ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 20 (2003) 1561–1577

www.elsevier.com/locate/ynimg

Reliability of functional localization using fMRI

Khena M. Swallow, Todd S. Braver, Abraham Z. Snyder, Nicole K. Speer, and Jeffrey M. Zacks*

Washington University, St. Louis, MO 63130-1125, USA

Received 26 February 2003; revised 10 July 2003; accepted 14 July 2003

Abstract

Neuroimaging researchers increasingly take advantage of the known functional properties of brain regions to localize them and probe changes in their activity under different conditions. The utility of this approach depends in part on the reliability of the methods used to define these regions of interest. Two operations may affect the reliability of functionally identified regions: spatially normalizing data to a stereotactic atlas and statistically combining data across participants to form a composite region (as opposed to identifying individual regions for each participant). The effect of these two operations on reliability was evaluated for two functionally identifiable regions: the MT complex and the frontal eye fields. Spatial normalization had almost no effect on within-subject reliability, while grouping across participants negatively affected retest measures of the activation and location of regions defined on separate occasions. We conclude that, for typical sample sizes and numbers of observations per subject, functional localization is most reliable when performed for each individual using data in atlas space.

Functional magnetic resonance imaging (fMRI) offers researchers the ability to associate cortical regions with discrete cognitive processes. Data from neuroimaging, clinical case studies, and nonhuman primate research have identified a number of brain regions with well-known functional properties. This mapping allows one to ask how a functional region responds to a range of situations and tasks. A region of interest (ROI) is first identified by isolating a cortical area based on its activity during a *localizer* task (such as the perception of moving stimuli or assessment of the familiarity of a human face). The behavior of this region is then examined during the performance of another experimental task. If the magnitude or time course of activity in this region is significantly modulated by the performance of the experimental task, its function may be a component process of the cognitive task under investigation. Several researchers routinely use functional localization to identify ROIs in the posterior lateral occipital–temporal cortex (e.g., Beauchamp et al., 2002), the fusiform face area (e.g., Tong et al., 2000), the lateral occipital complex (e.g., Kourtzi and Kanwisher, 2001), and the parietal place area (e.g., Downing et al., 2001) in individual participants before submitting the time courses of the ROIs to statistical analysis.

Alternatively, anatomical localization may be used when a functional region is consistently located relative to structural landmarks (Brett et al., 2002; Petersson et al., 1999). Such circumstances may arise when single cell recording studies in nonhuman primates and clinical neuropsychological studies in humans demonstrate that a structurally defined cortical region is involved in a particular cognitive function. Structural landmarks are typically identified by skilled neuroscientists and can be fairly labor and resource intensive. The demanding nature of this method may therefore make it impractical for many research projects. Furthermore, functional regions may not reliably respect sulcal landmarks and cytoarchitectonic structure may be difficult or impossible to locate in every individual in the sample (Lancaster et al., 2000).

However, functional localization has several problems of its own. A critical assumption about the consistency of the localizer task must be made to make the types of inferences described above. The extent, location, and magnitude of the

* Corresponding author. Department of Psychology, Washington University, Box 1125, St. Louis, MO 63130. Fax: +1-319-935-7588.
E-mail address: jzacks@artsci.wustl.edu (J.M. Zacks).

1053-8119/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1053-8119(03)00436-1

response to the localizer task must be reliable; that is, these properties of BOLD activity must be consistent from one run or scanning session of the localizer task to another. Fundamentally, the reliability of any measure, instrument, or methodology will directly affect the power of an experimental test through the introduction of unexplained error and uncontrolled variability (Cohen and Cohen, 1993). Within the context of functional localization, a procedure for identifying an ROI is reliable if multiple repetitions of the scanning protocol and statistical analysis on a given individual produce stable estimates of the location, extent, and responsiveness of the ROI.

During statistical analysis, two operations may substantially affect the reliability of an ROI. The first is *spatial normalization,* which registers the raw data to a stereotactic atlas. Though a number of researchers have the intuition that functional localization is best performed without spatial normalization, at least one direct, quantitative comparison of registered and nonregistered data has suggested that spatial normalization has no effect on the reproducibility of BOLD data (Miki et al., 2000; for other comparisons see Nobre et al., 1997; Downing et al., 2001; Dumoulin et al., 2000; Grosbras et al., 1999). The second operation that may affect reliability is statistically combining data across observers to form a composite region for the group as a whole. To our knowledge, the effect of *statistical grouping* on reliability has not previously been investigated.

Previous research indicates that BOLD activation varies substantially both across individuals and within individuals across scanning sessions. This variability depends in part on spatial filter size (Rombouts et al., 1998), flip angle (Moser et al., 1996), cortical region (Machielson et al., 2000), and experimental task (McGonigle et al., 2000; Specht et al., 2003). These earlier studies show that error in the localization of a region may be both common and large, but can be ameliorated by careful design and analysis choices.

Changes in ROIs identified through different localization runs reflect the combination of a true response to the localizer task and error introduced from random, methodological, interindividual, and/or intraindividual sources (Hunton et al., 1996). In addition to environmental and scanner changes across sessions (McGonigle et al., 2000), interindividual sources of error arise from the variability of structural and functional anatomy across individuals. Brains vary structurally in the relative placement, configuration, shape, and size of the sulci and gyri. Functional variability across individuals may be reduced by structural registration procedures, but may not be eliminated. To further complicate matters, some functional regions do not exist or cannot be identified in all individuals (Lancaster et al., 2000; Hunton et al., 1996). Intraindividual sources of error arise from changes in an individual's pattern and magnitude of cortical activation from one occasion of performing a task to another. Inconsistent task performance, uncorrected head motion, scanner noise, and misregistration of individual functional images to their structural images contribute to this source of error (Hunton et al., 1996; McGonigle et al., 2000) and should serve to decrease the reliability of functional localization.

These sources of localization error interact differently with spatial normalization and statistical grouping. Advantages of spatial normalization include the ability to report results in terms of standardized reference frame and the quantitative comparison of individual data at the voxel level. In their recent review of functional localization, Brett and colleagues (2002) asserted that the method used to spatially normalize functional images is crucial to accurate localization and that the appropriate method varies across regions, experimental objectives, and the availability of computational resources. Here we refer to spatial normalization as the process of aligning and warping functional images of individual brains to a standard stereotactic atlas. One method of spatial normalization is accomplished through volume matching procedures that align gross anatomical features and warp individual brains to the standard atlas volume. Small differences in structural anatomy across individuals, particularly those involving the absence of sulci, cannot be corrected through this process. Such anatomical variability is often demonstrated in investigations of anatomical landmarks for functional areas, such as that conducted by Grosbras et al. (1999) on the supplementary eye fields (SEF). These authors used a self-paced saccade task to identify the SEF in each individual relative to their cortical anatomy and relative to standard stereotactic coordinates. Their results suggest that the location of SEF relative to stereotactic coordinates is more variable across individuals than the location of SEF relative to an anatomical landmark identified for each individual.

The effects of spatial normalization on reliability depend upon the size of the resampled voxel and the number of resampling steps involved in registration. With a single resampling step that corrects for head motion and registers data to an atlas, smaller voxels should preserve more information by increasing resolution. When compared to native space data that have been corrected for head motion this may help or have neutral effects on reliability. However, if the data are resampled more than once (to first correct for head motion and then again for atlas registration), there will be a cumulative loss of high spatial frequency information, a by-product of resampling. If more than one resampling step is used, spatial normalization may result in the loss of high-frequency components of the signal and decrease the reliability and power of methods utilizing it. Spatial normalization, then, may help or have neutral effects on functional localization when it is performed in one step but may introduce error into functional localization by removing crucial components of the task-related BOLD signal when performed in multiple steps.

Following spatial normalization, individual functional data may be combined to form a single image through statistical grouping. The primary advantage of statistical grouping is an increase in power. An effect that may be too

weak to identify at the individual level may be detectable when data are pooled. This is accomplished through the use of statistical procedures, most commonly through *t* tests of specific contrasts or correlation analyses, to combine individual data in a single, group image. Because most research is primarily concerned with generalization to the population, these analyses increasingly employ random effects rather than fixed effects procedures. (For more on this topic see Friston et al., 1999.) As a result, group images represent the mean signal change of every individual in the sample relative to between participant variability. A second important advantage of statistical grouping is practical: a single ROI is easier to summarize and communicate to others.

Anatomic variability between individuals argues against functional localization of ROIs using group images. ROIs defined for each individual can accommodate the variations in functional and sulcal anatomy described by Brett et al. (2002) and others (Petersson et al., 1999; Hunton et al., 1996; Culham et al., 2001; Grosbras et al., 1999; Dumoulin et al., 2000). These variations are averaged out of group images (they are essentially treated as random error) and are not accounted for when the group ROI is later applied to individual data (Petersson et al., 1999). Individual ROIs provide more precision than group ROIs as well as the opportunity to measure variability in region location (Nobre et al., 1997). Every individual in a sample may show a strong, focal response to the localizer task. However, as Brett et al. (2002) point out, these responses may not align in atlas space even if their location relative to a sulcus is consistent (Grosbras et al., 1999; Downing et al., 2001). This could result in the circumstance where the group ROI cannot be identified (between subject variability is too great) or is so extensive that the effect of a task in each individual is washed out by voxels whose activity is not related to the cognitive process under study (Brett et al., 2002). Thus, statistical grouping can lead to including voxels that for some individuals should be excluded and excluding voxels that for some individuals should be included.

To summarize, both spatial normalization and statistical grouping offer potential advantages and pitfalls. It is therefore not clear *a priori* whether either procedure is desirable in typical functional neuroimaging studies using localizer tasks. To evaluate the effect of each of these operations, we examined the reliability of functional ROIs identified individually using normalized and nonnormalized data and the reliability of functional ROIs identified for the group as a whole. Two functional ROIs were created using data from separate localizer runs from the same scanning session. This presented the opportunity to examine the test–retest reliability of these methods using the same tasks and participants while avoiding any bias that may result from using the same localizer runs to both define and evaluate the ROIs.

In this study, we examined two functionally defined ROIs: the frontal eye fields (FEF) and the MT complex (MT+). A number of studies have characterized the FEF as a region between the central and precentral sulcus and anterior to the hand representation in the motor cortex (Paus, 1995) that is active during both overt and covert shifts of visual attention (Corbetta and Shulman, 2002; Petit and Haxby, 1999; Rosano et al., 2002). While the subject is performing saccades to different locations in the visual field, this region shows increased single unit activity in macaques and increased BOLD activity in humans (Paus, 1995). Because of their well-documented behavior during the performance of a wide variety of oculomotor tasks, the FEF are relatively simple to functionally localize and are likely to be reliably activated across several localization runs and across individuals.

The human MT complex is an extrastriate visual region that is selectively responsive to motion in visual stimuli. As with FEF, the behavior of this region is well documented in both macaques and humans (Dukelow et al., 2001; Morrone et al., 2000; Tootell et al., 1996) and is predictable in its response and sensitivity to a variety of motion stimuli (Culham et al., 2001). In a comprehensive study of the response properties of human MT+, Tootell and colleagues (1995) confirmed that in humans this region shows less contrast sensitivity than surrounding visual areas, decreased activity for stimuli of equiluminance, and variability in its location across participants. In a more recent analysis, Dumoulin et al. (2000) found that the location of MT+ in stereotactic coordinates is variable across participants, though its location relative to anatomical landmarks (junction of the inferior temporal sulcus and the ascending limb of the inferior temporal sulcus) is relatively consistent.

With these issues in mind, the current study quantitatively compared the reliability of functional ROIs as a function of spatial normalization and group averaging. A test–retest measure of reliability was used in which two separate functional data sets were collected from the same participants and compared. Reliability was assessed by asking two questions. First, if a functional activation measure is used to define ROIs based on one data set, how well do those ROIs perform on that same measure in the second data set? Second, how well do ROIs from the two data sets align spatially? For both questions we evaluated the effects of (1) spatial normalization and (2) group averaging on reliability. The sample sizes and protocols were typical of current practice, maximizing the applicability of our results to plausible experimental designs.

Methods

Participants

Eleven participants (20–51 years old, 5 female) who were part of an earlier study (Zacks et al., 2001) were asked to participate in a follow-up study that included the functional runs analyzed here. Participants received $25 an hour for their participation and gave informed consent for these experimental procedures, which were approved by the

Washington University Medical School Human Studies Committee.

Design and procedure

Apparatus

Stimuli were projected with an LCD projector onto a white screen located at the head of the scanner bore. Participants viewed the stimuli through a mirror mounted on the head coil. The stimulus subtended 13.5° of the horizontal visual field and had a resolution of 640 × 480 pixels. Stimulus presentation was controlled by an Apple Power Macintosh computer (Cupertino, CA) with PsyScope experimental software (Cohen et al., 1993).

Behavioral tasks

FEF. Two stimulus conditions were used to localize FEF: visual *fixation* and cued *saccades*. The fixation stimulus consisted of a white background with a red fixation cross (two perpendicular lines, each 0.563° long and 0.084° wide) displayed at the center of the screen (see Fig. 1). The same fixation cross was used for the saccade condition, but rather than remaining stationary, the fixation cross appeared in a new random position (using the PsyScope random number generator) on the screen once per second to cue a saccade. For these localization runs, participants were asked to focus their eyes on the red fixation cross at the center of the screen but to move their eyes to fixate on the cross whenever it appeared in a new position. As such, the saccade condition required the participant to quickly move their eyes to new, random locations.

MT+. Three stimulus conditions were used to functionally localize MT+: *fixation*, a *still* visual pattern, and visual *motion* (see Fig. 1). During the fixation stimulus the same fixation cross used in the FEF localization tasks was presented on the screen. In addition to the fixation cross, the still condition consisted of 100 randomly positioned stationary black dots (0.084° radius). Under the motion condition the dots moved toward the edges of the screen at an angle determined by their radial position at a rate of 0.844° of the visual field per second. Each dot disappeared once it reached the edge of the screen and was replaced by a new dot at the center of the screen moving along the same radial path. Participants were asked to focus on the fixation cross throughout the entire localization run.

Data acquisition

Imaging was performed with a 1.5-T Siemens Vision MRI Scanner (Erlangen, Germany). Structural imaging included a high-resolution (1 × 1 × 1.25 mm) T1-weighted sagittal MP-RAGE (TR, 9.7 ms; TE, 4 s; flip angle, 10°; T1, 20 ms; TD, 200 ms) and a T2-weighted turbo spin echo (TSE) scan. The high-resolution MP-RAGE was used to compute each subject's definitive atlas transformation. The TSE data served as a registration intermediate (see below). A coarse (2-mm^3 voxel, 1.3-min) pre-fMRI MP-RAGE was used as the basis for automatic computation of functional imaging slice tilts and offsets parallel to the anterior–posterior commissure (AC-PC) plane. The functional data were acquired using an asymmetric spin–echo echo–planar (EPI) pulse sequence (flip angle, 90°; T2* evolution time, 50 ms). Whole brain coverage was achieved with 16 contiguous 8-mm slices parallel to the AC-PC plane (in-plane resolution of 3.75 mm). Each functional volume took 2.16 s to acquire.

FEF. Functional ROIs for FEF were obtained from separate localizer runs. The scanning sequences for these localizer runs are illustrated in Fig. 1. Stimuli alternated between blocks of the fixation task and blocks of the saccade task, with eight blocks of the saccade task in each run. An additional block of fixation was added to the beginning and end of the stimulus sequence in order to ensure that both localization runs began and ended with fixation. (The final block of fixation was not included in the scanning sequence for the first participant.) Condition order was counterbalanced across runs but not across participants: after the initial fixation block, the first localizer run always started with a block of the saccade condition and the second localizer run always started with a second block of the fixation condition. Each block consisted of eight 2.16-s volumes (frames). Four volumes of fixation were added to the beginning of each localizer sequence to allow for magnetization stabilization. Each FEF localizer run consisted of 148 volumes (56 volumes of the saccade task and 84 volumes of the fixation task) and lasted approximately 5 min and 20 s.

MT+. During the same scanning session, two BOLD runs were performed to functionally localize MT+. The scanning sequences for the MT+ localizer runs are illustrated in Fig. 1. The stimuli were presented in a blocked design, with four repetitions of each type of nonfixation block and nine repetitions of the fixation block. The presentation order of the still and motion conditions was counterbalanced across runs but not across participants; the sequence for the first run always started with the fixation followed by the still condition while the sequence for the second run always started with the fixation followed by the motion condition. For all but the first participant, 5 volumes lasting one TR were collected during the fixation blocks. Fixation blocks were presented before and after the still and motion stimulus blocks, during which 10 volumes were collected. Forty volumes of the still condition, 40 volumes of the motion condition, and 45 volumes of the fixation condition were collected in each MT+ localizer run. This protocol was the same for the first participant except that the fixation blocks lasted 4 volumes and the motion and still blocks lasted 8 volumes (resulting in 32 still volumes, 32 motion volumes, and 36 fixation volumes in each run). Four volumes of fixation were added to the beginning of each run to allow magnetization to stabilize. The MT+ localizer runs con-

sisted of 129 volumes and lasted approximately 4 min and 38 s.

Data analysis

Image processing followed standard procedures for our laboratory (Michelon et al., in press; Zacks et al., 2001, 2002, 2003). Preprocessing included (1) compensation for slice-dependent time shifts (135 ms per slice), (2) elimination of odd/even slice intensity differences due to interpolated acquisition, and (3) realignment of all data acquired in each subject within and across runs to compensate for rigid body motion (Ojemann et al., 1997).

For the atlas and group analyses, we registered the functional data to a standard stereotactic space by computing a sequence of affine transforms: first frame EPI to T2-weighted TSE to MP-RAGE to atlas representative target. The MP-RAGE to atlas transform was 3D affice (12 free parameters). The EPI to T2-weighted TSE transform allowed in-plane stretch to partially compensate for EPI distortion. Including the TSE in the sequence of registrations (EPI to TSE to MP-RAGE) reduced systematic EPI to MP-RAGE misregistration due to distortion and signal dropout (Ojemann et al., 1997). All transforms were combined by matrix multiplication and then applied to native space data. Reslicing the functional data in conformity with the atlas while simultaneously correcting for head movement involved only one interpolation. (The native space data also were resampled once during correction for head movement.)

Our atlas representative target image was made to conform to the Talairach Atlas (Talairach and Tournoux, 1988) using the method of Lancaster et al. (1995) starting with coregistered MP-RAGE data acquired in 12 normal young adults. Certain of the above enumerated transformation steps (EPI to TSE, TSE to MP-RAGE) involved cross-modal image registration. Here we used a locally developed algorithm based on the gradient matching strategy of Andersson et al. (1995). Our method differs from the original technique in taking into account relative gradient orientation in addition to gradient magnitude.

Following preprocessing, BOLD signals during each localizer run were estimated using the general linear model (Friston et al., 1995). A 6-mm full-width half-maximum (FWHM) Gaussian kernel was used to spatially smooth the data (Smith, 2001). Linear trends over single runs were removed by regression but otherwise the time series was not temporally filtered.

ROI localization

Our localization procedure was unique in two ways. First, a single scan was used to identify an ROI, while one to four scans are typical in the literature (Kourtzi et al., 2002; Huk et al., 2002). Because condition order was counterbalanced across scans in our design, the different localizer runs (and therefore the ROIs) may be affected by condition order, fatigue, and habituation. Second, we chose to emphasize a more automated approach to functional localization in order to take advantage of its independence from highly specified anatomic criteria.

We conducted reliability analyses for three methods for functional localization. For the *native space individual* analyses, functional ROIs were identified for each individual based on their functional data prior to atlas normalization. For the *atlas space individual* analyses, functional ROIs were identified for each individual based on the atlas-normalized functional data. For the *group* analyses, ROIs were identified for the group as a whole using a within subjects random effects analysis. In this section we describe the methods for identifying ROIs using each of these methods in detail.

Anatomical restriction

Although regions were defined primarily based on functional criteria for all analyses, we restricted the regions with broad anatomic criteria. MT+ excludes frontal areas that are responsive to movement (Culham et al., 2001), and FEF excludes parietal areas known to be involved in eye movements and attention (Petit and Haxby, 1996). We restricted the search space for FEF to those voxels anterior to the central sulcus and dorsal to the line connecting the anterior and posterior commissures. We restricted the search space for MT+ to those voxels posterior to the Sylvian fissue and the central sulcus that were not part of the cerebellum or deep brain structures. These anatomical landmarks were identified on the group atlas brain and applied to each individual's data after atlas normalization. To define individual ROIs in native space, these same atlas-based landmarks were applied to each individual's data in acquisition space by reversing the algorithms used to align the individual's data to the standard atlas.

Native space individual localization

In these analyses, regions were identified for each participant using data in the native scanner acquisition space ($3.75 \times 3.75 \times 8$ mm voxels). To identify FEF, we computed the value of the contrast comparing saccades to fixation for each voxel within the anatomically defined search space. The contrast value at each voxel was converted to a *t* statistic by dividing the value by its estimated standard deviation and then converted to an equiprobable *z*. Voxels with a *z* statistic greater than 4.5 (corresponding to $P < 0.001$) were included in the ROIs. To identify MT+, we employed a two-stage thresholding process. First, to identify voxels that responded to motion we computed the value of the contrast comparing moving dots to still dots for each voxel within the anatomically defined search space. These contrasts were converted to equiprobable *z* statistics and thresholded at $z = 4.5$. Second, to exclude voxels with a large response to nonmoving stimuli (Tootell et al., 1996), we computed the value of the contrast comparing still dots

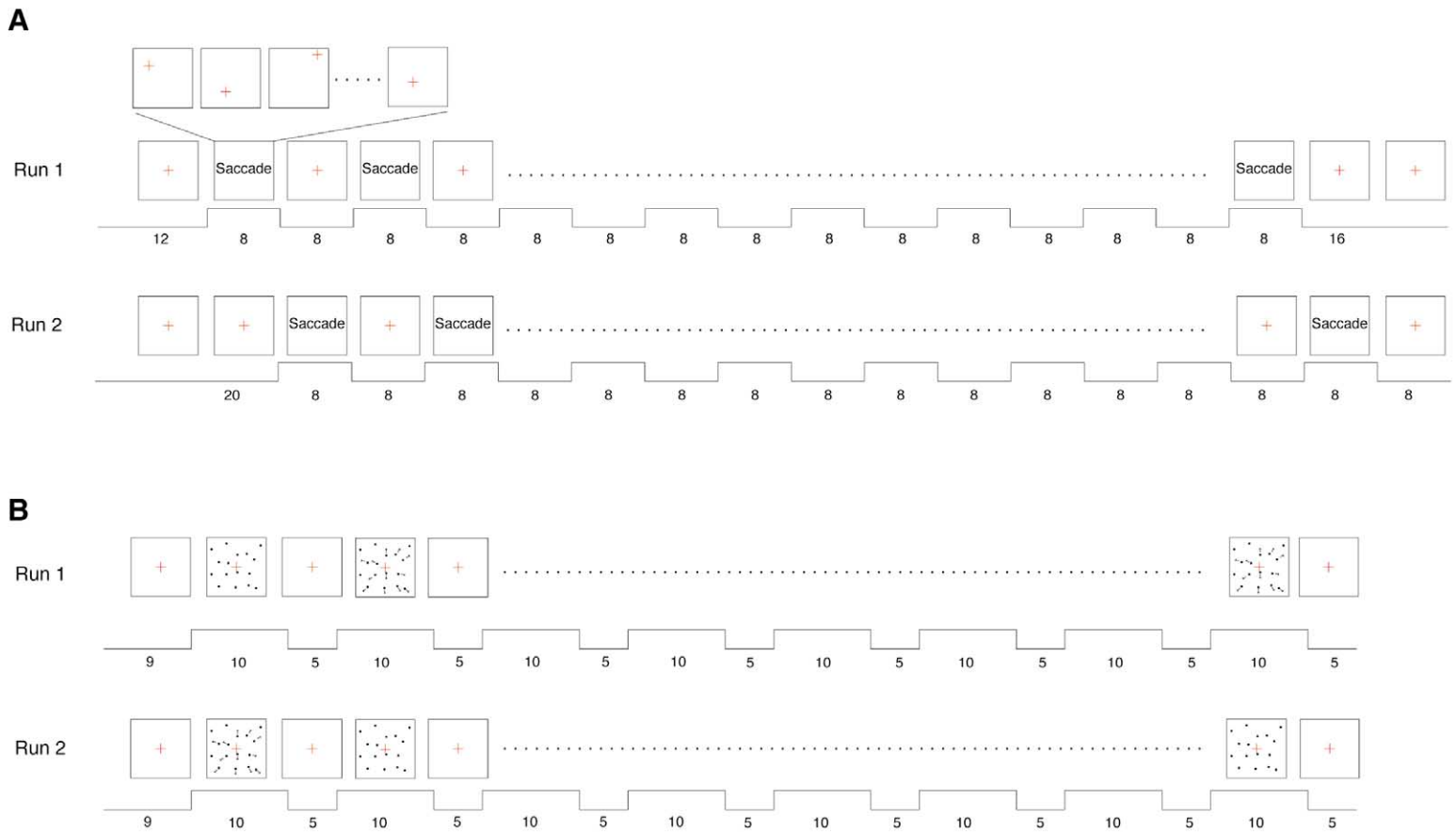


Fig. 1. Stimuli and tasks used to functionally localize the FEF (A) and MT+ (B) ROIs were presented in a block design during functional imaging runs. All functional runs began with 4 volumes of fixation immediately followed by 1 complete block of fixation. In addition, all functional runs ended with a block of fixation. Numbers under each line segment indicate the number of volumes within the block. Blocks in the FEF localization sequence were 8 volumes long and alternated between fixation and saccade. Within each run, there were 10 blocks of the fixation task and 8 blocks of the saccade task. Three types of blocks were used to localize MT+: fixation, still, and motion. Every block of the still and motion conditions lasted 10 volumes and was immediately followed by 1 block of fixation lasting 5 volumes. Blocks alternated between the still and motion conditions, with the fixation blocks interspersed between these.

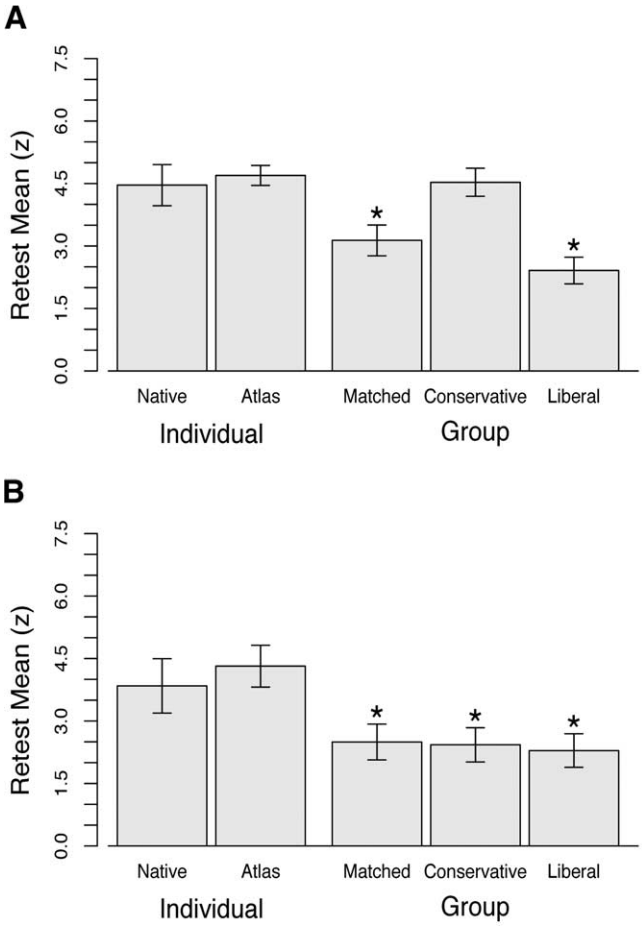


Fig. 2. Average retest means of the individual contrasts within each FEF (A) and MT+ (B) ROI. Error bars represent 1 SEM. Asterisks (*) indicate values that are significantly different than the individual atlas space value, $P < 0.01$.

to visual fixation for each voxel and excluded those with a *z* statistic *greater* than 2.0 (corresponding to $P = 0.023$) from the ROIs. This exclusion reduced the likelihood of including in the MT+ ROIs visual areas that are not specifically sensitive to motion, e.g., V1, V2, and V3 (Tootell et al., 1995). Thresholds were adjusted until the resulting ROIs appeared to be the appropriate size, contained mostly contiguous voxels, and, in the case of MT+, excluded the calcarine cortex for the majority of participants.

Atlas space individual localization

In these analyses, regions were identified for each participant using data in stereotactic space. As with native space analyses, the data were resampled (3D cubic spline) once to correct for head movement. However, this resampling also mapped the data into the stereotactic atlas space (3 mm^3), rather than retaining the native acquisition space (3.75 × 3.75 × 8 mm). The main difference between the atlas and native space analyses was in the different voxel dimensions. The change in coordinate frames from native to atlas space could also affect localization.

Group localization

Group ROIs were calculated using the same contrasts as those for the individual ROIs. The same linear models were used as for the atlas space individual analyses. However, rather than calculating *z* statistics for each individual, the contrast magnitudes of each voxel were submitted to group-level random effects *t* tests.[1] That is, for each voxel the contrast magnitude for each of the 11 participants was calculated, and the mean of this sample distribution of contrast magnitudes was divided by its standard error to construct a *t* statistic. For each voxel, a *t* statistic was calculated using the mean of the difference between activity under the two conditions for each individual and the standard error of that difference across participants. These *t* statistics were converted to *z* statistics and thresholded.

Because the *z* statistic images generated through this analysis are based on estimates of variability across participants rather than across runs for each participant, these images are not directly comparable to those generated at the individual level. This leaves open the possibility that differences between the individual and group analyses could arise from incidental differences in the conservativeness of the thresholds. To address this concern, we created three sets of thresholds for FEF and for MT+, one set *matched* in volume as closely as possible to the individual analyses, one more *liberal*, and one more *conservative*. The thresholds for the matched ROIs were determined through an iterative procedure whereby we equated the volumes of the group ROIs with the mean volumes of the individual ROIs. Liberal and conservative group ROIs were then created by relaxing and restricting these thresholds, respectively. Using this procedure, the matched group FEF ROI was created with the threshold for the saccade–fixation *z* statistic image set to 2.7 ($P = 0.003$). For the MT+ matched group ROI, we adjusted the still-fixation threshold until the mean number of voxels excluded from both group ROIs from both runs was as close as possible to the mean number of voxels excluded from the individual atlas space ROIs. This resulted in the still-fixation threshold for the matched group ROI being set to 1.9 ($P = 0.029$). With these voxels excluded, we then matched the ROIs by adjusting the motion–still contrast *z* statistic threshold, resulting in a threshold of 2.8 ($P = 0.003$).

Thresholds for the liberal and conservative group ROIs were determined by adjusting the statistical thresholds until

[1] A fixed effects analysis is appropriate when generalizations about a voxel's activity are made to the sample of participants rather than to the population from which they were drawn. However, it is strongly influenced by relatively large changes in activity unique to an individual in the sample. We performed a parallel set of analyses using group ROIs identified through a fixed effects analysis in order to determine whether this approach would lead to different results. These results, however, were comparable to those we obtained when using the group ROIs identified through the random effects analysis. Because we were wary of defining group ROIs that were potentially driven by the task-related activity of a voxel in a single individual, we felt the random effects analysis was more appropriate.

the size of the ROIs was at the large (liberal) or small (conservative) end of the distribution of individual ROIs. This resulted in the liberal thresholds for the group FEF ROIs being set to 2.2 ($P = 0.014$) for the saccade–fixation contrast. The liberal group ROIs were created for MT+ with thresholds set to 2.4 ($P = 0.008$) for the still-fixation contrast and 2.3 ($P = 0.011$) for the motion–still contrast. The conservative group ROIs were created for area FEF by setting the random effects saccade–fixation threshold to 3.5 ($P < 0.001$). A set of conservative group MT+ ROIs was created by setting the still-fixation threshold to 1.4 ($P = 0.919$) and the motion–still threshold to 3.3 ($P < 0.001$).

Measures

Three measures were used to evaluate reliability: *retest mean*, *overlap*, and *distance*. For the retest mean measure, each individual FEF or MT+ ROI was defined by data from one of the two scanning runs (the *test* run) as described previously. This ROI was then applied to data from the same participant's other scanning run (the *retest* run). The data used to calculate the retest mean of an ROI were therefore independent of the data used to define the ROI. For FEF, we calculated the mean value within the ROI of the contrast comparing the saccade condition to the fixation condition during the independent retest run. For MT+ we calculated the mean value within the ROI of the contrast comparing moving dots to still dots. To the extent that a region identification method is reliably identifying its target area, these values should be high. For each ROI, analyses were performed using the first run as the test run and the second run as the retest run, and vice versa, and the two were averaged.

For the overlap measure, one FEF or MT+ ROI was defined with data from each of the two scanning runs. We then calculated the proportion of overlapping voxels by dividing the number of voxels included in both ROIs by the total number of voxels included in either ROI. In other words, this measure was equivalent to the volume of the intersection of the two ROIs divided by the volume of the union of the two ROIs. This value should be highest when the location and extent of the ROI remain the same across runs, indicating a highly reliable region identification method. To emphasize, this measure is influenced by the difference in the volume of the two ROIs as well as the difference in location. A small ROI could be completely encompassed by a larger one and yield a low degree of overlap.

For the ROI defined by each run, we calculated its center of mass by averaging the location of each voxel in the ROI weighted by its *z* statistic for the relevant contrast (saccade–fixation for FEF, motion–still for MT+). Center of mass was used to obtain data for three different analyses. The first looked at the effects of spatial normalization and statistical group on the distance between the centers of mass of ROIs created before and after spatial normalization and statistical grouping (*distance analysis*). The second compared the distance between two individual atlas space ROIs to the distance between individual and group ROIs (*distance to group analysis*). The third and last analysis examined how well the group ROI represented the central tendency of the individual atlas space ROIs (*location analysis*). This analysis used multivariate analysis of variance (MANOVA) to determine whether the difference between the coordinates of the individual centers of mass and the group centers of mass was equal to zero.

The distance between ROI centers of mass was always computed using ROIs from separate localizer runs. The centers of mass were treated as points in three-dimensional Euclidean space and the distance between them was calculated by entering their coordinates into the following formula: $D = [(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2]^{1/2}$. For the distance analysis, these values were averaged across runs and hemispheres, resulting in one measure of distance for each individual atlas and native space ROI and for each type of group ROI. Highly reliable localization methods should produce regions whose location and activity change little across localization runs: the distance between the centers of mass should be minimal. Similarly, for the distance to group analysis, the best localization procedure will produce the best estimate of the center of mass of the ROI on future localization runs. For this analysis we calculated the distance from the center of mass of the individual ROI from one localizer run to the center of mass of the group ROI from the other localizer run, averaging across runs and hemispheres. This resulted in one measure of individual to individual distance and individual to group distance for each participant.

Results

In the sections that follow, we first report comparisons between regions defined based on individual data in native acquisition space and in atlas space. This comparison provides a test of the effect of spatial normalization on reliability. To address the effect of group averaging on reliability, we then report comparisons between regions defined with individual data in atlas space to those defined with group data (also in atlas space). For all three reliability measures, the native space and atlas space individual localization methods produce one observation per ROI per participant. Paired samples *t* tests were used to compare the overlap, distance, and retest means obtained for each individual when using data in native space and when using data in atlas space. When comparing the individual atlas space method to the group method, the retest mean and distance to group measures produce one observation per participant, so paired samples *t* tests were again used. However, for the overlap and distance measures the group localization methods produce only one observation per ROI. One-sample t

Table 1
Talairach and Tournoux (1988) coordinates for centers of mass of individual and group functional ROIs

Region	Method	Left			Right		
		x	*y*	*z*	*x*	*y*	*z*
FEF	Individual	−22 (1.1)	−3 (1.4)	51.3 (1.1)	23 (1.5)	0.6 (1.6)	50.5 (1.7)
	Group						
	Matched	−23.1	−2.3	47.4	23.6	−3.4	53.8
	Conservative	−23.1	−7.4	51.3	24.4	−4.2	57
	Liberal	−23.3	0.9	46.5	24.1	−0.9	51.4
MT+	Individual	−36 (3.3)	−71.8 (2)	5.5 (2.1)	36.3 (3)	−69.4 (2.6)	5.5 (1.5)
	Group						
	Matched	−37.6	−73.3	2.2	36.7	−68.7	4.6
	Conservative	−39.4	−77.2	3.2	37.8	−68.7	4.3
	Liberal	−36.1	−68.3	4.5	35.9	−66.8	3.8

Note. Reported values were averaged across ROIs from separate localizer runs. Means are reported for the individual ROIs in atlas space with standard errors in parentheses.

tests were used to compare the sample of overlap and distance measures from the individual atlas space ROIs to those single measures, using the value of that measure from the group ROI as the mean of the population under the null hypothesis. This is analogous to the method typically used to compare a single case study (in this case, the group ROI) to a sample (in this case, the individual ROIs). We note that this technique allows for generalization to other samples of individual ROIs, but does not model variability in the estimate of the group ROI.

Effects of spatial normalization

Individual FEF and MT+ regions could be localized in both hemispheres in at least one run for all observers. For FEF, data from native space missed one hemisphere during one run for two participants. Data from atlas space missed one hemisphere during one run for one participant. For MT+, data from native space failed to identify a region during one run for two participants and missed one hemisphere during one run for two participants. Data from atlas space missed one hemisphere during one run for two participants. Separate analyses on the left and right hemispheres produced similar results to those reported below, where measures from both hemispheres have been averaged together.

Transforming the functional data to stereotactic space prior to analysis had little effect on the regions obtained. As is detailed below, the two sets of ROIs showed similar levels of reliability. The ROIs based on data in native space were somewhat larger.

Retest mean

The retest means and standard errors for the FEF and MT+ native and atlas space ROIs are shown in Fig. 2. No differences in the average retest mean for the ROIs identified in atlas and in native space were found in FEF ($t(10) = 0.596$, $P = 0.564$) or in MT+ ($t(10) = 1.781$, $P = 0.105$).

Distance

No significant differences were found in the distances between the centers of mass of the FEF ROIs identified in atlas space (M = 11.172, SD = 7.822) and native space (M = 11.33, SD = 3.565; $t(8) = 0.115$, $P = 0.911$). However, the distance between the centers of mass of MT+ ROIs identified in atlas space (M = 6.898, SD = 5.628) was significantly smaller than between those identified with data in native space (M = 8.46, SD = 6.328; $t(6) = -2.837$, $P = 0.03$). The centers of mass of the atlas space ROIs can be found in Table 1. The distances between the centers of mass in the FEF and MT+ ROIs are illustrated in Fig. 3.

Overlap

The mean proportion of overlapping voxels in the MT+ and FEF ROIs can be found in Table 2. No differences in the mean proportion of overlapping voxels in the individual ROIs identified in atlas and in native space were found for either the FEF ROIs ($t(9) = -0.479$ $p = 0.643$) or the MT+ ROIs ($t(8) = 1.249$, $P = 0.247$).

Region size

The mean sizes and standard errors of the individual native and atlas space ROIs are reported in Table 3. The mean size of the FEF native space ROIs was larger than the mean size of the FEF atlas space ROIs when created with data from the first localizer run ($t(10) = -2.281$, $P = 0.046$) and from the second localizer run ($t(10) = -3.014$, $P = 0.013$). Though on average the MT+ native space ROIs were larger than the MT+ atlas space ROIs, this difference only approached statistical significance (run 1, $t(10) = -1.877$, $P = 0.09$; run 2, $t(10) = -1.916$, $P = 0.08$). Substantial individual differences existed in the individual ROIs for both FEF and MT+, five examples of which are illustrated in Fig. 4.

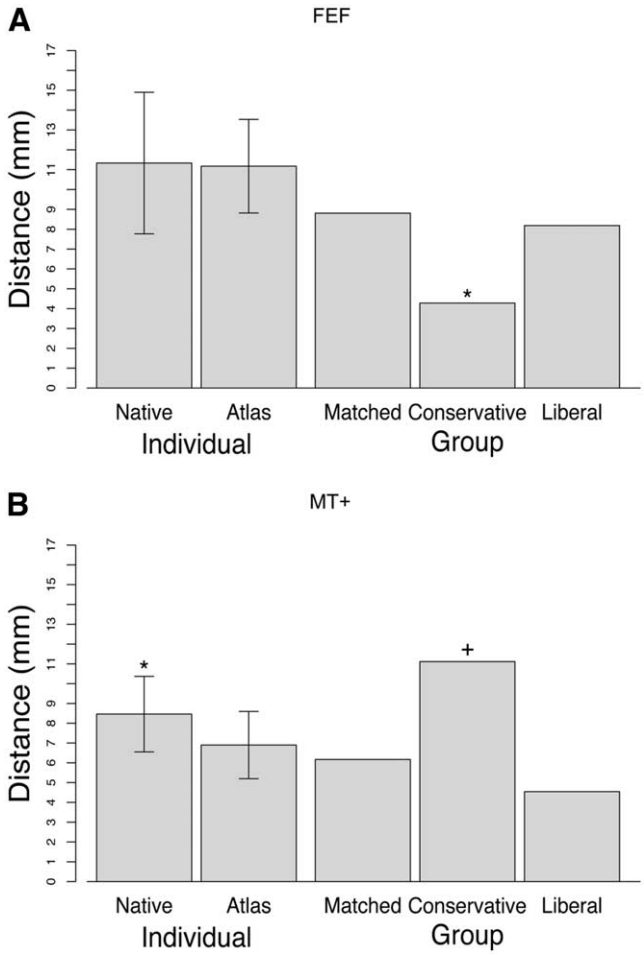


Fig. 3. Mean distance between the centers of mass of ROIs defined by two separate localizer runs for the FEF (A) and MT+ (B) regions. Error bars represent 1 SEM. Asterisks (*) indicate values that are significantly different than the individual atlas space value, $P < 0.05$. The cross (+) indicates a difference that was nearly significant, $P = 0.055$.

Effects of statistical grouping

Left and right MT+ and FEF ROIs were identified based on the matched, liberal, and conservative criteria. Overall, region identification at the individual level outperformed that at the group level across the matched, liberal, and conservative thresholds. For all statistical comparisons, we report the results from the comparison of the individual analysis to the best performing group analysis; if that difference was not statistically significant, we also indicate which of the others did differ significantly.

Retest mean

The retest means of the saccade–fixation contrasts for those voxels within the individual and group FEF ROIs are illustrated in Fig. 2. Fig. 2 also illustrates the retest means of the motion–still contrasts within the MT+ ROI. The individual atlas space ROIs yielded the highest retest means in the saccade–fixation contrast in the FEF ROI (M = 4.693, SD = 0.0795) and the motion–still contrast in the MT+ ROI (M = 4.319, SD = 1.663). However, the retest means from the conservative group FEF ROI (M = 4.529, SD = 1.121) were not significantly different than those obtained when using the individual FEF ROIs ($t(10) = 0.764$, $P = 0.462$). A statistically significant difference was found when comparing the matched group FEF ROIs (M = 3.133, SD = 0.0795) to the individual ROI retest means ($t(10) = 6.358$, $P < 0.001$). When applied to the individual motion–still contrasts, the retest mean of the matched group MT+ ROI (M = 2.497, SD = 1.428) was significantly less than that of the individual MT+ ROIs ($t(10) = 6.565$, $P < 0.001$).

Table 2
Proportion of overlapping voxels of functional ROIs, each defined with data from a single localization run

Region	Individual		Group		
	Native	Atlas	Matched	Conservative	Liberal
FEF	.349 (.052)	.340 (.044)	0.268	0.164	0.265
MT+	.227 (.055)	.263 (.055)	0.064	0.005	0.243

Note. Mean values and their standard errors (in parentheses) are reported for the FEF ROIs of 10 participants and MT+ ROIs of 9 participants. Individual data were removed if an ROI could not be identified in either run when the data were in either native or atlas space. Data removed due to a failure to find an ROI in native space were replaced for the individual to group comparison. This had a small effect on the mean and standard error of the atlas space FEF ROIs (M = 0.313, SE = 0.049) and MT+ ROIs (M = 0.219, SE = 0.177). The proportion of overlapping voxels was defined as the number of voxels common to both functional ROIs divided by the number of voxels in either ROI and reflects differences in both the location and extent of the ROIs.

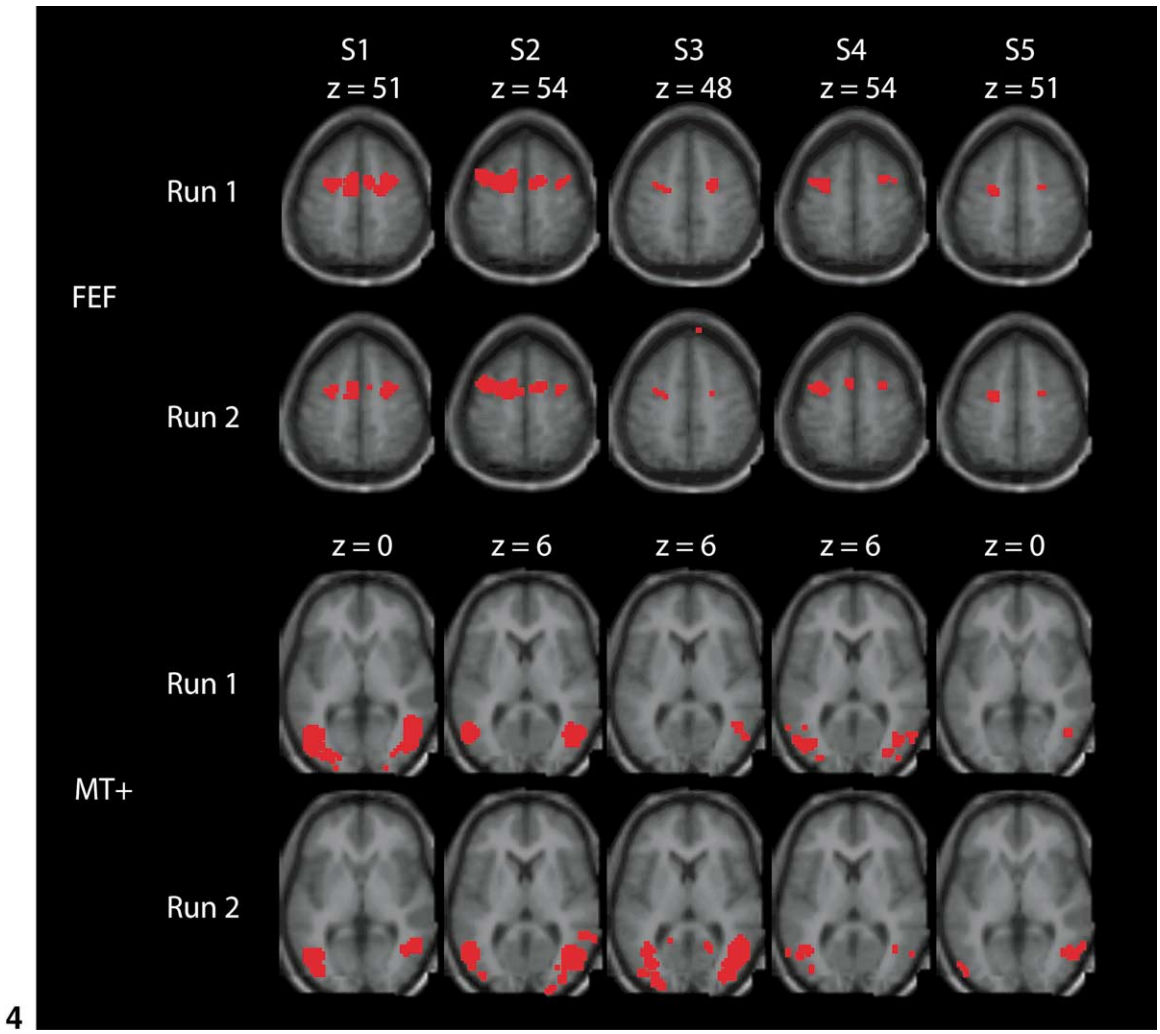
Distance to group

The Talairach and Tournoux (1988) coordinates of the individual atlas space ROIs and the group ROIs are listed in Table 1. Additionally, the centers of mass of every ROI identified in atlas space are illustrated in Fig. 5.

Table 3
Volume of functional ROIs (cm^3)

Region	Individual		Group		
	Native	Atlas	Matched	Conservative	Liberal
FEF					
Run 1	10.95 (2.74)	8.15 (1.7)	8.75	1.43	16.82
Run 2	6.61 (2.45)	5.34 (1.47)	3.51	0.49	8.21
Change	4.694 (1.428)	3.373 (.887)	5.238	0.945	8.613
Overlap	4.79 (1.78)	3.6 (1.12)	2.59	0.27	5.24
MT+					
Run 1	13.45 (4.81)	10.87 (3.61)	5.37	0.59	22.71
Run 2	16.36 (5.54)	13.16 (3.89)	19.44	4.81	45.33
Change	14.124 (4.471)	9.062 (2.965)	14.067	4.212	22.626
Overlap	6.43 (2.16)	5.35 (1.68)	1.49	0.03	13.31

Note. Volume change and the volume of the overlapping portions of these ROIs are reported to further illustrate the consistency of each localization method. Reported values are in cubic centimeters, with means and standard errors for the individually defined ROIs in atlas and native space.



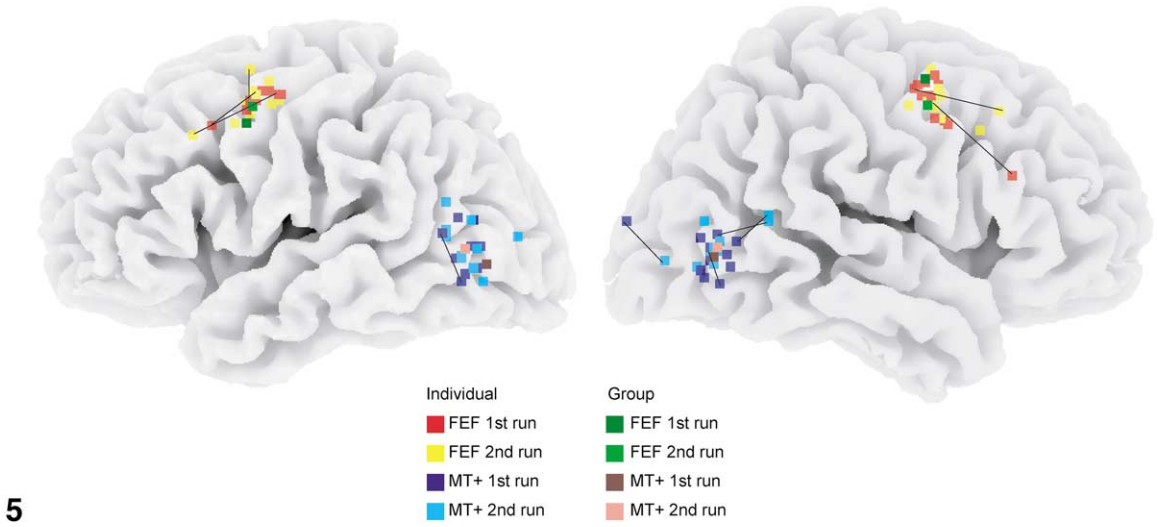


Fig. 4. Examples of individual ROIs in atlas space. Selected slices of ROIs created from the first and second localization runs for FEF and MT+ are illustrated for five participants. These ROIs are superimposed on an anatomical image created by averaging the anatomy of all participants in the sample. These participants were selected to demonstrate variability across localization runs and are not representative of the entire sample. The location and extent of the individual ROIs showed some variability across localization runs, particularly those for MT+. This is probably due to less power in the MT+ localization runs than the FEF localization runs, since these were divided among three rather than two task conditions. In addition, the calcarine cortex was not entirely excluded from the MT+ ROIs for all participants.

Fig. 5. Centers of mass of individual and matched group ROIs. Centers of mass for each individual atlas space ROIs and matched group ROIs are plotted over lateral views of the right and left hemispheres. Note that most of the ROIs for FEF and MT+ are located in the depth of a sulcus. Individual ROIs that moved more than 10 mm across runs are connected to indicate correspondence. This figure was produced using the CARET software (Van Essen et al., 2001; Van Essen, 2002).

Of primary interest when evaluating the effects of statistical grouping on characterizing the location of an individual ROI is the question: Is the reduction in variability achieved by grouping across participants advantageous given the reduced degree to which the group location captures the location of any individual's ROI? This can be answered by comparing the distance between one of an individual's ROIs and their other ROI to the distance between that individual's ROI and the group ROI. As noted previously, individual ROIs from one run were compared to the group ROI from the other run to provide an unbiased estimate. Though significant differences were found in only one comparison, the distance between the individual ROIs tended to be smaller than the distance from the individual ROIs to the group ROIs. Fig. 6 illustrates that the distances between the individual FEF ROIs were not significantly different than their distances to the group ROIs (matched group, M = 10.712, SD = 0.857; $t(9) = 0.272$, $P = 0.791$). However, it also illustrates a different outcome in the MT+ ROIs. In this region the liberal group ROI showed the best performance (M = 9.631, SD = 0.983). The distance between this group ROI and the individual ROIs showed a slight trend toward being greater than the distance between the two individual ROIs ($t(8) = -1.853$, $P = 0.101$). The distance between the individual MT+ ROIs was significantly less than the distance between the individual and conservative group MT+ ROIs (M = 11.203, SD = 1.151; $t(8) = -3.765$, $P = 0.006$).

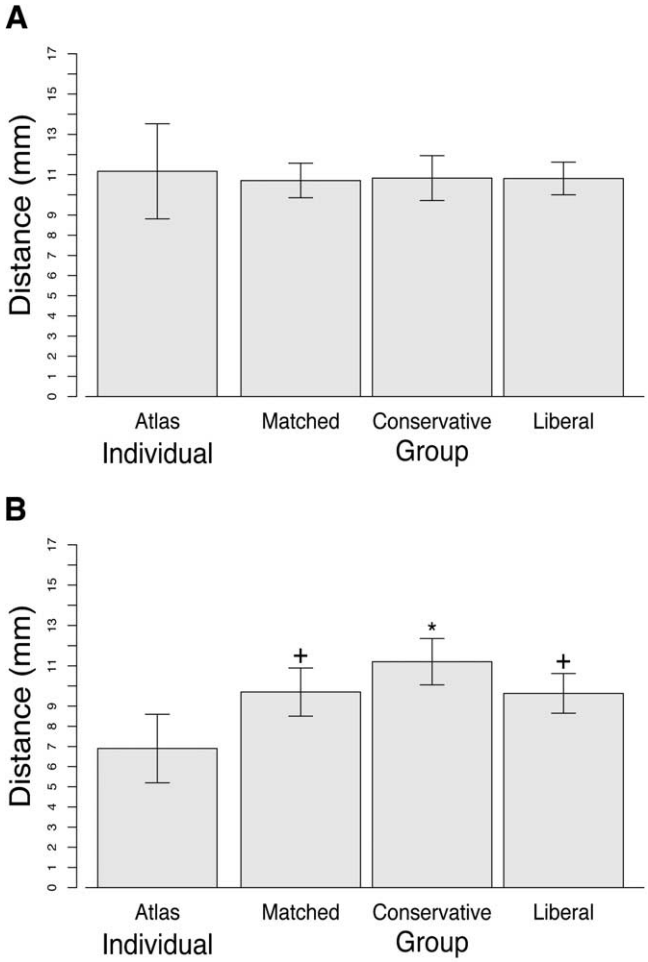


Fig. 6. Distance from the individual ROIs from one localization run to the individual and group FEF (a) and MT+ (b) ROIs from the other localization run. Error bars represent 1 SEM. Asterisks indicate values that are significantly different than the individual atlas space value, $P < 0.01$. Crosses indicate values that are marginally reliably different than the atlas space value, $P < 0.1$.

Distance

The absolute distance between group ROIs based on the two different data sets is of secondary interest, but it provides a straightforward reliability measure for the group localization approach. As illustrated in Fig. 3, the conservative group FEF ROI was the most consistent for this region in terms of the location of its center of mass across runs (distance, 4.281 mm). This distance was significantly better than the mean distance of the centers of mass of the individual FEF ROIs (M = 11.172, SD = 7.822; $t(9) = 2.786$, $P = 0.021$). The opposite pattern can be seen for the distances between the MT+ ROIs, also illustrated in Fig. 3. For MT+, the distance between the liberal group ROI was the smallest, 4.541 mm. This value was not significantly smaller than the mean center of mass distance of the individual MT+ ROIs (M = 6.898, SD = 5.628; $t(8) = 1.256$, $P = 0.244$). However, the distance between the conservative group MT+ ROIs, 11.113 mm, was greater than the mean individual distance ($t(8) = -2.247$, $P = 0.055$).

Overlap

The proportions of overlapping voxels for all ROIs are listed in Table 2. The values reported in this table are a reflection of the differences in both the location and extent of the two ROIs and may appear low due to the changes in ROI volume across runs, which are listed in Table 3. The group ROIs with the highest proportion of overlapping voxels for FEF and MT+ were not statistically greater than the individual atlas ROIs. In addition, the best set of group ROIs was different for FEF and MT+. The matched group FEF ROIs showed the most overlap (0.268) across runs. No statistically significant difference was found between the proportion of overlapping voxels of these ROIs and the mean proportion of overlapping voxels of the individual FEF ROIs (M = 0.313, SD = 0.162, $t(10) = 0.91$, $P = 0.383$). Likewise, the group ROI with the highest proportion of overlapping voxels for MT+, the liberal group, did not significantly differ from the individual MT+ ROIs (M = 0.219, SD = 0.177; $t(10) = -0.449$, $P = 0.662$). However, the proportion of overlapping voxels in the individual MT+ ROIs was significantly greater than that of the matched group ($t(10) = 2.915$, $P = 0.015$). This difference was also statistically significant when comparing the individual FEF ROIs to the worst set of group FEF ROIs, the conservative set ($t(10) = 3.047$, $P = 0.012$).

Table 4
Differences between the center of mass of individual ROIs in atlas space and the group ROIs

Region	Method	Mean location difference (mm)			Standard error		
		x	*y*	*z*	*x*	*y*	*z*
FEF							
	Matched	−0.395	−2.349	−0.027			
	Conservative	−0.025	−5.236	−3.622	1.003	1.231	1.018
	Liberal	−0.231	0.546	1.626			
MT+							
	Matched	0.309	−2.448	2.444			
	Conservative	−0.059	−4.385	2.106	0.646	1.19	1.751
	Liberal	0.686	1	1.717			

Note. Positive values indicate the group ROI was located to the left of (*x*) or posterior (*y*) or dorsal (*z*) to the individual centers of mass.

Location

Finally, to determine whether the location of the group ROIs differed significantly from those identified for each individual, we performed a MANOVA on the difference between the *x*, *y*, and *z* coordinates of the individual ROI centers of mass and the coordinates of the group ROI centers of mass. The mean differences in the location of the group and individual ROIs are reported in Table 4. The results of these analyses indicate that the locations of the group centers of mass in FEF and MT+ represented the central tendency of the individual ROIs. With the exception of the conservative group ROIs, the locations of the individual FEF and MT+ ROIs appear to be similar to the location of the group ROIs. The mean difference in the location of the liberal group ROI was less than 2 mm in all three dimensions for FEF (−0.231, 0.546, 1.626) and for MT+ (0.686, 1, 1.717). The liberal group ROI's location was not significantly different than the mean location of the individual FEF ROIs (see Table 1 for these values; Wilks's $\lambda = 0.756$, $F(3,7) = 0.753$, $P = 0.555$) or the individual MT+ ROIs (Wilks's $\lambda = 0.716$, $F(3,6) = 0.794$, $P = 0.54$). However, the conservative group ROI was significantly different than the central tendency of the individual FEF ROIs (Wilks's $\lambda = 0.318$, $F(3,7) = 5$, $P = 0.037$) and showed a trend toward significance in MT+ (Wilks's $\lambda = 0.353$, $F(3,6) = 3.673$, $P = 0.082$).

Region size

A set of analyses was performed to verify that the thresholding procedures we used produced the group ROIs we expected. The volumes of the group ROIs are reported in Table 3 along with the means and standard errors of the individual atlas space ROIs. As can be seen in the table, the matching procedure was successful in creating matched group ROIs that closely tracked the typical size of the individual ROIs: the liberal and conservative threshold adjustments successfully moved the ROI sizes to the upper and lower bounds of the individual size distributions, respectively.

The average volume of the two matched group FEF ROIs was not significantly different than the volume of the individual FEF ROIs averaged across runs ($t(10) = 0.411$, $P = 0.689$). The average volume of the liberal group FEF ROIs was greater than the individual FEF ROIs ($t(10) = -3.851$, $P = 0.003$), which were greater than the average volume of the conservative group FEF ROIs ($t(10) = 3.862$, $P = 0.003$). The thresholding procedures were equally successful for the group MT+ ROIs. Here again, the average volume of the matched group MT+ ROIs was not significantly different than the average volume of the individual MT+ ROIs ($t(10) = -0.124$, $P = 0.903$), the average volume of the liberal group MT+ ROIs was significantly greater than the individual MT+ ROI volume ($t(10) = -6.968$, $P < 0.001$), and the average volume of the conservative group MT+ ROIs was significantly less than the individual MT+ ROI volume ($t(10) = 2.94$, $P < 0.015$). The matched group ROIs are illustrated in Fig. 7.

Discussion

Several data analysis options exist when investigating the role of a functionally defined ROI in the performance of a cognitive task. The purpose of the current research was to look at the effects of two of these options, spatial normalization and group averaging, on the reliability of functional localization as it is used in these sorts of investigations: as a methodological tool. The results argue against group averaging and are equivocal about the use of spatial normalization. Overall, functional ROIs defined for each individual in atlas space were more reliable than group ROIs and were essentially equivalent to ROIs defined in native acquisition space (see Table 5). Because spatial normalization confers advantages for reporting the location of ROIs and for making comparisons across individuals, identifying ROIs in individuals using spatially normalized data was the best method examined here.

Comparison of individual native space to individual atlas space ROIs

We found no advantages associated with defining ROIs using functional data in native space relative to functional ROIs defined for each individual using spatially normalized data. Consistency in the location of the centers of mass of the individual native and atlas space ROIs, as measured by overlap and distance, suggests that these two methods are equally good at identifying the same regions in separate localizer runs. Miki et al. (2000) found similar results with their examination of the effects of a two-step normalization process on the reproducibility of BOLD response to visual stimulation. Our results confirm and extend these findings to FEF and MT+ and demonstrate that one-step spatial nor-

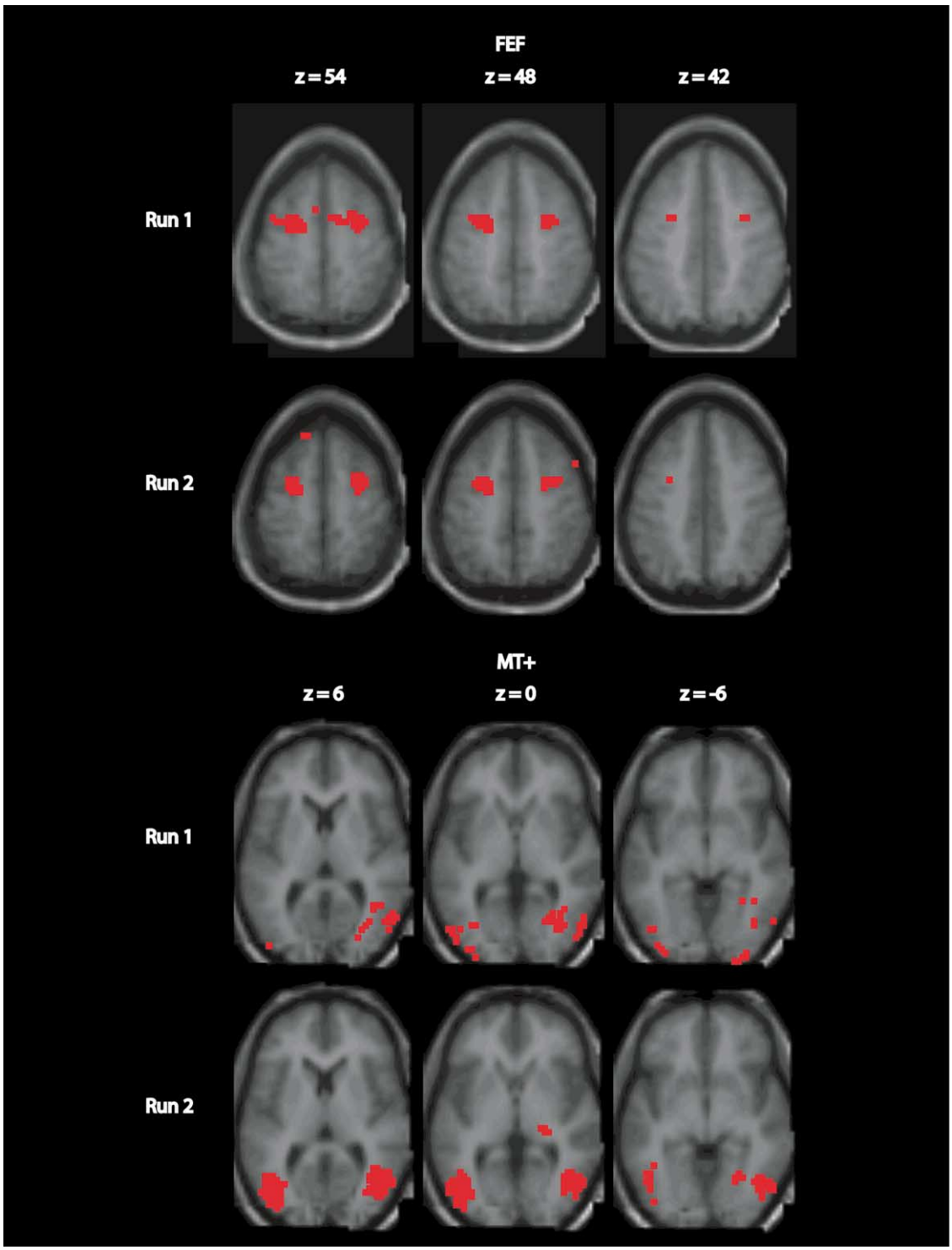


Fig. 7. Functional ROIs of the FEF and MT+ regions defined through a random effects group level analysis. Thresholds for these ROIs were set in order to match their volume as closely as possible to the mean volume of the individual ROIs. Three slices of each region created from both localization runs are shown. These ROIs are superimposed on an anatomical image created by averaging the anatomy of all participants in the sample.

malization has little effect on the ability of functionally localized ROIs to identify task-related activity in other imaging scans. This finding is crucial to the validity of inferences that must be made when examining task-related activity in functional ROIs. ROIs defined in atlas space benefit from the fact that they can be described in terms of standard coordinates and compared to ROIs in other individuals and in other studies.

Individual ROIs defined in native space were larger than those defined in atlas space. This result may reflect coarser

Table 5
Summary of results

Comparison	Individual		Group		
	Native	Atlas	Matched	Conservative	Liberal
Retest mean					
FEF	.	+		.	
MT+	.	+			
Overlap					
FEF	+	.	.		.
MT+	.	.	NA	NA	+
Distance					
FEF			.	+	.
MT+	.	.	.		+
Distance to group					
FEF	NA	.	+	.	.
MT+	NA	+	.		.

Note. A plus sign (+) indicates the method that yielded the most advantageous results (the highest retest mean and proportion of overlapping voxels and the smallest distance). Localization methods that did not differ significantly from this method are indicated with a period. Localization methods that were significantly different ($P < 0.05$) are left blank. Some comparisons could not be performed due to too few observations (overlap MT+ comparison of group observations) or the inability to compare coordinates in atlas space to coordinates in native space (distance to group comparison).

sampling ($3.75 \times 3.75 \times 8$ mm vs 3 mm^3). Another possibility is that the small amount of blurring introduced by spatial normalization eliminated isolated voxels with weak signals. These differences, however, were modest.

The present results support earlier work demonstrating large changes in BOLD activation across runs within the same scanning session and individual (Moser et al., 1996; Rombouts et al., 1997, 1998; Miki et al., 2000; McGonigle et al., 2000; Machielson et al., 2000; Specht et al., 2003). Within our data set, intraindividual variability is likely to have arisen from condition order, fatigue, and habituation to the task and stimuli.

One aspect of the overall reliability estimates may be somewhat misleading. As can be seen in Table 1, the overlap between the ROIs we defined on separate localizer runs was small. However, this was likely due to large differences in the *sizes* of the ROIs identified in each of the two different localizer runs (see Table 2). The upper bound on the overlap measure is the ratio of the size of the smaller region to the size of the larger region. Given that one fixed set of criteria was used across runs, changes in the size of the regions are not surprising. These overlap measures are in relatively good agreement with those of other studies (Rombouts et al., 1998; Miki et al., 2000).

Comparison of individual to group ROIs

Identification of a functional ROI was more reliable for individuals than for groups. This was true despite the fact that the number of data available to perform individual localization were limited to one acquisition sequence lasting less than 6 min. Individually defined ROIs in atlas space tended to show more agreement in their location as well as in the proportion of voxels that were included in both ROIs. When applied to data from another localizer run, these ROIs also encompassed more strongly activated voxels than ROIs defined through group averaging.

Importantly, the individual ROIs were not always more reliable than all three group ROIs. However, neither the matched, liberal, nor conservative group ROIs consistently performed well, and which of the three performed best varied across analyses. This inconsistency coupled with the anatomic implausibility of the liberal and conservative group ROIs suggests that individually defined functional regions should generally be preferred.

There are several purposes for which defining a functional ROI at the group level is desirable. Most important is the case in which one's aim is to characterize "typical" functional neuroanatomy. This has been by far the most common use of functional neuroimaging to date. Group localization may be preferable for testing the behavior of a functional ROI if the region is small or if the criterial response is weak (Downing et al., 2001). Even when the functional ROI is highly responsive to the localizer task and is relatively large, group averaging may be necessary when a small number of data are available to perform the localization for each individual. Finally, localization at the individual level presents the possibility of failing to identify an ROI in some individuals (McGonigle et al., 2000). This is an important issue to consider since the assumption of random and independent sampling may be violated if these participants are excluded from the analysis.

Further issues

It is also important to consider how thresholds for the ROI should be set: should they be held constant across participants or varied until the size of the individual ROIs are within a preset range? In fact, earlier clinical research suggests that reliability of activated regions does increase when thresholds are independently set for each individual (Rombouts et al., 1998; Specht et al., 2003). Are such *ad hoc* adjustments of significance level acceptable? Or should data-driven adjustments to alpha levels, and consequently Type I and Type II error rates, be accepted as a necessary step in the region finding process? Because of these issues and the increased complexity of analyzing functional data at the individual level, it may be more appropriate to use a functional ROI defined through group averaging if the normalization procedure is able to adequately match functional regions across individuals (but see Crivello et al., 2002).

Some reservations about our results should be kept in mind. First, it would be desirable to investigate the sensitivity of these results to variations in data preprocessing. Our preprocessing protocol registers individual structural and functional data to standard stereotactic space through what Brett and colleagues (2002) refer to as a volume

matching procedure. As these authors point out, this procedure does little to align the sulcal anatomy of the individuals in the sample. It is possible that a sulcal pattern matching algorithm might differentially benefit statistical grouping. Second, the method used here combined motion correction and atlas normalization in one resampling step. If the data are resampled twice, once for motion correction and again for atlas normalization, additional high-frequency information may be lost and localization impaired. However, the results of Miki et al. (2000), described above, imply that spatial normalization has little effect on reliability under these circumstances. Finally, it would be desirable to extend these results to other regions with differing consistency, responsiveness, and selectivity to their relative localizer tasks. In particular, regions with less anatomic variability should be more amenable to statistical grouping.

It is important to note that MT+ as localized here showed a relatively weak or inconsistent response in some participants, but individual ROIs still performed better than statistically grouped ROIs. This is suggested by the lack of contiguity in the regions identified at the group level. One possibility is that the localizer task employed here was not ideal. It is of concern that the group and mean individual MT+ centers of mass did not fall within the coordinate range identified by Tootell et al. (1995). More problematic, however, is that no clear left MT+ ROI could be identified in two participants in either run. In fact, one of these participants did not show the predicted response to the localizer task at all in the first run. This may be the result of having an insufficient amount of power in one run of this particular localizer task to clearly identify MT+ and may explain the large intersubject variability in the location of MT+. On its own, however, this should not be surprising since the location of MT+ is known to be variable (Dumoulin et al., 2000) and dependent upon the localization stimulus used (Morrone et al., 2000). Finally, the anatomic restrictions we used for MT+ may be too broad to have isolated one region responsive to the motion stimuli we used in our localizer task (Culham et al., 2001). Tighter anatomic criteria and retinotopic mapping of visual areas allow for more precise localization of MT+, but require considerably more data and more human-intensive analyses (Huk et al., 2002; Huk and Heeger, 2000; Tootell et al., 1995).

Conclusions

It may be helpful to conceptualize the difference between the group and individual localization methods as originating from two opposing sources: (1) a greater number of functional data to identify the region in the group analysis and (2) the inclusion of intersubject as well as interrun variability in the error term used to compute the *z* statistic images for the group contrasts. While more functional data should benefit the reliability of group ROIs relative to individual ROIs, intersubject variability will reduce the reliability of group ROIs when applied to individual contrasts. Although a greater number of observations and the inclusion of intersubject variability are useful in making inferences from samples to populations, the prediction of individual behavior from similar sources will necessarily lead to error. The accuracy of prediction in this case is negatively related to the amount of between-subjects variability in the population and the degree to which an individual's known score deviates from the group mean. In other words, using the behavior of the group to predict an individual's behavior will lead to more error when the behavior is variable across individuals. This principle was demonstrated in our data, suggesting that these results are likely to hold whenever uncertainty due to individual differences in functional-anatomic variability outweigh the predictive benefits of averaging over a large number of observations.

Researchers utilizing functional localization do so to predict the location and extent of future task-related activity. To infer that a region defined through some measure will always perform this function, the accuracy and reliability of this measure must be high: current results must reasonably predict future results. We conclude that for samples typical of those used in functional imaging studies the best predictions were made by defining regions at the individual level and that the most practical method of doing so utilized spatial normalization procedures.

Acknowledgments

This research was supported by a grant from the James S. McDonnell Foundation at Washington University in St. Louis. The authors thank Margaret Sheridan for her essential role in the collection of the data reported here. Portions of these data were presented at the 2003 annual meeting of the Cognitive Neuroscience Society in New York City.

References

Andersson, J.L., Sundin, A., Valind, S., 1995. A method for coregistration of PET and MR brain images. J. Nucl. Med. 36, 1307–1315.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34, 149–159.

Brett, M., Johnstrude, I.S., Owen, A.M., 2002. The problem of functional localization in the human brain. Nat. Rev. Neurosci. 3, 243–249.

Cohen, J., Cohen, P., 1983. Applied multiple regression/correlation analysis for the behavioral sciences, second ed. Erlbaum, New Jersey.

Cohen, J.D., MacWhinney, B., Flatt, M., Provost, J., 1993. PsyScope: a new graphic interactive environment for designing psychology experiments. Behav. Res. Methods, Instrum. Comput. 25, 257–271.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215.

Crivello, F., Schormann, T., Tzourio-Mazoyer, N., Roland, P.E., Zilles, K., Mazoyer, B.M., 2002. Comparison of spatial normalization procedures and their impact on functional maps. Hum. Brain Mapp. 16, 228–250.

Culham, J., He, S., Dukelow, S., Verstraten, F.A.J., 2001. Visual motion and the human brain: what has neuroimaging told us? Acta Psychol. 107, 69–94.

Downing, P., Liu, J., Kanwisher, N., 2001. Testing cognitive models of visual attention with fMRI and MEG. Neuropsychologia 39, 1329–1342.

Dukelow, S.P., DeSouza, J.F.X., Culham, J.C., van den Berg, A.V., Menon, R.S., Vilis, T., 2001. Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. J. Neurophysiol. 86, 1991–2000.

Dumoulin, S.O., Bittar, R.G., Kabani, N.J., Baker, C.L., Le Goualher, G., Pike, G.B., Evans, A.C., 2000. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb. Cortex 10, 454–463.

Friston, K.J., Holmes, A.P., Worsely, K.P., 1999. How many subjects constitute a study? NeuroImage 10, 1–5.

Friston, K.J., Holmes, A.P., Worsely, K.P., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Grosbras, M.H., Lobel, E., Van de Moortele, P.F., LeBihan, D., Berthoz, A., 1999. An anatomical landmark for the supplementary eye fields in human revealed with functional magnetic resonance imaging. Cereb. Cortex 9, 705–711.

Huk, A.C., Dougherty, R.F., Heeger, D.J., 2002. Retinotopy and functional subdivision of human areas MT and MST. J. Neurosci. 22, 7195–7205.

Huk, A.C., Heeger, D.J., 2000. Task-related modulation of visual cortex. J. Neurophysiol. 83, 3525–3536.

Hunton, D.L., Meizin, F.M., Buckner, R.L., van Mier, H.I., Raichle, M.E., Petersen, S.E., 1996. An assessment of functional-anatomical variability in neuroimaging studies. Hum. Brain Mapp. 4, 122–139.

Kourtzi, Z., Bülthoff, Erb, Grodd. 2002. Object-selective responses in the human motion area MT/MST. Nat. Neurosci. 5, 17–18.

Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital complex. Science 293, 1506–1509.

Lancaster, J.L., Glass, T.G., Lakipalli, B.R., Downs, H., Mayberg, H., Fox, P.T., 1995. A modality-independent approach to spatial normalization of tomographic images of the human brain. Hum. Brain Mapp. 3, 209–223.

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach atlas labels for functional brain mapping. Hum. Brain Mapp. 10, 120–131.

Machielson, W.C.M., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Witter, M.P., 2000. fMRI of visual encoding: reproducibility of activation. Hum. Brain Mapp. 9, 156–164.

McGonigle, D.J., Howseman, A.M., Athwal, B.S., Friston, K.J., Frackowiak, R.S.J., Holmes, A.P., 2000. Variability in fMRI: an examination of intersession differences. NeuroImage 11, 708–734.

Michelon, P., Snyder, A.Z., Buckner, R.L., McAvoy, M., Zacks, J.M., 2003. Neural correlates of incongruous visual information: An event-related fMRI study. NeuroImage 19, 1612–1626.

Miki, A., Raz, J., van Erp, T.G.M., Liu, C.J., Haselgrove, J.C., Liu, G.T., 2000. Reproducibility of visual activation in functional MR imaging and effects of postprocessing. Am. J. Neuroradiol. 21, 910–915.

Morrone, M.C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., Burr, D.C., 2000. A cortical area that responds specifically to optic flow, revealed by fMRI. N. Neurosci. 3, 1322–1328.

Moser, E., Teichtmeister, C., Diemling, M., 1996. Reproducibility and postprocessing of gradient-echo functional MRI to improve localization of brain activity in the human visual cortex. Magn. Reson. Imaging 14, 567–579.

Nobre, A.C., Sebestyn, G.N., Gitelman, D.R., Mesulam, M.M., Frackowiak, R.S.J., Frith, C.D., 1997. Functional localization of the system for visuospatial attention using positron emission tomography. Brain 120, 515–533.

Ojemann, J.G., Akbudak, E., Snyder, A.Z., McKinstry, R.C., Raichle, M.E., Conturo, T.E., 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. NeuroImage 6, 156–167.

Paus, T., 1995. Location and function of the human frontal eye-field: A selective review. Neuropsychologia 34, 475–483.

Petersson, K.M., Nichols, T.E., Poline, J.B., Holmes, A.P., 1999. Statistical limitations in functional neuroimaging II. Signal detections and statistical inference. Philos. Trans. R. Soc. London Ser. B 354, 1261–1281.

Petit, L., Haxby, J.V., 1999. Functional anatomy of pursuit eye movements in humans as revealed by fMRI. J. Neurophysiol. 81, 463–471.

Rombouts, S.A.R.B., Barkhof, F., Hoogenraad, F.G.C., Sprenger, M., Scheltens, P., 1998. Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. Magn. Reson. Imaging 16, 105–113.

Rombouts, S.A.R.B., Barkhof, F., Hoogenraad, F.G.C., Sprenger, M., Valk, J., Scheltens, P., 1997. Test-retest analysis with functional MR of the activated area in the human visual cortex. Am. J. Neuroradiol. 18, 1317–1322.

Rosano, C., Krisky, C.M., Welling, J.S., Eddy, W.F., Luna, B., Thulborn, K.R., Sweeney, J.A., 2002. Pursuit and saccadic eye movement subregions in human frontal eye field: A high-resolution fMRI investigation. Cereb. Cortex 12, 107–115.

Smith, S.M., 2001. Preparing fMRI data for statistical analysis. inJezzard, P., Matthews, P.M., Smith, S.M. (Eds.), resonance imaging: an introduction to methods, Functional magnetic. Oxford, Univ. Press, Oxford, pp. 229–241.

Specht, K., Willmes, K., Shah, J., Jäncke, L., 2003. Assessment of reliability in functional imaging studies. J. Magn. Reson. Imaging 17, 463–471.

Talairach, J., Tournoux, P., 1988. Co-planar stereotactic atlas of the human brain. Theime, New York.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., Kanwisher, N., 2000. Response properties of the human fusiform face area. Cogn. Neuropsychol. 17, 257–279.

Tootell, R.B.H., Dale, A.M., Sereno, M.I., Malach, R., 1996. New images from human visual cortex. Trends Neurosci. 19, 481–489.

Tootell, R.B.H., Reppas, J.B., Kwong, K.J., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., Belliveau, J.W., 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J. Neurosci. 15, 3215–3230.

Van Essen, D.C., Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H., Drury, H.A., 2001. An integrated software system for surface-based analyses of cerebral cortex. J. Am. Med. Inform. Assoc. 41, 1359–1378.

Van Essen, D.C., Drury, H.A., Harwell, J., Hanlon, D. 2002. CARET: Computerized Anatomical Reconstruction and Editing Toolkit [Computer software and manual]. Retrieved from http://brainmap.wustl.edu/caret.

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., Buckner, R.L., Raichle, M.E., 2001. Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 4, 651–655.

Zacks, J.M., Ollinger, J.M., Sheridan, M.A., Tversky, B., 2002. A parametric study of mental spatial transformations of bodies. NeuroImage 16, 857–872.

Zacks, J.M., Vettel, J.M., Michelon, P., in press. Imagined viewer and object rotations dissociated with event-related fMRI. J. Cogn. Neurosci.