statistical thresholds. We also found proportionally decreasing FFA–FFA overlap with increasing center-of-FFA distance, resolving an apparent contradiction between the consistency of the location of the FFA and the seemingly low FFA–FFA overlap. Finally, Monte Carlo simulations revealed the most stable formula—the most resistant to ROI size variations—to be the *average* of the two single-ROI-denominator-based overlap indices. In sum, ROI overlap analysis is not a reliable tool for assessing category specificity, and caution should be exercised with regard to ROI overlap definition, underlying assumptions, and interpretation. ■

INTRODUCTION

For a wide variety of theoretical reasons, measures of the convergence of functional magnetic resonance imaging (fMRI)-derived activations across different task conditions, or *cognitive conjunction* (Friston, Holmes, Price, Buchel, & Worsley, 1999; Price & Friston, 1997), has been widely used in the neuroimaging field. To give just a few examples, researchers have identified common activations for perception and imagery (Kosslyn, Ganis, & Thompson, 2001; Mellet, Petit, Mazoyer, Denis, & Tzourio, 1998), object and motion perception (Kourtzi, Bulthoff, Erb, & Grodd, 2002), primary and secondary language processing in bilinguals (Chee et al., 2000), and between working memory and spatial attention networks (Hopfinger, Buonocore, & Mangun, 2000; LaBar, Gitelman, Parrish, & Mesulam, 1999). Despite its popularity, conjunction analysis has recently been challenged regarding the degree of common cognitive process involved (Caplan & Moo, 2004) and issues regarding the statistical assumptions (Friston, Penny, & Glaser, 2005;

Nichols, Brett, Andersson, Wager, & Poline, 2005). These discussions have greatly enriched our understanding and raised practical concerns about how to best perform and interpret conjunction analyses.

In a similar vein, the present study focuses on a particular form of conjunction analysis: the region-of-interest (ROI) overlap comparison. Specifically, an ROI overlap index is the number of overlapping voxels between two ROIs divided by a common denominator. By comparing the overlap percentage across different tasks or conditions, a number of studies have used this overlap index difference to infer the degree of convergence or segregation between functionally defined ROIs. For example, O'Craven and Kanwisher (2000) measured the percentage of overlap of the fusiform face area (or FFA; Kanwisher, McDermott, & Chun, 1997) derived from participants imagining faces and from participants actually viewing faces. They reported a high degree of overlapping voxels in the FFA between the two conditions, and argued that this result supported a shared representation for face perception and face imagery.

In another study, Grill-Spector, Knouf, and Kanwisher (2004) compared the FFA derived from a localizer scan (passive viewing of faces vs. objects) and from an identification task using a variety of stimuli, such as faces, birds, flowers, houses, guitars, and cars. In their

[1]Brown University, [2]Princeton University, [3]California State University, Fullerton
*Current Address: Department of Psychology, Princeton University, Princeton, NJ 08544.

study, two overlap indices were calculated. Based on our interpretation of Table 1 (p. 559), the numerator was the number of overlapping voxels between the FFA from the localizer and the specific category that was tested; the denominator was either the voxels for the FFA localizer or the FFA task. Both indices showed significantly higher overlap between the FFA localizer and the FFA task compared to the overlap between the localizer and the other task-derived object-selective areas. However, their ROI overlap comparison was derived from novices' data ($n = 5$), so whether car experts, for example, might show higher overlap between FFA and fusiform car-selective areas was unclear.

With similar results to those of Grill-Spector et al. (2004), Rhodes, Byatt, Michie, and Puce (2004) reported the overlap between the FFA and a fusiform Lepidoptera area (Lepidoptera are commonly know as butterflies and moths), or FLA, in both Lepidoptera novices (Experiment 1) and experts (Experiment 2). They scanned people with an initial FFA localizer (or LO), which consisted of passive viewing of faces and objects, followed by a similar passive viewing (or PV) of faces, objects, and Lepidoptera, and a subsequent individuation (or IN) task, requiring participants to recognize the items they viewed in the previous session (half of them were from the previous session or "old"). The FFA–FLA overlap index was defined as the overlap between face- and Lepidoptera-selective voxels in the fusiform gyrus (FG) divided by the face-selective voxels in the FG or (FFA ∩ FLA)/FFA.[1] For the purpose of comparison, the FFA–FFA overlap across the three tasks (LO, PV, and IN) was divided into two values, $(FFA_{pv} \cap FFA_{lo})/FFA_{pv}$ and $(FFA_{pv} \cap FFA_{in})/FFA_{pv}$, and was calculated for each participant to provide the baseline overlap index, which was a mean of 26.2% in Experiment 1 and 28.3% in Experiment 2 (across both left and right hemispheres, Table 1, p. 194). In comparison, the average FFA–FLA overlap, or the mean of $(FFA_{pv} \cap FLA_{pv})/FFA_{pv}$ and $(FFA_{in} \cap FLA_{in})/FFA_{in}$, across both hemispheres, was 10.7% in Experiment 1 and 6.7% in Experiment 2. The significant difference between the FFA–FFA and FFA–FLA overlap was taken as support for the face specificity hypothesis (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Kanwisher, 2000).

In addition to the above studies, which are directly related to the decade-long debate between the face specificity and the perceptual expertise hypotheses (Gauthier & Bukach, 2007; Bukach, Gauthier, & Tarr, 2006; Kanwisher, 2006; McKone & Robbins, 2006; McKone & Kanwisher, 2005; Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Kanwisher et al., 1997), a recent commentary (Peelen & Downing, 2005a) also reported the ROI overlap of imagined extrastriate action-related area (ARA; Astafiev, Stanley, Shulman, & Corbetta, 2004) and the extrastriate body-part area (EBA; Downing, Jiang, Shuman, & Kanwisher, 2001) as only 14%, implying a

possible segregation between ARA and EBA. In their later reply, Astafiev et al. (2004) pointed out that a low index may be biased by the use of the larger ROI (here EBA) as the denominator. They suggest that the overlap may have been larger if the smaller ARA was used. This raises two important concerns in interpreting ROI overlap results: (1) the percentage of overlap should be compared to an appropriate baseline condition, just as the FLA–FFA overlap was compared to the repeated FFA overlap used in the Rhodes et al. (2004) study; (2) the choice of the denominator can greatly affect the calculated overlap percentage. This second point will be elaborated in the next section.

The Choice of the Denominator in ROI Overlap Comparison

Although the ROI overlap has been used in a number of studies for a variety of purposes, a careful examination of the extant studies that adopted the ROI overlap analyses reveals one surprising finding: Although the numerator used in the overlap equation was *always* the total number of overlapped voxels between the two ROIs, the chosen denominator varied from study to study. As mentioned earlier, Peelen and Downing (2005a), Grill-Spector et al. (2004), Rhodes et al. (2004), and O'Craven and Kanwisher (2000) used a single ROI as the denominator, whereas other studies used the sum (Ganis, Thompson, & Kosslyn, 2004), the mean (Nieto-Castanon, Ghosh, Tourville, & Guenther, 2003), or the union (Sawamura, Georgieva, Vogels, Vanduffel, & Orban, 2005) of the two ROIs as overlap denominators. In most of these studies, the reasons for the use of a particular denominator were not discussed, nor were alternatives considered. In a real example of how the choice of denominator can drastically affect the overlap ratio, we recalculated the data from Table 1 of O'Craven and Kanwisher (2000, p. 1016). This study compared the overlap for imagining and viewing faces. When the voxel overlap (mean size = 9) was divided by the face imagery condition (mean size = 11 voxels), the mean overlap was 84%. However, when the voxel overlap was divided by the face perception condition (mean size = 63 voxels), the average overlap dropped to 22%. Therefore, it is reasonable to infer that in the Rhodes et al. (2004) study, the low overlap value (<30%) may have been due to the use of the comparatively larger ROI as the denominator; they reported a larger FFA_{pv} than FFA_{lo} (1440 vs. 813 mm^3) for the FFA–FFA overlap, and a larger FFA than FLA in both the FFA_{pv}–FLA_{pv} and FFA_{in}–FLA_{in} overlaps (1440 vs. 650 and 1759 vs. 1346 mm^3, respectively).

Agreeing that the volume (e.g., FFA vs. FLA) differences may account for some variation in the degree of overlap between faces and Lepidoptera, Rhodes et al. (2004) emphasized that "…it cannot account for the relatively modest degree of overlap between faces and

Lepidoptera (7%) compared with the two sets of face stimuli (27%)'' (p. 194). Indeed, although the choice of a single larger ROI as the denominator seems unable to explain the consistently larger FFA–FFA overlap compared to FFA–FLA overlap, the *magnitude* of the overlap difference (27% vs. 7%) can surely be adjusted by the denominator chosen. Figure 1 shows the results of three different ROI overlap indices using data from the Rhodes et al. report. When the choice of denominator changes from a single ROI (what Rhodes et al. used) to either the union or the sum of the two ROIs, the overlap *difference* drops from 20% to 16% to 13%, respectively. Although a 4–7% decrease in overlap difference may appear trivial, it does call for the need to systematically examine the effect of the denominator in the ROI overlap comparison, and to evaluate which formula might fare best in the current ROI overlap scenarios in assessing functional selectivity of the ROI.

How Consistent is the FFA (across Runs and Sessions)?

Another interesting issue raised in the Rhodes et al. (2004) study was their somewhat contradictory finding of low FFA–FFA overlap (27%), and a high consistency in the location of the FFA across tasks, runs, and sessions reported in the literature (Peelen & Downing, 2005c; Grill-Spector et al., 2004; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Kanwisher et al., 1997). Rhodes et al. state that ''given the apparent robust consistency of previous research demonstrating activation of FG [fusiform gyrus] to faces, it should be noted that the measure of overlap

used here is a conservative index as it is the degree of overlap in activated regions thresholded to the same conservative criterion for both tasks'' (p. 194). First, it is not clear why their uncorrected p value thresholds of .001 should be considered ''conservative'' compared to other published FFA studies which used p value thresholds of .0001 (uncorrected) or less (Kanwisher, Tong, & Nakayama, 1998; Kanwisher et al., 1997). Second, even if the threshold was indeed conservative, further explanation is needed regarding how statistical threshold affects the FFA–FFA overlap, including whether a lenient statistical threshold elevates the ROI overlap and whether a more stringent threshold decreases it, and if so, by how much. It is unclear whether changes in the threshold would have any substantial effect on the overlap index. Any increase of the threshold, say from $p = .05$ to $p = .10$, will likewise increase the sizes of both ROIs in the denominator and their overlap in the numerator (and similarly, a decrease in the threshold would also be perpetuated equally), in many cases, rendering the final outcome equipotent.

In the current study, we will (1) further explore the relationship between the FFA sizes, distances, and their resulting overlaps (defined by various formulae); and (2) compare the FFA–FFA and FFA–FXA overlap (X being the object of expertise category) under several statistical thresholds to see whether there is any systematic relationship. The latter point not only addresses the use of different statistical thresholds and the effect on the low FFA–FFA overlap specifically but also more generally tests the effect of statistical criterion on various ROI overlap comparison results.

Figure 1. Illustrations of how ROI percentage overlap in Rhodes et al. (2004) varies with various denominators: (1) single ROI (FFA from the passive viewing session, or FFA$_{pv}$), (4) the other single ROI (FFA from the localizer session, or FFA$_{lo}$), (2) the union, and (3) the sum of FFA$_{pv}$ and FFA$_{lo}$. As shown, the differences of the ROI overlap (FFA–FFA minus FFA–FLA overlap) decreased from 20% (27%−7%) in (1), to 16% (21%−5%) in (2), or 13% (18%−5%) in (3). In addition, the FFA–FFA overlap increased from the original 27% to 49% by just switching the denominator from FFA$_{pv}$ to FFA$_{lo}$. These FFA values were derived from Table 1 of Rhodes et al. (2004, p. 194).

Mean ROI overlap comparison in Rhodes et al. (2004)

FLA$_{pv}$ 650 mm^3

FFA$_{pv}$ 1440 mm^3

FFA$_{pv}$ ∩ FLA$_{pv}$ 100 mm^3

FFA$_{pv}$ ∩ FFA$_{lo}$ 400 mm^3

FFA$_{lo}$ 813 mm^3

Denominator: (1) single ROI

$$\text{FFA}_{pv} - \text{FFA}_{lo} \text{ overlap} = \frac{400}{1440} = 27\%$$

$$\text{FFA}_{pv} - \text{FLA}_{pv} \text{ overlap} = \frac{100}{1440} = 7\%$$

(2) union

$$\frac{400}{1440 + 813 - 400} = 21\%$$

$$\frac{100}{1440 + 650 - 100} = 5\%$$

(3) sum

$$\frac{400}{1440 + 813} = 18\%$$

$$\frac{100}{1440 + 650} = 5\%$$

(4) FFA$_{lo}$

$$\text{FFA}_{pv} - \text{FFA}_{lo} \text{ overlap} = \frac{400}{813} = 49\%$$

Goal of the Present Study

Based on the above review and discussion, the goal of the present study is to examine: (1) the effect of the denominator choice (or the formula used) on ROI overlap comparisons; (2) the relationship between FFA location and the overlap index, and (3) how ROI overlap is affected by the choice of statistical threshold. To achieve this, we tested 10n Greeble novices with three runs of a localizer scan (consisting of blocks of faces, objects, and Greebles). Five of these participants were subsequently given 2 weeks of training to become "Greeble experts" (Gauthier & Tarr, 1997). These five laboratory-trained experts were further scanned in two additional sessions with the same localizer, once during training and again immediately following training (training procedure = 2 weeks; Rossion, Kung, & Tarr, 2004; Gauthier, Williams, Tarr, & Tanaka, 1998). With the data we sought to (1) replicate previous results using an overlap analysis similar to the one used by Rhodes et al. (2004); (2) further examine six different denominators and compare their results; (3) reconcile the issue of FFA location consistency across runs and sessions and corresponding FFA–FFA overlap measures; and (4) evaluate the effect of varying statistical thresholds on ROI overlap formula and their values.

In our assessment of FFA–FFA overlap, we compared the same task, face stimuli, and epoch length across three separate scanning sessions. We expected no difference in the FFA size across scans and predicted no differences regardless of which single FFA served as the denominator. In addition, we predicted that the size of the FFA would be larger than the FGA ("fusiform Greeble-selective area"), and this size difference would lead to inconsistencies in the ROI overlap measure, dependent on the particular formula used. More specifically, when using only the larger face area as the denominator, we predicted that the FFA–FFA overlap would be significantly larger than the FFA–FGA overlap, consistent with the face specificity prediction. However, when a combination of both FFA and FGA are used as the denominator, we predicted that the resulting mean overlap between FFA–FFA and FFA–FGA would not be significantly different, consistent with the prediction of the perceptual expertise hypothesis. To test these predictions, we used six different denominators. The numerator was the same across all formulae, that is, the number of overlapping voxels (represented as $ROI_a \cap ROI_b$ in the formulae below).

$$(a)(ROI_a \cap ROI_b)/ROI_a$$

$$(b)(ROI_a \cap ROI_b)/ROI_b$$

For formulae a and b, the number of voxels for a single category is used as the denominator. For example, the number of active voxels in the FG calculated for a face localizer task could be used (ROI_a), which might be measured as 1000 voxels. Alternatively, the number of active voxels for a Greeble localizer task could be used as the denominator (ROI_b), which might be measured as 50 voxels. If our number of overlapping voxels for faces and Greebles was 25 voxels, this would yield an overlap index of 2.5% in the case of formula a, and 50% in the case of formula b.

$$(c)(ROI_a \cap ROI_b)/(ROI_a + ROI_b)$$

For formula c, the denominator is the sum of the active voxels for two categories. For example, if there were 1000 voxels for faces and 50 for Greebles, the denominator would be 1050 voxels. With 25 overlapping voxels, this would yield a percent overlap of 2.38% using formula c.

$$(d)(ROI_a \cap ROI_b)/[(ROI_a + ROI_b)/2]$$

In formula d, the mean volume of two ROIs is calculated. For example, if there were 1000 voxels for faces and 50 for Greebles, the denominator would be 525 voxels. Using our example of 25 overlapping voxels, this would yield a percent overlap calculation of 4.76% in the case of formula d.

$$(e)(ROI_a \cap ROI_b)/(ROI_a \cup ROI_b)$$

In formula e, the union of both ROIs is calculated and used as the denominator. This differs from the sum of voxels in that the shared voxels for Greebles and faces are not counted twice. For example, if the face area was 1000 voxels and the Greeble area was 50 voxels, but 25 of these voxels are shared, then the denominator would be 1025 voxels (975 unique face voxels + 25 unique Greeble voxels + 25 shared voxels). Using our example of 25 overlapping voxels again, this would yield a percent overlap calculation of 2.43% using formula e.

$$(f)[((ROI_a \cap ROI_b)/ROI_a) + ((ROI_a \cap ROI_b)/ROI_b)]/2$$

Formula f is the average of formulae a and b. Thus, for this formula, we take the mean overlap index for formula a (2.5%) and formula b (50%), yielding an overlap index of 26.25%. From these examples, we can see that the overlap index varies widely with the use of different denominators. However, the critical issue here is how much these different overlap denominators affect the comparison of the overlap for repeated measures of the face-selective area (e.g., FFA–FFA overlap) versus the overlap between face-selective and another category-selective area, in this case, Greebles (e.g., FFA–FGA overlap). We examined each of these formulae using fMRI data for Greebles and faces.

METHODS

Participants

Ten undergraduate or graduate students at Brown University participated in the one-session scan (6 women, 1 left-handed, mean age = 25.9 years, SD = 4.2 years). Five of them (3 women, mean age = 25 years, SD = 4.74 years) continued with two additional scans, a week apart, during the middle and end of 2 weeks of Greeble expertise training. Behavioral performances in both picture-naming and picture-name verification tasks were monitored to make sure the participants achieved expertise criterion, for example, no significant difference between basic- and subordinate-level objects in accuracy and response time (Tanaka & Taylor, 1991). Participants typically achieved criterion by the eighth or ninth training session (Rossion et al., 2004). All the participants gave informed consent as approved by the Brown University Institutional Review Board (IRB).

Stimuli

The stimuli consisted of 90 full-color Caucasian front-view face photographs (half male), 100 full-color object photos (sampled from commercial object image CDs), and 80 asymmetric Greeble images created using 3-D Studio Max R4 (www.discreet.com). The complete set of the Greeble stimuli, both images and 3-D models, is available at www.tarrlab.org/. The asymmetric Greebles were created by randomly repositioning the ''boges,'' ''quiff,'' and ''dunth'' within a particular placement range. In interviews before and after training, participants indicated that the asymmetric Greebles did not appear face-like. The stimuli, which subtended approximately 8° of visual angle, were projected onto an opaque screen and were viewed by participants through a mounted mirror on the head coil.

fMRI Procedure

Each scan session began with an MPRAGE structural scan (1 mm³) and followed by three runs of functional FFA localizers. The FFA localizers consisted of blocks of faces, objects, and Greebles (20 images per block), interleaved with the fixation baseline (Figure 2). Participants were instructed to press a button with their dominant hand when they detected a stimulus repetition (1−back identity task). Forty-five 3-mm axial EPI slices (in-plane resolution = 3 × 3 mm², TR = 3600 msec, TE = 38 msec, flip angle = 90°, gap = 0 mm, 76 measurements) were acquired in the 1.5-T Siemens scanner located at Memorial Hospital of Rhode Island.

fMRI Data Analysis

We used BrainVoyager 2000 v4.96 (Brain Innovation) to (1) discard the first two volumes in each run; (2)
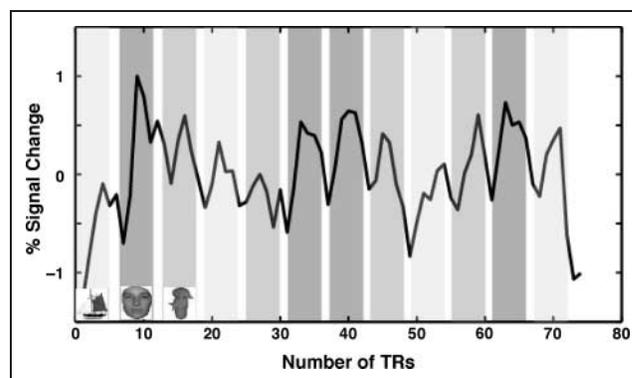

Figure 2. A representative subject's time course and sample stimuli from our FFA localizer study. Horizontal axis denotes TR number, vertical axis denotes percent signal change.

preprocess using 3-D motion correction, slice scan time correction (interleaved ascending), linear trend removal, and high-pass filtering at 3 cycles/sec. Structural images were warped into Talairach space for group analysis. The volume time-course files were analyzed using a general linear model (Boynton, Engel, Glover, & Heeger, 1996) with three convolved predictors: faces, objects, and Greebles, corrected for autocorrelation (AR1), and restricted by the gray matter mask. Subsequent contrasts of faces minus objects (p = .05, corrected for main experiment, with p = .01 and .10 for additional analyses) were applied either to each run (for all 10 Greeble novices), or to each session collapsing over three runs (for the five Greeble experts). The face-selective voxel cluster, or FFA, was determined by confirming the co-localization of (1) a reasonable anatomical range (x TAL: 35 to 40; y TAL: −40 to −60; z TAL: −15 to −25), (2) a higher average BOLD response for faces than for other object categories (Figure 2), and (3) a significant mean percent signal change (PSC) of about 1% for faces compared to fixation, and a smaller PSC for objects and Greebles.

RESULTS

Within- and Between-session FFA Spatial Consistency

Our first analysis sought to reconcile the discrepancy between the low FFA–FFA overlap (M = 27%) reported by Rhodes et al. (2004), and the consistent within-subject FFA spatial location measured across runs and sessions commonly reported in the literature (Peelen & Downing, 2005c; Gauthier et al., 1999; Kanwisher et al., 1997). All the participants' FFA locations (center-of-mass Talairach coordinates) and size (mm³) are listed in Table 1. Examples of the spatial location of the FFA are shown for four participants in Figure 3. The top two rows show the data by run and the bottom two rows show data by

Table 1. FFA Volume (mm³) and the Center-of-Mass Talairach Coordinates (x, y, z) of the Present Experiment, Across-Run ($n = 10$, One Session) and Across-Session ($n = 5$, 3 Sessions)

(a)	FFA—run 1		FFA—run 2		FFA—run 3	
S	TAL (x, y, z)	Size (mm³)	TAL (x, y, z)	Size (mm³)	TAL (x, y, z)	Size (mm³)
1	38, −55, −11	218	38, −55, −10	105	40, −54, −13	308
2	44, −48, −17	140	45, −44, −19	83	46, −40, −21	1017
3	40, −54, −13	33	38, −52, −10	36	38, −53, −11	22
4	37, −43, −20	1301	35, −41, −22	1032	36, −35, −21	533
5	35, −52, −12	46	36, −56, −15	6	35, −51, −14	10
6	40, −45, −14	95	41, −45, −14	139	41, −43, −13	194
7	48, −46, −17	227	51, −42, −16	286		
8	44, −48, −17	26	45, −44, −19	102	46, −40, −21	482
9	42, −45, −9	2202	38, −42, −10	6	41, −46, −12	1678
10	41, −41, −23	733	42, −42, −23	729	41, −43, −23	1004
Average (n = 10)		502		252		583

(b)	FFA—session 1		FFA—session 2		FFA—session 3	
S	TAL (x, y, z)	Size (mm³)	TAL (x, y, z)	Size (mm³)	TAL (x, y, z)	Size (mm³)
1	39, −55, −12	346	38, −56, −13	641	40, −54, −11	365
2	44, −42, −21	1181	44, −40, −22	499	44, −44, −14	503
3	38, −52, −11	72	38, −50, −12	46	37, −60, −13	214
4	37, −39, −23	2206	32, −42, −21	2308	34, −41, −22	1784
5	35, −54, −12	71	32, −56, −13	409	37, −59, −14	252
6	41, −44, −14	370				
7	47, −41, −16	364				
8	39, −45, −13	618				
9	42, −44, −10	1679				
10	41, −43, −22	1431				
Average (n = 5)		775		780		623
Average (n = 10)		834				

session. Several points can be made about these data. First, there are considerable between-subject differences for the overlap indices, in both the between-run and between-session calculations. For example, using formula f, two participants exhibited overlap in the range of ∼60% to 70% (Rows 1 and 3), whereas the two remaining participants exhibited overlap of only ∼3% (Row 2) and ∼20% (Row 4). Second, as expected, the larger the FFA–FFA distance, the smaller the overlap index. This inverse relationship was significant for all the correlations between the FFA–FFA distances and the six overlap indices (a, b, c, d, e, and f), both across-run [$n = 10$, $r(29) = −.50, −.52, −.76, −.76, −.76$, and $−.66$; all

$p < .05$] and across-session [$n = 5, r(14) = −.64, −.80, −.78, −.78, −.74$, and $−.83$; all $p < .05$]. One example of the correlation results for a single formula (formula f) is shown in Figure 4. Third, the by-session analysis combined data over three runs, resulting in FFA sizes that were larger than the by-run analysis [$t(9) = 3.36, p = .008$ for 10 novices; and $t(14) = 2.71, p = .016$ for 5 experts who underwent 3 scans] and more consistent in size, which was indicated by the highly significant FFA–FFA volume correlation [$r(14) = .91, p < .0001$] combining all 15 possible pairings for the across-session analysis (5 participants × 3 sessions). In comparison, the across-run FFA volume correlation [$r(29) = .39, p > .05$] was

Figure 3. The axial slice map of the FFA location (Talairach coordinates and FFA sizes, in mm³, shown on top of each slice) of four representative subjects (upper two rows for within-session, lower-two rows for across-session analysis). With a revised overlap index (formula f) which averages two single-ROI-as-denominator overlap indices, the mean FFA–FFA overlap was: 64% and 3% for the upper two, and 67% and 27% for the bottom two rows. In general, our results are congruent with Peelen and Downing (2005b), who found that, for individual subjects across runs or sessions, the location and size of the FFA is relatively consistent, with some across-subject variation.

much smaller, possibly reflecting the combined effects of fewer data points per run and within-subject noise (including adaptation, fatigue, etc.), which jointly contribute to larger variability in FFA size across runs. Finally, the range of center-of-FFA Talairach coordinates are consistent with what has been reported in the literature: Across Tables 1 and 2, the coordinates are in the range of: $x = 35$ to 45, $y = -40$ to -60, $z = -10$ to -25.

Our FFA results, in terms of the location and size, were both within the ranges reported in the literature, therefore supporting the qualitative notion of FFA "consistency." However, the substantial differences in calculated overlap indices between individuals reveal several important factors related to overlap calculations. Consistency in the spatial location of the FFA can be linked to "small" center-of-FFA distance variations (e.g., ±5 mm in the y-axis) and high correlations among the calculated sizes of the FFA (at least across sessions). In contrast, the FFA–FFA overlap index was concurrently "low" (e.g., based on the equation shown in Figure 4, with 5 mm FFA-center distance, the expected mean overlap is ~35%), yielding the impression of a discrepancy between spatial location and overlap. This apparent "inconsistency" can be resolved by a quantitative comparison. We found a significant negative correlation between FFA-center distance and the corresponding
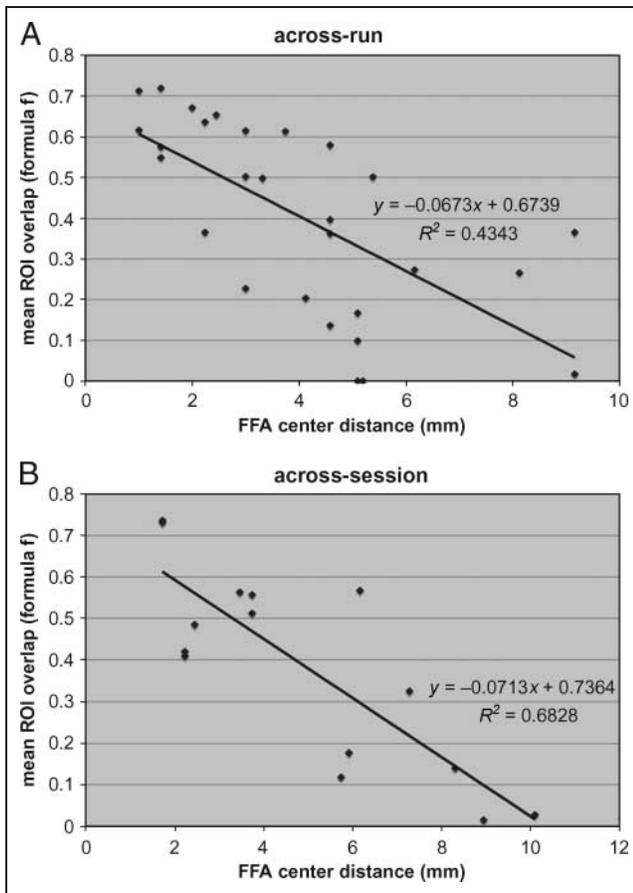
Figure 4. The significantly negative correlations between FFA-center distance and the corresponding FFA–FFA overlap (using formula f only here) in the current study. Upper row (A) represents across-run analysis ($n = 10$, each had 3 runs), and bottom row (B) represents across-session analysis ($n = 5$, each had 3 sessions). The mean FFA–FFA overlap decrease, based on the two regression equations, was about 10% per mm distance increase.

overlap, shown in Figure 4. These data indicate that, as you would expect, as the distance between the center points of calculated face-selective areas increases, the amount of overlap decreases. Our results, therefore, suggest that FFA consistency depends on what one defines as the measure of consistency, and where you place the cutoff for ''consistent'' versus ''inconsistent.'' If consistency is measured as center-of-FFA distance, several millimeters is thought to be trivial and data are interpreted as indicative of as a stable, functionally defined area. However, when measured as FFA–FFA overlap, the result appears much smaller (e.g., <50%), which might be interpreted as indicative of a less stable area, even though these measures are significantly correlated.

Comparison of FFA–FFA vs. FFA–FGA Overlap

As a basis for later comparisons with different overlap formulae, the first step of our FFA–FFA versus FFA–FGA comparison was to replicate previous results showing larger FFA–FFA overlap compared to the overlap between FFA and the category-selective responses arising from a second domain of expertise, including Lepidoptera (Rhodes et al., 2004) and cars (Grill-Spector et al., 2004). Here we used asymmetric Greebles, and created experts with a training paradigm that has been shown to be effective in a number of published studies (Rossion et al., 2004; Gauthier et al., 1998, 1999). Across 10 Greeble novices, using a single FFA measure as the denominator and $p = .05$ as the statistical threshold for the face minus object contrast (the middle bars for each graph in Figure 5), we found a significantly larger mean FFA–FFA overlap (40.8%) compared to the mean FFA–FGA area overlap (21.7%), $t(9) = 2.81$, $p = .02$. This result is

Table 2. Data from Selected Studies that Reported the Location of Representative FFA (in Talairach Coordinates), Average FFA Size (if Documented), and the FFA Localizer Tasks

| Study	FFA Localizer Task	Right FFA Talairach Coordinates	Right FFA Size (mm^3)
McCarthy, Puce, Gore, & Allison, 1997	View faces or flowers among objects	40, −59, −22	143
Kanwisher et al., 1997	Face vs. object passive viewing	40, −55, −10	1000
Gauthier et al., 1999	Face vs. object passive viewing	41, −55, −10	990
O'Craven & Kanwisher, 2000	Perceiving and imaging face vs. houses	37, −36, −18	2187
Joseph & Gathers, 2002	Picture detection task (among faces, natural and man-made objects, and fixations)	42, −45, −27	346
Druzgal & D'Esposito, 2003	Face vs. object passive viewing		820
Rhodes et al., 2004	Face vs. object passive viewing	44, −48, −20	1422
Grill-Spector et al., 2004	Face vs. object passive viewing	39, −40, −16	
Lehmann et al., 2004	Face vs. chair passive viewing	44, −48, −20	117
Grill-Spector et al., 2006	Face vs. object passive viewing	35, −52, −13	
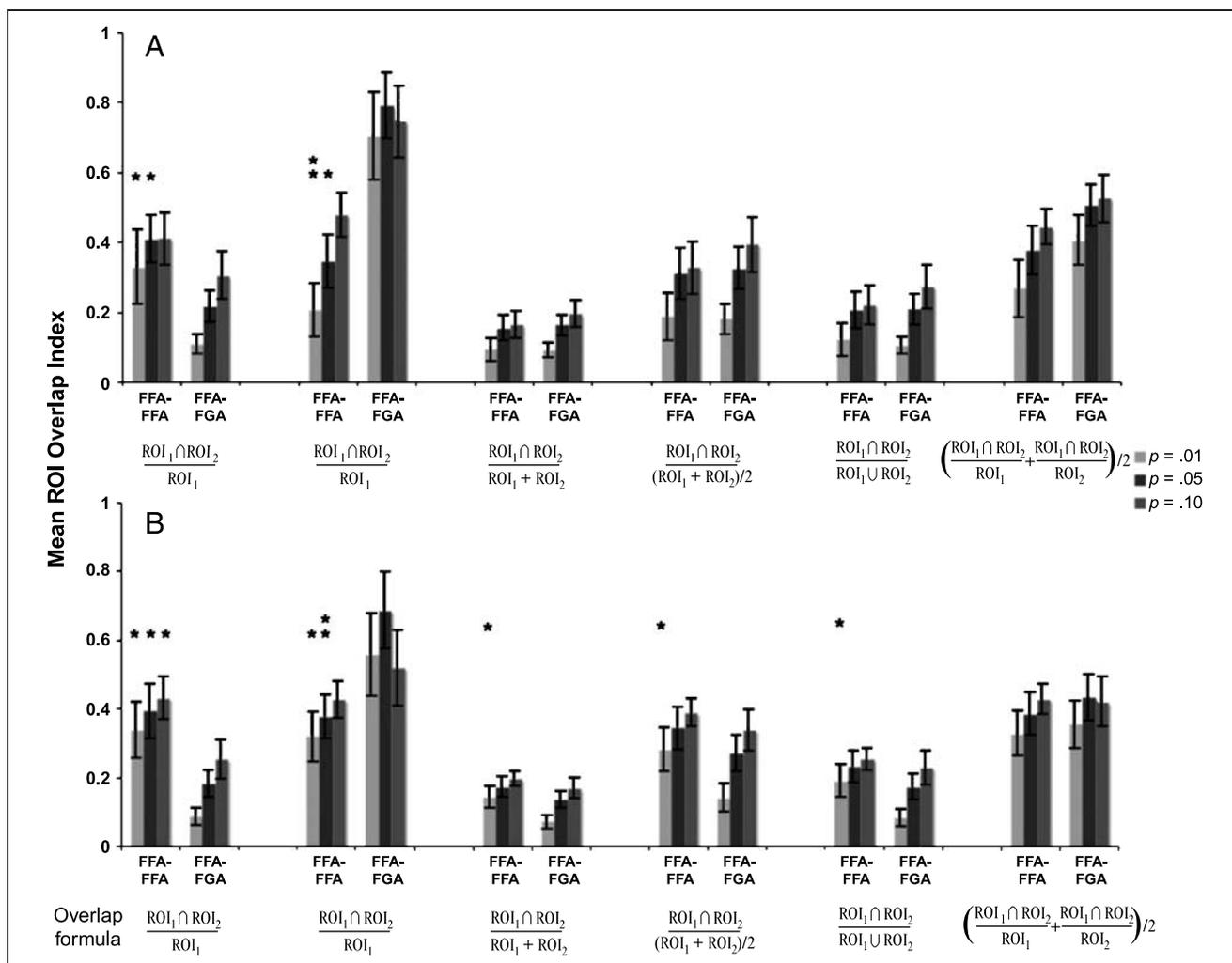
Figure 5. The six chosen ROI overlap indices (various denominators) calculated either across- or within-sessions (across runs) for FFA–FFA and FFA–FGA ("fusiform Greeble area") overlap analyses, with three levels of statistical thresholds ($p = .01$, $.05$, and $.10$, all corrected), using our fMRI data from 10 subjects. The upper row (A) represents the within-session, the lower row (B) the cross-session ROI overlap comparison. Across the two rows, both the first and second panels use a single ROI as the denominator of ROI overlap index and reveal significant FFA–FFA and FFA–FGA overlap differences (e.g., consistent with the face specificity prediction). Using denominators that take both ROI sizes into account (panels 3–6), we find no systematic differences in FFA–FGA versus FFA–FFA overlap. *$p < .05$, **$p < .001$.

consistent with Rhodes et al. (2004) and Grill-Spector et al. (2004). Note that Rhodes et al. used FFA activation as measured in an initial passive viewing session as the denominator, rather than using the FFA activation for faces measured during the experimental tasks (e.g., individuation for the trained novices, Experiment 1 or an average of passive viewing and individuation for the untrained novices in Experiment 1 and the experts in Experiment 2). In our case, because there was no systematic size difference for face activation across the three runs, we averaged the FFA–FFA overlap across all three comparisons (FFA$_{r1}$–FFA$_{r2}$, FFA$_{r2}$–FFA$_{r3}$, and FFA$_{r1}$–FFA$_{r3}$).

In order to compare the effect of different denominators on ROI overlap, we calculated six plausible variations of the overlap index by fixing the numerator as the number of overlapping voxels for ROI$_a$ and ROI$_b$, and then varying the denominator from formula a to f (see Methods). Figure 5A summarizes the mean results of the six different within-session FFA–FFA and FFA–FGA overlap comparisons. Using the $p = .05$ threshold level, we find significant FFA–FFA and FFA–FGA overlap differences [for a, $t(9) = 2.81$, $p < .05$; for b, $t(9) = -4.66$, $p < .001$], but only when overlap is calculated with a single ROI as the overlap denominator (the two leftmost comparisons in Figure 5). However, there were no significant overlap differences for the remaining overlap formulae [for c and d, $t(9) = -0.19$, $p = .84$; for e, $t(9) = -0.03$, $p = .97$; for f, $t(9) = -1.7$, $p = .12$], in which the denominator included both ROIs. Thus, at least using $p = .05$ for the threshold level, we replicated previous results, finding a significant overlap difference between FFA–FFA and FFA and another class of objects. However, it is important to note that this difference is obtained

only when we rely on a formula that uses only a single ROI as the denominator. In contrast, overlap formulae that take both ROIs into account reveal no significant differences between the overlap of two face localizers and the overlap of faces and Greebles.

In Figure 5A, the FFA–FFA overlap was calculated for three separate run comparisons, FFA_{r1}–FFA_{r2}, FFA_{r2}–FFA_{r3}, and FFA_{r1}–FFA_{r3}, and then averaged together. However, due to the limited number of Greeble-selective voxels in each run, the FFA–FGA overlap was averaged across three runs within a single session, FFA_{r123}–FGA_{r123}, and then compared. Because this was not a direct comparison (across-run in the FFA–FFA comparison vs. a combined-run comparison in FFA–FGA overlap), we also calculated the overlap comparison for the five Greeble experts over three scanning sessions so that across-session FFA–FFA and FFA–FGA overlap comparisons, calculated identically for both, were possible. These across-session results are shown in Figure 5B. Again for a $p = .05$ threshold, we observed a significant FFA–FFA and FFA–FGA overlap difference [for formula a, $t(14) = 2.58, p = .02$; for formula b, $t(14) = −2.64, p = .019$] in the first two columns. For the remaining four overlap comparisons, in which both ROIs were taken into account in the denominator, there were no significant differences between the overlap for face areas and the overlap for face and Greeble areas [for formulae c and d, $t(14) = 1.26, p = .22$; for e, $t(14) = 1.31, p = .20$; for f, $t(14) = −0.4, p = .96$].[2] A comparison of Figure 5A and B shows highly similar FFA–FFA and FFA–FGA overlap results for each overlap formula, regardless of how the sessions were combined. This demonstrates a consistency in the results across runs *and* sessions in our study. Together, these data strongly support the proposal that ROI overlap varies significantly across different calculation methods using different denominators.

It is worth pointing out that in Figure 5, the FFA–FFA versus FFA–FGA overlap comparisons for two formulae (a and b) only differed in the FFA–FGA comparison [21% in a vs. 79% in b, $t(9) = −0.68, p < .0001$ for the across-run analysis; 18% in a vs. 60% in b, $t(14) = −4.41, p = .0005$ for the across-session analysis], but did not differ substantially in the FFA–FFA overlap [41% in a vs. 35% in b, $t(9) = 1.81, p = .10$ for the across-run analysis; 40% in a vs. 38% in b, $t(14) = 0.23, p = .81$ for the across-session analysis]. This suggests a reverse pattern (e.g., "FFA–FFA > FFA–FGA" in a, vs. "FFA–FFA < FFA–FGA" in b) by merely changing the overlap denominator from the larger FFA in a to the smaller FGA in b. Another minor point of interest is, similar to the point raised in Figure 1, whether the denominator size alone can explain the current ROI overlap comparison pattern. The observation that the mean denominator sizes (for the FFA–FGA overlap in both across-run and across-session analyses) in formulae a (774 mm^3) and b (212 mm^3) were indeed smaller than in c (986 mm^3) and in e (828 mm^3) seems to support the importance of

denominator size per se in explaining the pattern of overlap comparison across formula, despite the case of formula d (493 mm^3), whose denominator size was in between those of formulae a and b, but with a similar overlap comparison pattern as with formulae c and e. More complicated, the results of formula b also signify the problem in changing the denominator across FFA–FFA and FFA–FGA comparisons (because the FFA was the denominator for calculating FFA–FFA overlap, and the FGA was the denominator for calculating FFA–FGA overlap), whose significant size differences surely contributed to the final significant overlap difference [mean FFA vs. FGA: 834 vs. 230 mm^3, $t(9) = 3.23, p = .01$ for the across-run analysis; 714 vs. 194 mm^3, $t(14) = 3.23, p = .005$ for the across-session analysis]. Therefore, inspections of formulae (a, b) versus (c, d, e, and f) in Figure 5 suggest that it is the combination of several factors, including at least the denominator size and the use of the identical denominator for the baseline and category-selective overlap calculations, that affects the final overlap comparison results.

Effects of Statistical Threshold on Overlap Analysis

Because the definition of ROI overlap is based on the ROI sizes and their area of overlap, all of which are determined by the given statistical threshold in the fMRI data analysis procedure (e.g., face–object contrast after the general linear model), overlap analysis is tremendously dependent on the statistical threshold. So far, we have only discussed the results for a threshold of $p = .05$, corrected, a common threshold used in many studies (Downing et al., 2006; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). To qualify Rhodes et al.'s (2004) claim that the conservative threshold can, in part, explain low ROI overlap calculations (<30%), and, more importantly, to address how statistical threshold affects the ROI overlap in general, we carried out two additional overlap analyses using $p = .01$ and $p = .10$, both corrected. These data are shown alongside the $p = .05$ bars in Figure 5. The general pattern is that for a threshold of $p = .01$, the overlap value slightly decreases, and for a threshold of $p = .10$, the overlap value increases, when compared to the original $p = .05$ analysis. Examined individually, a threshold increase (from $p = .05$ to $p = .01$) yielded three significant FFA–FFA and FFA–FGA overlap differences in the by-session analysis: formulae c and d, $t(14) = 2.28, p = .038$; and formula e, $t(14) = 2.29, p = .037$. In contrast, the threshold decrease (from $p = .05$ to $p = .10$) yielded nonsignificant or marginally significant results for formula a in the by-run analysis [$t(9) = 1.62, p = .13$], and formula b in the by-session analyses [$t(14) = −1.88, p = .07$]. In order to further examine the interaction between statistical threshold and overlap formula, we

ran a 3 ($p = .01$, $p = .05$, and $p = .10$) × 6 (the six formulae presented in the Methods) analysis of variance. With the mean ROI overlap difference as input, the interaction between statistical threshold and overlap formula was not significant [$F(10, 162) = 0.38$, MS $= .02$, $p = .95$ for the by-run analysis; $F(10, 72) = 0.23$, MS $= .01$, $p = .99$ for the by-session analysis], suggesting that quantitatively, the overall effect of statistical threshold change did not interact with the formulae chosen (refer to Figure 5). Taken together, these results suggest that the overall effect of statistical threshold on the mean ROI overlap is, indeed, contingent on the statistical threshold chosen. When examined individually, these threshold shifts can result in changes in whether an ROI overlap index passes or fails a significance test. Thus, the result of the *local* overlap comparison does depend on the specific threshold chosen. However, the *global* effect of threshold shift on ROI overlap is about the same across the different overlap formulae, with no significant interaction found between threshold and the different overlap formulae.

Lastly, our additional analyses only found partial support for Rhodes et al.'s (2004) claim that the more conservative threshold decreases the overlap value (and a more lenient threshold increases the overlap value). In our study, such increases or decreases were small or moderate, and clearly not the only contributing factor to the overall low overlap values. It is more likely that the low FFA–FFA overlap reported in their study (27%) was primarily due to the large FFA chosen as the denominator, not the conservative statistical criteria they adopted.

Monte Carlo Simulations

A comprehensive evaluation of the overlap analysis is not complete without an objective assessment of the expected overlap range under each overlap formula. Computing the mean overlap index using various ROI sizes provides the chance to compare the results from the fMRI data analysis with those from the simulation. Also, the simulation serves as an important selection criterion for determining the ''best'' overlap formula for use in the type of study presented here. We define ''best'' as the formula that is least biased for the comparison of two ROIs, regardless of whether the ROI sizes are quite similar (e.g., FFA–FFA) or largely different (FFA–FXA, X being any given object category of expertise).[3] Here, we used a Monte Carlo simulation to determine the range of mean expected values under systematic ROI size variations.[4]

Figure 6 shows the simulation results of 10,000 iterations for each of the six formulae used in our ROI overlap comparisons. The mean ROI overlap is represented along the vertical axis, and the horizontal axis indicates the size difference between the two virtual ROIs. To simulate the potential size relationships between category-selective brain regions, we used 50-unit increments for ROI_1 (ranging from 50 to 1000 units) and held ROI_2 constant at 1000 units. This produced 19 steps, with the ROI *differences* becoming systematically smaller (e.g., 950 to 0) as the simulations progress toward the right in Figure 6. Three observations are apparent. First, the overlap index varied by formula when the ROI sizes were equal (ROI_1 and $ROI_2 = 1000$). For formulae a, b, d, and f (Figure 6A, B, D, and F) it was 50%, for formula c it was 25% (Figure 6C), and for formula e it was 33% (Figure 6E). Second, the effect of ROI size differences on the estimated overlap index differed substantially across formula, as represented by the slopes of the bar graphs. For example, when the denominator was a single ROI, the overlap index varied substantially with ROI size difference (Figure 6A and B). Formulae c, d, and e yielded intermediate slopes (Figure 6C, D, and E), which showed a plateau at an ROI_1 size of 500 units (ROI size difference of 500). In contrast, formula f was robust to changes in the size difference, yielding the same mean overlap (50%) across all ROI size differences (Figure 6F). Not surprisingly, our empirically established ROI overlap results fall within the range of values predicted by the Monte Carlo simulations. For example, in the third panel of Figure 5, the mean overlap values hovered around 0.20, close to the Monte Carlo simulation result in Figure 6C. Similar ranges of mean ROI overlap values can also be found in the fourth panel of Figure 5 (0.2–0.4) versus Figure 6D (0.2–0.5), the fifth panel of Figure 5 (0.2–0.3) versus Figure 6E (0.2–0.3), and the sixth panel of Figure 5 (0.3–0.4) versus Figure 6F (~0.5). Figure 6A and B, in contrast, clearly show a more drastic change in the overlap index as the size difference varies when a single ROI is used as the denominator, reflected by a larger slope of change (compared with c, d, e, and f). This variability makes it more difficult to compare the simulations with real fMRI data.

DISCUSSION

The present study examined whether the functional properties of category-selective brain regions can be inferred on the basis of the degree to which ROIs overlap. Under this proposal, we expect a high overlap index for tasks or object classes that are thought to recruit common functional systems, for example, the FFA in perceiving and imagining faces (O'Craven & Kanwisher, 2000), or in passively viewing two sets of faces (Rhodes et al., 2004). At the same time, we expect a low degree of overlap for categories or tasks that we believe are processed by distinct functional systems. Obviously, this argument relies on the assumption that spatially localized brain regions correspond, at least roughly, to functional mechanisms. Although this logic is widely accepted in the neuroimaging literature, spatial localization (or

Figure 6. Six simulation results (over 10,000 iterations each) of changing the size of ROI_1 in 50-unit steps (from 50 units to 1000 units) shown along the horizontal axis (ROI_2 was kept constant at 1000 units); the vertical axis plots the mean overlap index. Given that each different formula has different asymptote overlap value (ranging from 25% to 50%) and a different degree of stability over ROI size changes, the most consistent overlap formula is the average of A ∩ B/A and A ∩ B/B (F), which also seems best suited for comparing FFA–FFA to FFA–FGA (typically having different-sized ROIs) in the current study.
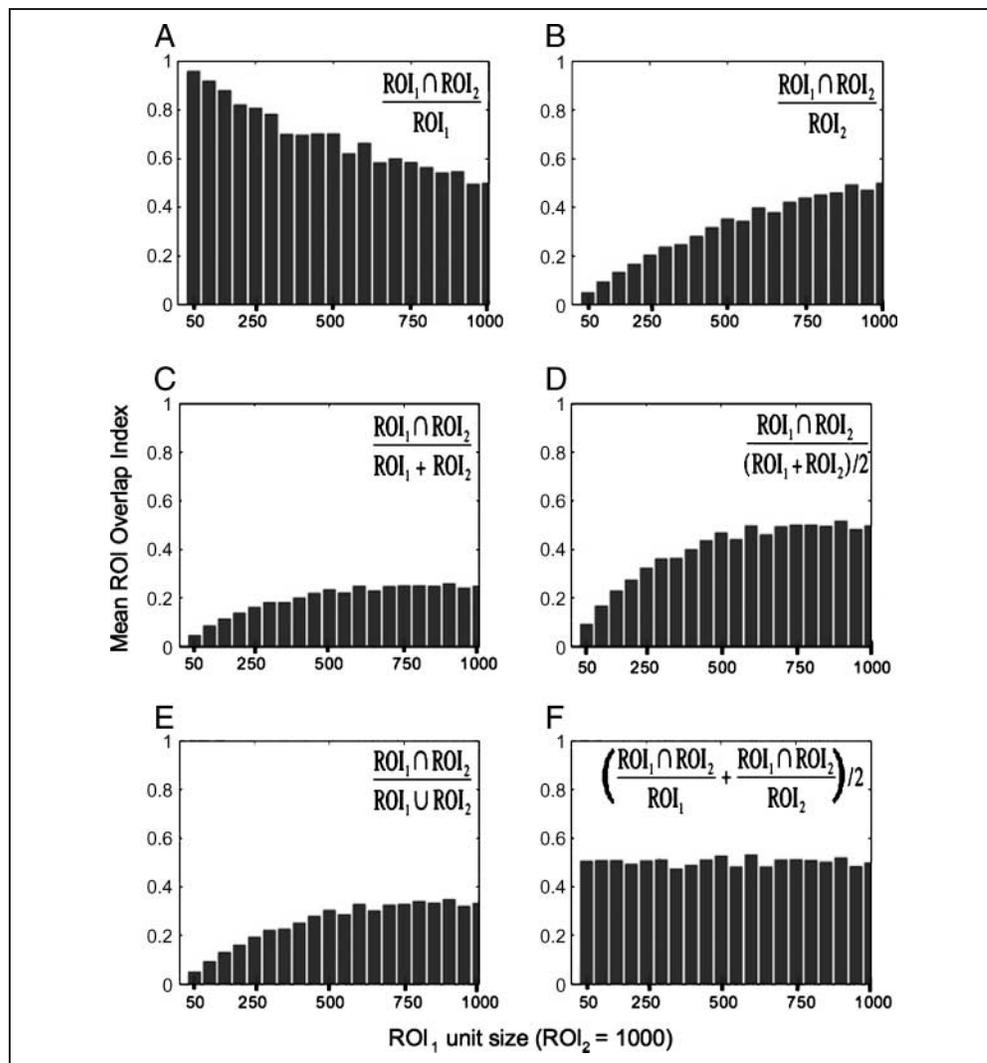
co-localization) may not entirely address the issue of functional commonalities. For example, as will be discussed in more detail later, higher-resolution imaging might reveal functional separability at a finer level of analysis (e.g., Grill-Spector, Sayres, & Ress, 2006). As Astafiev et al. (2004) suggest, a low overlap index may be the result of using a large, single area as the denominator—a proposal tested systematically by our current study.

Consistent with Astafiev et al.'s (2004) claims, the results of our FFA–FFA and FFA–FGA ("fusiform Greeble area") overlap comparisons reveal a clear influence of the choice of denominator: When a single FFA is used, the FFA–FFA overlap is significantly larger than the FFA–FGA overlap (replicating Rhodes et al., 2004). Importantly, this overlap difference disappeared altogether when the overlap denominator was a combination of both ROIs, be it the sum, the union, or the average. Monte Carlo simulations corroborated this empirical result, and further revealed that, compared to a single ROI denominator, when a combination of the two ROIs was used, the overlap index was much more resistant to

ROI size variations, and the average of two single ROIs as the denominator was the best among the six formulae tested in the current context (comparing FFA–FFA and FFA–FXA overlap), in which within-subject ROI size variation was a central issue.[5]

In addition to finding significant changes in the overlap index depending on the formula chosen, our results also address the issue of consistency in the FFA across tasks. Although both 1–back identity and passive viewing tasks are commonly used as default FFA localizer tasks, and some studies have found no significant difference between the two tasks (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000), the nature of passive viewing may still vary within the task context. In the Rhodes et al. (2004) study, participants were told to "passively view" the stimuli in both the initial FFA localizer and the ensuing experimental scans with faces. However, the mean FFA size ratio for the localizer (FFA_{lo}) was almost half of that in the experimental face scans (FFA_{pv}) (813 mm^3 vs. 1440 mm^3). Rhodes et al. suggested that this difference was the result of using

different stimuli and epoch lengths. In contrast, several studies have found that the FFA is consistent across sessions and/or experimental block length (Gauthier et al., 1999; Kanwisher et al., 1997). In our study, we compared face activation in the same task (1−back identity) across three runs within a session, or three separate scanning sessions. Consistent with these early reports, we found little size variation (mean FFA size: 775 mm^3 in FFA$_1$, 780 mm^3 in FFA$_2$, and 623 mm^3 in FFA$_3$) and a high correlation across three scan sessions [$r(14) = .91$, $p <$ .0001], indicating that FFA activation was relatively stable, at least across sessions, in terms of small center-of-FFA distance variation (several millimeters). In addition, we found a proportionate decrease in terms of ROI overlap indices as the center-of-FFA distance increased, both across runs and across sessions (Figure 4). Thus, the issue of FFA ''consistency'' depends on how consistency is defined, either in terms of distance between center-of-mass Talairach coordinates (relatively small) or in terms of ROI overlap indices (larger decrements).

Our results are congruent with the majority of the FFA literature demonstrating the reliable consistency of repeated FFA measures. In addition, we are able to explain what has been an incongruence between overlap measures and center-distance measures. Despite the apparent disparity between these measures, we found a systematic relationship between FFA-center distance and corresponding FFA–FFA overlap. Furthermore, the correspondence between ∼40% FFA–FFA overlap calculated from the fMRI data, and the 50% overlap calculated from the Monte Carlo simulation (both from formula f, which is least resistant to ROI size variations), suggests that even in a highly consistent region like the FFA, the expected overlap across runs or sessions is only about 50%. In support of this finding, if we recalculate the overlap for several studies using formula f, we find similar results. The average FFA–FFA overlap for O'Craven and Kanwisher (2000) is 53%, and for Rhodes et al. (2004) is 38%. These calculations are similar to our findings using a 1−back identity task, both across-run (38%) and across-session (39%). The only higher-than-expected overlap index for the FFA–FFA was found for Grill-Spector et al. (2004). Using formula f to recalculate their data, we find a mean overlap of 72%. Among other possibilities, this high overlap may be due to differences in task, or to the small sample size ($n = 5$) in their study.

It is also important to note that, although our data were generally consistent with the predictions of the ''perceptual expertise hypothesis'' (Bukach et al., 2006), our participants only exhibited behavioral evidence for Greeble expertise by the third scan (Gauthier et al., 1999; Gauthier & Tarr, 1997). Thus, it is somewhat surprising that we found no significant difference in the overlap index for Greebles and faces from the initial scans onward. This apparent dissociation may be explained, in part, by previous studies that have shown that fusiform activity is related to task demands (Rogers,

Hocking, Mechelli, Patterson, & Price, 2003) and high within-family homogeneity of the stimulus class (Gauthier, Tarr, et al., 2000; Gauthier et al., 1997). In particular, we may have failed to find a change in the degree of FFA–FGA overlap with training because we employed a 1−back identity task—a task that has been associated with increased focal fusiform activity relative to passive viewing (Kanwisher et al., 1998). Such task-related activation might also help account for previous results which show a nonlinear relationship between behavioral measures of expertise and FFA activity for objects of expertise (e.g., car experts viewing cars) when using a 1−back identity task (Gauthier, Skudlarski, et al., 2000). This heightened activation for an object category in the 1−back task may explain why we found similar overlap for the FFA–FGA and FFA–FFA in participants' first scans, prior to expertise training. Regardless of one's preferred explanation, the significant FFA–FFA and FFA–FGA overlap we observed before expertise training is certainly not easily explained by current instantiations of the face specificity hypothesis.

Last year, two fMRI studies explored the effect of ''expertise training'' of novel objects in the whole brain (Moore, Cohen, & Ranganath, 2006) or in the extrastriate cortex (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). After training, participants in Moore et al.'s study showed increased activation in the dorsolateral prefrontal, inferior parietal, and occipito-temporal cortices. In Op de Beeck et al.'s study, they found increased activation in the lateral occipital complex (LOC). Both studies failed to find a significant increase in FFA activity after training. A closer look at the training paradigms for these studies, however, reveals disparate tasks used in training and scanning. In previous Greeble training studies (Rossion et al., 2004; Gauthier et al., 1998), object naming tasks (e.g., what is this Greeble's name?) and shape–name verification (Is this Greeble's name ''Tezi''?) were alternated. This alternation of different tasks was intended to increase the speed of subordinate-level access, a characteristic of expert-level recognition (Tanaka, 2001). During scans, a simple delayed-matching or passive-viewing task was typically used (Gauthier et al., 1999). In contrast, in the Moore study, simultaneous match-to-sample, delayed recognition, family placement, and family discrimination tasks were used during training, and a typical working memory task (cue–delay–probe) was used during scan. In the Op de Beeck study, discrimination learning was used during training, and a demanding color-change detection task was used during the scan. Therefore, the task differences and, more importantly, the disparate underlying representations the participants may have acquired in these studies, may explain their failure to find significantly increased FFA activity after training. Such conjecture, however, calls for future studies that systematically compare various training paradigms on (a) specific ''aspects of expertise'' acquired during training, (b) the different neural manifestations underlying training process, and (c) the different patterns of neuronal reorganization.

Finally, as mentioned earlier, a recent discussion of high-resolution fMRI studies of the FFA (Baker, Hutchison, & Kanwisher, 2007; Grill-Spector, Sayres, & Ress, 2007; Simmons, Bellgowan, & Martin, 2007; Grill-Spector et al., 2006) further reinforces our arguments regarding the importance of formula definition for the particular results and ensuing conclusions. Grill-Spector et al. (2006) scanned five novices and used [Preferred − Nonpreferred] / [Preferred + |Nonpreferred|] as the category selectivity index. They found evidence, at 1-mm³ voxel resolution, for category-selective voxels to faces, animals, cars, and even sculptures. However, as Baker et al. (2007) and Simmons et al. (2007) pointed out, the analysis method used by Grill-Spector et al. was flawed due to the voxel selection procedure (dependent vs. independent) and the lack of a baseline adjustment. Especially relevant to our study is the commentary of Simmons et al., in which they argue that using the absolute value of the nonpreferred condition in the selectivity index can inflate the voxel selectivity for other categories within the FFA. Sidestepping the issue of whether the FFA is homogeneously face selective or intermingled with other non-face-selective voxels (Grill-Spector et al., 2007; Peelen & Downing, 2005b; Schwarzlose, Baker, & Kanwisher, 2005), our study highlights the importance of choosing a sound index and verifiable rationale, and then cautiously interpreting the results.

To conclude, an overlap index, when appropriately used, can provide useful information and complement other analyses. However, it is not a suitable measure to assess category specificity on its own, as demonstrated in our study. We found that changing the denominator using six perfectly plausible alternatives significantly changed the results of the overlap comparison. These data suggest that finding a significant difference between FFA–FFA overlap and FFA–FXA (X being any given category of domain experts) overlap may be largely a result of choosing a denominator that biases the results in this direction, rather than reflecting a robust difference between face areas and areas of expertise. Furthermore, using ROI overlap comparisons to test the face specificity and perceptual expertise hypotheses is not ideal, in particular, because we found equivalent FFA–FFA and FFA–FXA overlap even in novices. In addition, the mean overlap value for different face measures, using the least-biased overlap formula, was only about 50%. Therefore, it is our contention that different models of FFA function, including face specificity (McKone & Kanwisher, 2005; Grill-Spector et al., 2004; Spiridon & Kanwisher, 2002; Kanwisher, 2000), distributed (O'Toole, Jiang, Abdi, & Haxby, 2005; Hanson, Matsuka, & Haxby, 2004; Haxby et al., 2001) and perceptual expertise accounts (Gauthier & Bukach, 2007; Bukach et al., 2006; Tarr & Gauthier, 2000), may be better served by focusing more on why certain stimuli, tasks, analysis methods are chosen, and how these particular concatenations do or do not give rise to category selectivity.

Reprint requests should be sent to Chun-Chia Kung, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544, or via e-mail: chunkung@princeton.edu, or to Jessie J. Peissig, Department of Psychology, California State University, Fullerton, P.O. Box 34080, Fullerton, CA 92834-9480, or via e-mail: jpeissig@fullerton.edu.

Notes

1. Instead of using a functionally defined FFA as the unit of the face-selective cluster, Rhodes et al. (2004) actually used an anatomical approach, which summed the total activated voxels in the FG as the unit of their overlap analysis. The anatomical approach in defining FFA is seldom used in the literature, and may miss some FFA voxels that extend into adjacent gyri or sulci. Despite this, in order to better compare our results with those of Rhodes et al., we drew the face-selective voxels in FG anatomically for one participant and found that, although the number of individual voxels included in the FFA increased, the overall pattern of the overlap comparisons was relatively unchanged.

2. Because of the small sample size ($n = 5$ experts) in this cross-session overlap comparison, the expected differences between mean FFA–FFA and FFA–FGA overlap with a single ROI as denominator were not significant [(FFAs$_1$ ∩ FFAs$_2$)/FFAs$_1$ = 0.595, (FFAs$_1$ ∩ FGAs$_1$)/FFAs$_1$ = 0.212, (FFAs$_2$ ∩ FFAs$_3$)/FFAs$_2$ = 0.204, (FFAs$_2$ ∩ FGAs$_2$)/FFAs$_2$ = 0.111; (FFAs$_3$ ∩ FFAs$_1$)/FFAs$_3$ = 0.376, (FFAs$_3$ ∩ FGAs$_3$)/FFAs$_3$ = 0.216, all $p > .05$]. It was significant only after combining these three across-session comparisons [$t(14) = 2.56$, $p < .02$], suggesting an advantage for combining the three across-session comparisons. The same pattern holds for the remaining five overlap formulae.

3. We have to emphasize that the Monte Carlo simulation, under the current context, was used to test which overlap index was best when comparing the *relative stability across various ROI size variations*. The point here is whether size variation is an issue; when it is not, many studies have used the sum of the two ROIs as the denominator (e.g., the high-resolution FFA study by Grill-Spector et al., 2006). They defined the selectivity index as (Preferred − Nonpreferred)/ (Preferred + |Nonpreferred|). Although there may not be a single overlap index suitable for all situations, our simulation reveals the robustness of each formula in the present context.

4. Matlab code for Monte Carlo simulations is provided below:

```
for j = 1:20
  a(j) = 1000; % cluster 1
  b(j) = 50*j; % cluster 2

  tot(j) = a(j) + b(j);

  for i = 1:10,000 % number of iterations
    v = zeros(1,tot(j));
    r = randperm(tot(j));
    v(r(1:a(j))) = v(r(1:a(j))) + 1;
    r = randperm(tot(j));
    v(r(1:b(j))) = v(r(1:b(j))) + 1;
    ovp(i) = length(find(v==2));
  end
```

```
%computing mean ROI overlaps in 6 formula
mean_ovp1(j) = mean(ovp(i))/a(j);
mean_ovp2(j) = mean(ovp(i))/b(j);
mean_ovp3(j) = mean(ovp(i))/tot(j);
mean_ovp4(j) = mean(ovp(i))/(tot(j)/2);
mean_ovp5(j) = mean(ovp(i))/(tot(j) − mean(ovp(j)));
mean_ovp6(j) = (mean(ovp(i))/a(j)+mean(ovp(i))/b(j))/2;
end
```

5. With the recommended formula of $d' = \frac{[M_{\text{preferred}} - M_{\text{nonpreferred}}]}{\sqrt{\frac{\sigma^2_{\text{preferred}} + \sigma^2_{\text{nonpreferred}}}{2}}}$

(Afraz, Kiani, & Esteky, 2006), we also adopted and recalculated the overlap index with the formula $\frac{[\text{ROI}_a \cap \text{ROI}_b]}{\sqrt{\frac{\text{ROI}_a^2 + \text{ROI}_b^2}{2}}}$ (the standard

deviation in the denominator was not available because each ROI had only the size information here, unlike the case of d' with four responses: faces, objects, sculptures, and animals, for each voxel). The results of the FFA–FFA versus FFA–FGA overlap is 0.291 versus 0.363, $t(9) = 0.64, p = .53$ for across-run, and 0.332 versus 0.397, $t(14) = −0.696, p = .497$ for across-session analysis, consistent with the results from formulae c, d, e, and f.

REFERENCES

Afraz, S. R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature, 442,* 692–695.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience, 7,* 542–548.

Baker, C. I., Hutchison, T. L., & Kanwisher, N. (2007). Does the fusiform face area contain subregions highly selective for nonfaces? *Nature Neuroscience, 10,* 3–4.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *Journal of Neuroscience, 16,* 4207–4221.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences, 10,* 159–166.

Caplan, D., & Moo, L. (2004). Cognitive conjunction and cognitive functions. *Neuroimage, 21,* 751–756.

Chee, M. W., Weekes, B., Lee, K. M., Soon, C. S., Schreiber, A., Hoon, J. J., et al. (2000). Overlap and dissociation of semantic processing of Chinese characters, English words, and pictures: Evidence from fMRI. *Neuroimage, 12,* 392–403.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16,* 1453–1461.

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

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience, 15,* 771–784.

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage, 10,* 385–396.

Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *Neuroimage, 25,* 661–667.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Brain Research, Cognitive Brain Research, 20,* 226–241.

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology, 7,* 645–651.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103,* 322–330.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3,* 191–197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a ''greeble'' expert: Exploring the face recognition mechanism. *Vision Research, 37,* 1673–1682.

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,* 568–573.

Gauthier, I., Tarr, M. J., Moylan, J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2000). Does visual subordinate-level categorisation engage the functionally defined fusiform face area? *Cognitive Neuropsychology, 17,* 143–163.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training ''greeble'' experts: A framework for studying expert object recognition processes. *Vision Research, 38,* 2401–2428.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7,* 555–562.

Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience, 9,* 1177–1185.

Grill-Spector, K., Sayres, R., & Ress, D. (2007). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience, 10,* 133.

Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a ''face'' area? *Neuroimage, 23,* 156–166.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303,* 1634–1640.

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

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3,* 284–291.

Joseph, J. E., & Gathers, A. D. (2002). Natural and manufactured objects activate the fusiform face area. *NeuroReport, 13,* 935–938.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3,* 759–763.

Kanwisher, N. (2006). Neuroscience. What's in a face? *Science, 311,* 617–618.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68,* B1–B11.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews: Neuroscience, 2,* 635–642.

Kourtzi, Z., Bulthoff, H. H., Erb, M., & Grodd, W. (2002). Object-selective responses in the human motion area mt/mst. *Nature Neuroscience, 5,* 17–18.

LaBar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional MRI comparison within subjects. *Neuroimage, 10,* 695–704.

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,* 75–83.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9,* 605–610.

McKone, E., & Kanwisher, N. (2005). Does the human brain process objects of expertise like faces? A review of the evidence. In S. Dehaene, J. Duhamel, M. Hauser, & G. Rizzolatti (Eds.), *From monkey brain to human brain.* Boston: MIT Press.

McKone, E., & Robbins, R. (2006). The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition.*

Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *Neuroimage, 8,* 129–139.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience, 26,* 11187–11196.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25,* 653–660.

Nieto-Castanon, A., Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2003). Region of interest based analysis of functional imaging data. *Neuroimage, 19,* 1303–1316.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience, 12,* 1013–1023.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience, 17,* 580–590.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 26,* 13025–13036.

Peelen, M. V., & Downing, P. E. (2005a). Is the extrastriate body area involved in motor actions? *Nature Neuroscience, 8,* 125–126.

Peelen, M. V., & Downing, P. E. (2005b). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93,* 603–608.

Peelen, M. V., & Downing, P. E. (2005c). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping, 25,* 402–408.

Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *Neuroimage, 5,* 261–270.

Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience, 16,* 189–203.

Rogers, T. T., Hocking, J., Mechelli, A., Patterson, K., & Price, C. J. (2003). Fusiform activation to animals is driven by the process, not the stimulus. *Journal of Cognitive Neuroscience, 17,* 434–445.

Rossion, B., Kung, C. C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 101,* 14521–14526.

Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., & Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *Journal of Neuroscience, 25,* 4294–4306.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25,* 11055–11059.

Simmons, W. K., Bellgowan, P. S., & Martin, A. (2007). Measuring selectivity in fMRI data. *Nature Neuroscience, 10,* 4–5.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35,* 1157–1165.

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130,* 534–543.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3,* 764–769.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17,* 257–279.