Koku
Geruch

NeuroImage

Odore
Geur
رائحة
Smell
냄새

Odeur
गंध

气味

Human olfactory bulb

OT

1. olfactory nerve layer
2. glomerular layer
3. external plexiform layer
4. mitral cell layer
5. internal plexiform layer
6. granule cell layer
OT = olfactory tract

匂い

Обоняние
ריח

Olor
Μυρωδιά

Available online at www.sciencedirect.com
SciVerse ScienceDirect

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)
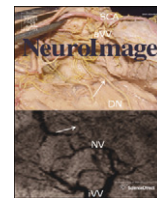
Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses

Alfonso Nieto-Castañón *, Evelina Fedorenko *

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Building 46, Room 3037G, Cambridge, MA 02139, USA

ABSTRACT

One important goal of cognitive neuroscience is to discover and explain properties common to all human brains. The traditional solution for comparing functional activations across brains in fMRI is to align each individual brain to a template brain in a Cartesian coordinate system (e.g., the Montreal Neurological Institute template). However, inter-individual anatomical variability leads to decreases in *sensitivity* (ability to detect a significant activation when it is present) and *functional resolution* (ability to discriminate spatially adjacent but functionally different neural responses) in group analyses. Subject-specific functional localizers have been previously argued to increase the sensitivity and functional resolution of fMRI analyses in the presence of inter-subject variability in the locations of functional activations (e.g., Brett et al., 2002; Fedorenko and Kanwisher, 2009, 2011; Fedorenko et al., 2010; Kanwisher et al., 1997; Saxe et al., 2006). In the current paper we quantify this dependence of sensitivity and functional resolution on functional variability across subjects in order to illustrate the highly detrimental effects of this variability on traditional group analyses. We show that analyses that use subject-specific functional localizers usually outperform traditional group-based methods in both sensitivity and functional resolution, even when the same total amount of data is used for each analysis. We further discuss how the subject-specific functional localization approach, which has traditionally only been considered in the context of ROI-based analyses, can be extended to whole-brain voxel-based analyses. We conclude that subject-specific functional localizers are particularly well suited for investigating questions of functional specialization in the brain. An SPM toolbox that can perform all of the analyses described in this paper is publicly available, and the analyses can be applied retroactively to any dataset, provided that multiple runs were acquired per subject, even if no explicit "localizer" task was included.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Cognitive neuroscience strives for generality: one of the main goals is to discover and explain properties common to all human brains, not just to a particular individual or set of individuals.[1] In functional magnetic resonance imaging (fMRI) studies, for example, it is important to determine whether a particular activation pattern is consistent across subjects. The traditional solution for comparing functional activations across different brains in fMRI is to align each individual brain to a template brain in a Cartesian coordinate system

(e.g., MNI coordinate space; Devlin and Poldrack, 2007; Poldrack et al., 2011) and then to examine the effects – across the subjects in the set – in each unit of analysis, i.e., a voxel. Random-effects and linear mixed model group analyses, which have become the standard since Holmes and Friston (1998), further allow the investigator to extend conclusions from the particular set of subjects to the larger population.

Because of inter-subject anatomical variability, however, alignment of functional activations across individual brains is not perfect (e.g., Miller et al., 2002; Wohlschläger et al., 2005). Two sources of anatomical variability plausibly contribute to this poor alignment. First, brains vary in their folding patterns (e.g., Duvernoy, 1991; Geschwind and Levitsky, 1968; Juch et al., 2005; Ono et al., 1990; Tomaiuolo et al., 1999). And second, brains vary in terms of the locations of the cytoarchitectonic zones – which plausibly correspond to function (e.g., Iwamura et al., 1983; Matelli et al., 1991; Rozzi et al., 2008) – relative to the sulci and gyri (e.g., Amunts et al., 1999; Brodmann, 1909; Fischl et al., 2008; Rajkowska and Goldman-Rakic, 1995a,b; Zilles et al., 1997). Because of this second kind of variability, even

* Corresponding authors. Fax: +1 617 258 8654.
E-mail addresses: alfnie@gmail.com (A. Nieto-Castañón), evelina9@mit.edu (E. Fedorenko).

[1] Another important goal of cognitive neuroscience is to explain the *inter-individual differences* in brain anatomy and function, and to relate those differences to behavior on the one hand, and genetic make-up, on the other. Although the discussion of the methods for investigating questions related to individual differences is beyond the scope of this paper, the methods advocated here are well suited for looking at inter-individual variability in functional activations.

the more advanced normalization methods, which aim to align sulci and gyri across individual brains (e.g., Fischl et al., 1999), are unlikely to lead to perfect alignment of functional activations (e.g., Frost and Goebel, 2012). Furthermore, to the extent that there exists functional heterogeneity *within* cytoarchitectonic areas (e.g., self-organizing neural spatial representations such as the ocular dominance columns in V1), there is no reason to expect these functional patterns to line up across individual brains due to their very fine spatial scale. As a result, methods that rely on inter-subject overlap of functional activations in stereotaxic space to make population-level inferences are likely to a) miss some effects that are present because functionally equivalent regions will not be aligned across subjects, as well as b) detect spurious activation overlap between conditions because spatially distinct activations will be blurred together in group analyses.

An alternative method that circumvents the need to align individual brains in a common stereotaxic space and yet enables the use of inferential statistics about the human brain in general (rather than one individual or a small set of individuals) emerged in the mid 1990s. In this method, regions of interest (ROIs) are defined *functionally* in each individual brain using a contrast targeting the cognitive function of interest (a "localizer" contrast).[2] Once a region is identified in each subject in this way, inferences can be made about the responses of this region to new experimental conditions by aggregating the responses across the voxels within these subject-specific areas (bottom panel of Fig. 1). This method was first applied to cortical regions engaged in visual processing (e.g., the motion-sensitive area MT and the object-selective lateral occipital complex—Tootell et al., 1995a, 1995b; Malach et al., 1995; retinotopic cortex—Tootell et al., 1998). For example, visual area MT is commonly defined by comparing the brain's response to radially moving dots relative to stationary dot arrays. Indeed, it has been the consensus in the vision fMRI community for a decade that each individual subject must be retinotopically mapped so that the experimental data can be analyzed and reported separately for each specific visual region (V1, V2, V3, etc.). Others subsequently applied this subject-specific functional ROI method to investigations of high-level visual processing. For example, an area selectively engaged in processing faces was discovered in the fusiform gyrus (the fusiform face area, FFA), using a contrast between faces and objects (Kanwisher et al., 1997; see also McCarthy et al., 1997). Several other high-level visual areas were discovered and characterized, using this approach, over the following several years, including the parahippocampal place area, PPA (Epstein and Kanwisher, 1998) and the extrastriate body area, EBA (Downing et al., 2001). Furthermore, this approach has now been successfully applied to even higher-level cognitive domains, such as theory of mind (e.g., Saxe and Kanwisher, 2003; Saxe and Powell, 2006), numerical processing (e.g., Pinel et al., 2007; Shuman and Kanwisher, 2004), language (e.g., Fedorenko et al., 2010; Hickok et al., 2009; January et al., 2009), and executive functions (e.g., Derrfuss et al., 2012).

Although the use of functional localizers is now quite widespread and although there have been several empirical demonstrations of the advantages of the subject-specific functional localization method compared to group-based methods (e.g., Brett et al., 2002; Fedorenko and Kanwisher, 2009, 2011; Fedorenko et al., 2010, 2012; Saxe et al., 2006), the method remains controversial (e.g., see the debate between Friston et al., 2006, and Saxe et al., 2006, *Neuroimage*, for an extensive discussion of the advantages and disadvantages of the method). The goal of the current paper is two-fold. *First*, we will formally demonstrate

the relationship between inter-subject variability in the loci of functional activations and the sensitivity and functional resolution of the analyses, and show that subject-specific functionally-defined ROIs (fROIs)[3] surpass subject-independent ROIs, with respect to both sensitivity and functional resolution, for a wide range of variability values. Importantly, this increase in sensitivity and functional resolution from fROI methods holds even when the same total amount of data is used for each analysis (i.e., the localizer plus the experimental runs in the fROI analysis, and all experimental runs in the traditional group analysis). Furthermore, we argue that because the use of functional localizers has been typically framed in the context of ROI analyses, their advantages have often been interpreted too narrowly, in terms of the relative merits of ROI- vs. whole-brain voxel-based analyses. Therefore, the *second* goal of the paper is to extrapolate the results from *ROI-based analyses* to *voxel-based analyses*. In particular, we will show that standard multi-subject voxel-based analyses, like ROI-based analyses, improve in sensitivity and functional resolution, and show a reduction in bias, when subject-specific localizer contrasts are used.

Before proceeding to the main part of the paper, one important issue deserves discussion. In particular, as we note above, the functional localization approach circumvents the need to align individual brains in a common stereotaxic space: all of the analyses can be performed in the subject's native anatomical space. However, individual-subject analyses can also be (and often are) performed in the common space. What are the relative advantages of the native vs. common space? Intuitively, it seems that functional localization should be performed using the subject's own anatomy (e.g., Swallow et al., 2003). This analysis method results in the least amount of data distortion, and the locations of the functional activations can be more clearly related to the macroanatomic landmarks (e.g., Grosbras et al., 1999). However, the use of the common space has some benefits as well: the results can be more easily related to other studies that rely on the use of the common space. This is especially important in the fields where traditional group methods still dominate the research landscape, like language or social cognition (cf. the field of vision research). Furthermore, common space is required in cases where the definition of individual fROIs is constrained by a group-level representation of the functional data (e.g., Fedorenko et al., 2010; Thirion et al., 2007). Although for the purposes of method comparison, all of the subject-specific analyses discussed here – both ROI-based and whole-brain voxel-based – rely on the use of the common space, fROI analyses performed in the native subject space (using subject-specific anatomical ROIs) would benefit from increased sensitivity and functional resolution in much the same way as those conducted in the common space. Reassuringly, a few previous studies that have examined the effects of spatial normalization on the individual-subjects' activations suggest that spatial normalization as in of itself does not have much of a detrimental effect (e.g., Miki et al., 2000; Swallow et al., 2003).

---

[2] In some cases (e.g., when working with certain subcortical structures or with cortical areas that have relatively good alignment with the cortical folds, like V1 (Hinds et al., 2009)), regions of interest can be defined *anatomically* in each individual brain. However, for most cortical regions anatomical definition is problematic. As a result, although anatomical ROIs defined in individual subjects have been shown to be superior to those defined in the common space (Nieto-Castañón et al., 2003; see also Grosbras et al., 1999), combining subject-specific anatomical ROIs with functional localization has a better chance of picking out the corresponding functional regions across subjects.

[3] There are several ways to define fROIs in individual brains. The traditional way involves examining each individual subject's activation map for the localizer contrast and selecting a set of active voxels using macroanatomic landmarks as guidelines. Fedorenko et al. (2010; see also Julian et al., 2012) have recently argued that a more objective way of defining ROIs is desirable in order to (a) avoid the subjectivity inherent in the traditional procedure, and (b) facilitate the standardization of ROI definition procedures across subjects, studies, and labs. A solution Fedorenko et al. (2010) proposed was to use spatial constraints ("parcels") derived from a group-level representation of functional activation data (e.g., a probabilistic overlap map or a random-effects map) to constrain the selection of subject-specific voxels. In particular, the group-level parcels are intersected with individual subjects' thresholded activation maps for the localizer contrast, and voxels that fall within the boundaries of each relevant parcel are used for subject-specific ROI definition. A related alternative discussed in Fedorenko et al. (2012) involves intersecting each individual subject's activation map for the localizer contrast with *anatomical* parcels from standardized atlases (e.g., Duvernoy, 1991; Maldjian et al., 2003; Tzourio-Mazoyer et al., 2002; Eickhoff et al., 2005; see Thirion et al., 2007, 2010a,b, for other approaches). Although these recent, more objective and automated, fROI definition methods may be adopted in the future, the arguments presented here apply to *all* subject-specific fROI methods, regardless of the details of the ROI definition procedure.
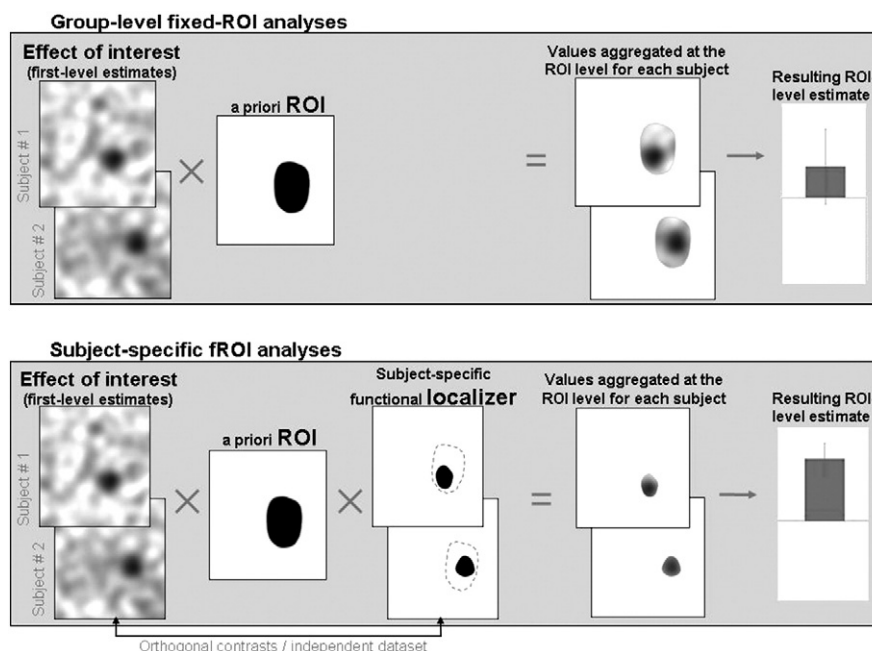
Fig. 1. A schematic illustration of ROI-based analyses. Top panel: an a priori ROI is intersected with each subject's activation map for the effect of interest. ROI-level measures are estimated by aggregating the BOLD data (or single-subject effect estimates) across all of the voxels within the ROI. Bottom panel: an a priori ROI is intersected with each subject's functional localizer mask. ROI-level measures are estimated by aggregating the BOLD data across all of the voxels within the resulting subject-specific areas. Note that the data used to derive the values that are aggregated at the ROI level for each subject must come from data left out of the localizer, or from a different, orthogonal contrast (Vul and Kanwisher, 2010).

Finally, we clarify some terminology that will be used throughout the paper. We classify multi-subject analysis methods as *group-level* vs. *subject-specific*. In the former, only group-level activation maps are used in the analyses (e.g., voxel-based stereotaxically registered group analyses, or ROI analyses that utilize fixed subject-independent ROIs), while in the latter, subject-specific functional activation maps are used at some stage in the analyses. We use the term *localizer contrast* to refer to a contrast between conditions/sets of conditions that is used for constraining the units of analysis, i.e., by limiting the analyses within each subject to a subset of voxels that show a particular functional characteristic, as indicated by supra-threshold effects in the localizer contrast. Unless otherwise noted, we use the term *sensitivity* to refer to the sensitivity or power (true-positives rate) of multi-subject analyses, characterizing the ability of a given analysis procedure to correctly infer a functional response when one is present in the population (see Appendix A for details). We use the term *functional resolution* to refer to the ability of a given analysis procedure to discriminate two spatially adjacent but functionally different neural responses (see Appendix E for details).

Methods and results

To evaluate the sensitivity and functional resolution of different analysis procedures, we considered the following general scenario. A researcher is interested in performing an inference on a primary contrast of interest *A* (e.g., functional response to high-pitch auditory stimuli compared to some low-level baseline condition, such as silence). As a secondary goal, the researcher would like to know whether a second and orthogonal contrast of interest *B* (e.g., functional response to low-pitch auditory stimuli compared to a low-level baseline) elicits responses in the same or different brain areas. Both of these questions exemplify typical inferences made from neuroimaging data, with the second question being of particular importance when investigating questions of functional specialization in the brain. We here evaluate the impact of anatomical variability on the ability of the researcher to correctly answer both of

these questions. We use two complementary measures: sensitivity (the ability of a given analysis procedure to correctly answer the first question; e.g., to infer a functional response to high-pitch stimuli when one is present in the population), and functional resolution (the ability of a given analysis procedure to correctly answer the second question; e.g., to infer a selective neural response to high-pitch but not low-pitch stimuli when the functional responses to these two stimuli are located in adjacent but non-overlapping regions[4]).

We begin with a demonstration of the impact of inter-subject variability in the loci of activation on group-level fixed-ROI analyses ('Standard group-level fixed-ROI analyses' section). We then discuss subject-specific fROI analyses and show how they circumvent the problem of inter-subject variability in the loci of activation, achieving higher sensitivity and functional resolution and a reduction in bias, compared to the fixed-ROI method ('Subject-specific functional ROI analyses' section). In the section 'Standard voxel-based analyses', we demonstrate the impact of inter-subject variability on voxel-based analyses, for unsmoothed and smoothed data. We show how voxel-based analyses that use smoothing can be considered as a general case of fixed-ROI analyses. The section 'Subject-specific localizers in the context of whole-brain voxel-based analyses' then discusses the application and advantages of subject-specific functional localizers in the context of voxel-based analyses. Finally, the section 'Simulation examples' presents a series of simulations that illustrate the points made in the section 'Standard group-level fixed-ROI analyses' up to the section

---

[4] Note that the second question in this example is typically addressed by a conjunction analysis, inferring a non-selective response to stimulus *A* if, within the same analysis unit (ROI or voxel), both effects *A* and *B* are found significant at a given threshold. Inter-subject variability in the loci of activation introduces the potential for false positives in this conjunction analysis, beyond those related to the threshold level, when two nearby areas respond selectively to stimuli *A* and *B*, respectively (where the researcher might incorrectly conclude that a given analysis unit responds to both stimuli).

'Subject-specific localizers in the context of whole-brain voxel-based analyses', and the section 'Some practical considerations' discusses some practical issues important to consider when performing subject-specific analyses.

Standard group-level fixed-ROI analyses

Here, we consider the case of group-level ROI analyses[5] that use a fixed (subject-independent) ROI across all subjects (top panel of Fig. 1). This fixed ROI can be defined functionally (using a group-level map for a localizer contrast, or a set of coordinates from another study) or anatomically (using macroanatomic landmarks or anatomical atlases; e.g., Duvernoy, 1991; Maldjian et al., 2003; Tzourio-Mazoyer et al., 2002; Eickhoff et al., 2005). First, we model the effect of inter-subject variability in the loci of activation on the sensitivity and functional resolution of group-level fixed-ROI analyses, and we show how the sizes of individual activations relative to the size of the ROI mediate this effect. Then we investigate the detrimental effect of inter-subject variability in the context of a common practice to accommodate this variability by using ROI definitions that fully encompass the expected extent of variability in activation loci (e.g., Mikl et al., 2008; White et al., 2001). We show that this practice only partially compensates for the loss of sensitivity, even when optimally placed and sized ROIs are used. Last, we discuss additional bias and functional resolution issues in group-level fixed-ROI analyses.

In Appendices A and E we derive the general expressions for the sensitivity and functional resolution of group-level fixed-ROI analyses. Inter-subject variability in the loci of activation has a detrimental effect on the sensitivity and functional resolution of group-level fixed-ROI analyses.[6] This detrimental effect can be characterized by the distribution of *partial-coverage* values across-subjects, i.e., the sizes of the true loci of activation relative to the size of the ROI. Partial coverage can be expressed as a percentage of the total number of voxels in the ROI (e.g., if a subject activates only 30 voxels of the 300 total voxels in the ROI, the partial coverage value for this subject will be 10%). Both (i) low average partial-coverage values (i.e., activations within the ROI that are small with respect to the total size of the ROI), as well as (ii) high variability in partial-coverage values (i.e., differences in the size of activations within the ROI across subjects), detrimentally affect the sensitivity of the group-level fixed-ROI analyses. The effect of different distributional properties of partial-coverage values on the sensitivity and functional resolution of group-level analyses is illustrated in Fig. 2 (top). Sensitivity and functional resolution values are computed using an ROI size of 1 resel (resolution element; Friston et al., 1994), and average within- and between-subject variability estimates from Desmond and Glover (2002).[7] As illustrated in the left panel of Fig. 2 (top), sensitivity of fixed-ROI analyses decreases with decreasing partial-coverage values. In practical terms

this means that if the true locus of activation is small compared to the ROI size, fixed-ROI analyses can fail to find a significant effect that is present within the ROI (because the effect of interest is washed out by the large number of voxels that do not show the effect). Similarly, as illustrated in the right panel of Fig. 2 (top), functional resolution decreases rapidly with decreasing partial-coverage values. In practical terms this means that if an ROI shows a response to two conditions, *A* and *B*, we will not know whether the ROI, or areas within it, respond to both *A* and *B*, or whether the ROI instead contains areas that respond selectively to each of these two conditions (e.g., an area that responds to *A* but not *B*, and a different area that responds to *B* but not *A*), unless the ROI is of similar size or smaller than the true loci of activation.

Although smaller ROI sizes can lead to higher functional resolution, the relationship between ROI size and sensitivity is more complex. In particular, a common solution for accommodating some level of inter-subject variability in the loci of activation in ROI analyses is simply to define an ROI that encompasses the expected variability across subjects (e.g., Mikl et al., 2008; White et al., 2001). Fig. 3 (left) illustrates the sensitivity of group-level analyses as a function of inter-subject variability in the loci of activation, assuming fixed ROIs centered in the optimal location (the average locus of activations across subjects), and considering different possible ROI sizes. The values of the within- and between-subject sensitivities used in this plot represent the higher range of sensitivity values from Desmond and Glover (2002). This plot illustrates that maximal sensitivity can be obtained by using ROI sizes that approximately match the overall extent of activation across subjects (consistent with the matched filter theorem; Turin, 1960). The dotted line represents the optimal sensitivity that can be achieved in group-level fixed-ROI analyses (by placing an optimally-sized ROI at the optimal location) across different extents of inter-subject variability. Critically, however, even if a researcher is able to a priori define optimally-sized and optimally-located fixed ROIs, the sensitivity of group-level analyses will still be detrimentally affected by inter-subject variability (the dotted line decreases dramatically with increasing inter-subject variability values). In other words, in the presence of inter-subject variability, increasing the ROI size to encompass the extent of observed or expected variability will be at best a partial solution; the analyses will still suffer from decreased sensitivity compared to that achievable in the presence of little/no inter-subject variability in the loci of activation, or compared to subject-specific analyses, as we will show in the section 'Subject-specific functional ROI analyses'.

In addition to these sensitivity and functional resolution limitations, it is also worth noting that by aggregating across both "active" and "inactive" voxels within the ROI in each subject, the effect sizes will typically be underestimated compared to those within the true loci of activation (by a factor proportional to the partial-coverage average; see Appendix A; see Saxe et al., 2006; Fedorenko et al., 2012, for empirical support). This adds an important bias to analyses that require accurate effect size estimation (e.g., power analyses), and it also adds an additional source of variance to other secondary analyses based on these effect size estimates (e.g., predictive analyses of behavioral measures).[8] All of these limitations are often used to argue against ROI-based analyses and in favor of voxel-based analyses (to avoid aggregating across voxels with non-homogeneous responses;

---

[5] We consider here ROI analyses that take the average time-series within an ROI as a summary measure of the ROI signal. Although some alternative approaches have been argued to better deal with functionally heterogeneous ROIs (e.g., extracting the first eigenvariate from a singular value decomposition of the within-ROI covariance, Friston et al., 2006), the analysis of these alternative methods goes beyond the scope of this manuscript.

[6] Although these analyses consider the specific contribution of inter-subject variability in the *loci* of activation, they also take into account the presence of other sources of between-subject variability (e.g., variability in the magnitude of the neural response across subjects; e.g., Aguirre and Detre, 2012).

[7] Because the estimates in Desmond and Glover (2002) do not attempt to discriminate different sources of between-subject variability (i.e., those associated with variability in the loci of activation vs. those associated with variability in effect size magnitude due to other sources), the between-subject variability estimates used in our analyses are likely overestimating the true amount of variability in effect size magnitude from other sources. In the absence of accurate measures of the relative contribution of different sources to the total between-subject variability, we chose this approach so that our conclusions would err on the conservative side (plausibly underestimating rather than overestimating the relative contribution of inter-subject variability in the loci of activation).

[8] Note, nevertheless, that the concept of bias is always relative to a particular model of neural activation, and we are considering here a simplified model of activation that considers a localized response within the ROI, with unknown strength, position, size, and shape, which may vary across subjects (see Appendix A for details). Biases are defined as mismatches between the reported effect sizes and the mean strength of the neural responses in this particular model of activation, and it is only with respect to this model that fixed-ROI analysis effect size estimates can be considered biased. In contrast, fixed-ROI analyses clearly would not be biased with respect to a simpler model where any ROI inhomogeneities are disregarded and the ROI behavior is simply modeled by its average effect.
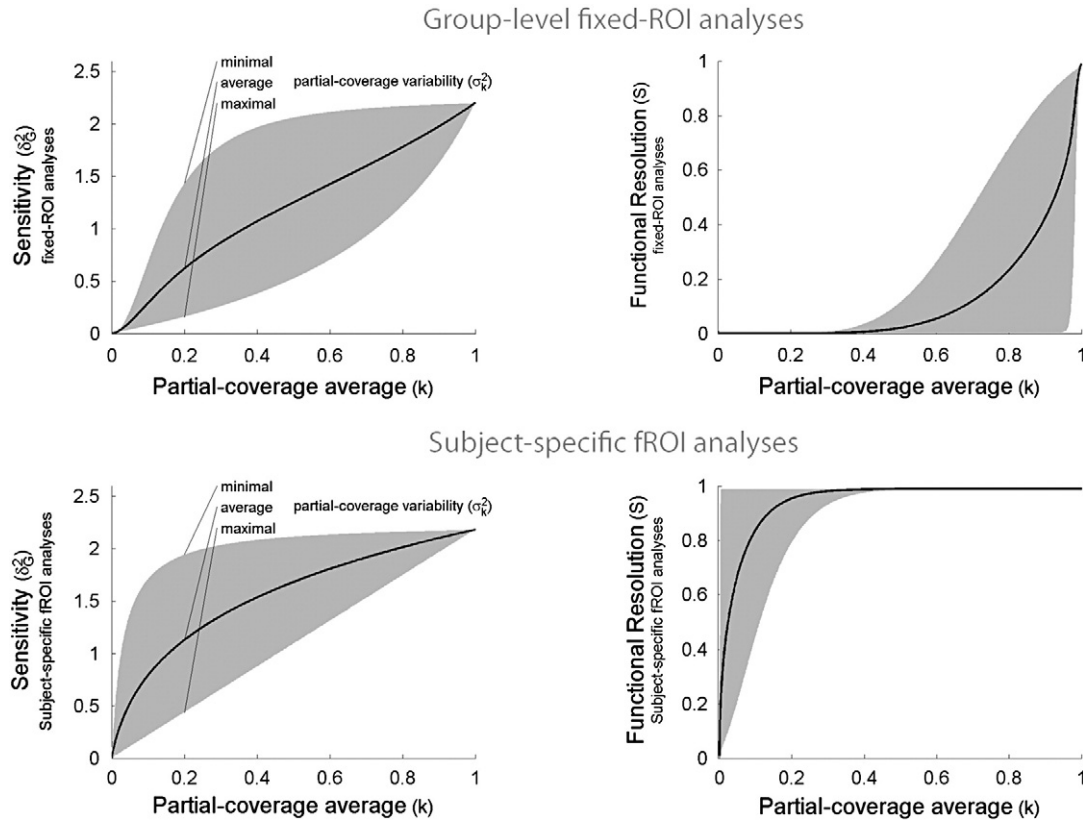
Group-level fixed-ROI analyses

Subject-specific fROI analyses

Fig. 2. Sensitivity (left) and functional resolution (right) of group-level fixed-ROI analyses (top) and subject-specific fROI analyses (bottom) as a function of partial-coverage values. Top panel: the sensitivity and functional resolution of group-level fixed-ROI analyses are detrimentally affected by inter-subject variability in the loci of activation. This detrimental effect can be characterized as a function of the average and variability of partial-coverage values across subjects. The sensitivity (y-axis, left plot) and functional resolution (y-axis, right plot) of group-level fixed-ROI analyses is plotted for different levels of partial-coverage average values (x-axis), and for all possible partial-coverage variability values (gray area). (Partial-coverage values represent the proportion of activated voxels within an ROI relative to the total number of voxels in the ROI.). Bottom panel: the sensitivity (y-axis, left plot) and functional resolution (y-axis, right plot) of subject-specific fROI analyses are plotted for different levels of partial-coverage average values (x-axis), and for all sensible partial-coverage variability values (gray area). Compared to the fixed-ROI case (top panel), both sensitivity and functional resolution show a reduced detrimental effect of lower partial-coverage values or higher partial-coverage variability.

e.g., Friston et al., 2006). The section 'Subject-specific functional ROI analyses' will show how functional localizers offer an alternative approach for increasing the sensitivity and functional resolution of these analyses, and the sections 'Standard voxel-based analyses' and 'Subject-specific localizers in the context of whole-brain voxel-based analyses' will further show how functional localizers can be incorporated

Group-level fixed ROI analyses

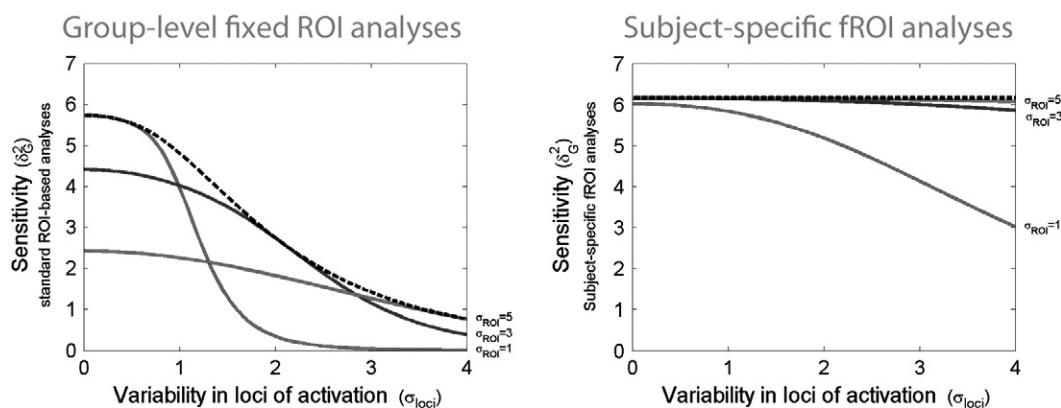## Subject-specific fROI analyses

Fig. 3. Sensitivity of group-level fixed-ROI analyses (left) and subject-specific fROI analyses (right) as a function of spatial variability in the loci of activations and as a function of the ROI size. Left: the sensitivity of multi-subject fixed-ROI analyses (y-axis) is plotted for different levels of inter-subject variability in the loci of activation (x-axis) for three different ROI sizes (solid lines). ROI sizes that approximately encompass the extent of activation across subjects (dotted line) are optimal, yet the sensitivity of such ROIs still decreases with increasing inter-subject variability in the loci of activation. Theoretical sensitivity values estimated for a spherical Gaussian-distributed activation with size parameter σ_{act} (fixed to a value of 1, representing approximately an activation extent of 12 voxels FWHM assuming 1 resel = 125 voxels) and with the loci of this activation varying randomly between subjects following a Gaussian distribution in center positions (with σ_{loci} varying from 0 to 4 characterizing the inter-subject variability in the loci of activation). The ROI is similarly characterized as encompassing an isotropic Gaussian sphere with size parameter σ_{ROI} (varying from 1 to 5). Right: the sensitivity of multi-subject subject-specific fROI analyses (y-axis) is plotted for different levels of inter-subject variability in the loci of activation (x-axis) for three different ROI sizes (solid lines). ROI sizes that minimally encompass the extent of activation across subjects are optimal (cf. matched filter theorem). In addition, the sensitivity of optimally-sized ROIs (dotted line) is not detrimentally affected by increasing inter-subject variability in the loci of activation.

into voxel-based analyses, increasing their sensitivity and functional resolution as well.

Subject-specific functional ROI analyses

The application of subject-specific functional localizers in the context of ROI analyses considers, instead of all of the voxels within a fixed ROI volume, only those voxels within the a priori ROI that are activated in each subject by the localizer contrast.[9] In this context, the a priori ROI serves as a spatial constraint on the location of the fROI in each subject, and the individual subjects' fROIs are defined by intersecting this a priori ROI and the subject-specific localizer mask (see Fig. 1, bottom). As in the previous section, we first model the effect of inter-subject variability in the loci of activation on sensitivity and functional resolution. We then show that when the a priori ROI fully encompasses the extent of inter-subject variability in activation loci, this method compensates for the detrimental effect of inter-subject variability on sensitivity, and we discuss how this method additionally addresses the functional resolution and bias limitations present in the fixed-ROI approach. Last, we discuss the practical conditions for which this method can be expected to outperform the group-level fixed-ROI method described in the previous section in terms of the sensitivity of multi-subject analyses.

In Appendices B and E we derive the general expressions for the sensitivity and functional resolution of subject-specific fROI analyses. Fig. 2 (bottom) illustrates the effect of different distributional properties of partial-coverage values on sensitivity and functional resolution. The ROI size and within- and between-subject variability values exemplified here are the same as those used in Fig. 2 (top), and we assumed an 80% power for the localizer contrast when using an FDR-corrected p<.05 false positive threshold (a relatively powerful localizer). This figure illustrates two key findings. First, in contrast to the fixed-ROI case, the functional resolution of subject-specific fROI analyses remains high over a wide range of partial-coverage values, which is a natural consequence of aggregating only across "active", or "responsive", voxels within the ROI. In practical terms this means that subject-specific fROI analyses can effectively identify whether areas within the ROI show a joint or selective response to two conditions *A* and *B* (e.g., it can identify whether there are areas within an ROI that respond to *A* but not *B* by using a localizer contrast *A* and examining the response to *B* across the suprathreshold voxels within this ROI, or vice versa). And second, unlike in the fixed-ROI case, this additional functional resolution does not come at the cost of reduced sensitivity. On the contrary, in this example the sensitivity of subject-specific fROI analyses is the same as or higher than that of the fixed-ROI analyses over a wide range of partial-coverage values. In the remainder of the section, we will investigate a) the effect of ROI size (i.e., of the size of the spatial constraint used to select the relevant subject-specific voxels) on subject-specific fROI analyses compared to fixed-ROI analyses; and b) whether the higher sensitivity and functional resolution values observed in the current example generalize to other conditions.

Fig. 3 (right) illustrates sensitivity at different levels of inter-subject variability in the loci of activation, when considering the same parameter values as in the example of group-level fixed ROIs (Fig. 3, left), and assuming an 80% power level for the localizer contrast. Similar to fixed-ROI analyses, optimally-sized ROIs in subject-specific fROI analyses encompass (at least) the overall extent of activation across subjects. In contrast to fixed-ROI analyses, however, the optimal sensitivity of subject-specific fROI analyses does not decrease with increasing extents of inter-subject variability in the loci of activation (dotted line remains approximately constant for the entire range of variability values), highlighting that this methodology compensates for the variability in the loci of activation without incurring an associated loss of sensitivity that characterizes the fixed-ROI methodology.

In addition to the increased robustness to inter-subject variability in the loci of activation, it is also worth noting that, by aggregating only across active voxels within the ROI in each subject, the effect sizes will typically be more accurately estimated than in a fixed-ROI methodology (see Appendix B). The simulations in the section 'Simulation examples' will illustrate the theoretical advantages of the subject-specific ROI methodology discussed in this section, showing how spatially non-homogeneous responses within the ROI can be differentiated, and how non-overlapping responses to different conditions can be correctly identified in these analyses.

It could be argued, however, that the advantages of using functional localizers are to be weighted against two potential disadvantages. First, using separate (orthogonal) contrasts to define the ROIs and to estimate the strength of the BOLD responses results in decreased power because some portion of the data is not being used to estimate the strength of the response (Neyman–Pearson lemma). Second, the effectiveness of the subject-specific approach relies on being able to (relatively accurately) estimate the location of the responses using a localizer contrast. Depending on the power of the localizer contrast, we might not be able to obtain accurate representations of the subject-specific locations of activation, which would adversely affect the sensitivity of the multi-subject analyses. In other words, one would wish to use a large portion of the data for the localizer contrast in order to obtain accurate representations of the location of activation for each subject. Yet one would also wish to use a large portion of the data for the estimation of the strength of the responses in order to maximize the within-subject sensitivity (minimize the estimation error impacting the sensitivity of multi-subject analyses). This issue could raise concerns that the sensitivity and functional resolution gains exemplified in Fig. 3 (right) might not generalize to other conditions. However, we show in Appendix B (Eq. (B2)) that regardless of how the data are split (including the use of cross-validation, discussed in the section 'Some practical considerations' below) and for any arbitrary distribution of partial-coverage values, it is *always* possible to find a minimal single-subject power that will result in improved sensitivity and functional resolution of the multi-subject analyses when using the subject-specific fROI method compared to using the group-level fixed-ROI method. In other words, for the same amount of total functional data, and no matter how much of the data we use for defining the localizer contrast, if the first-level design is sufficiently high-powered (e.g., the total amount of functional data acquired for each subject is large enough) the sensitivity of the resulting subject-specific fROI analyses always surpasses that of a fixed-ROI methodology applied to the entire dataset. Although this result is theoretically interesting, the question remains of what exactly is a "sufficiently" high-powered first-level design. We address this question next, by explicitly comparing the sensitivity and functional resolution of the fixed-ROI and subject-specific methods for the same amount of experimental data.

Following the description of the general experimental setup at the beginning of the section 'Methods and results', we consider and experiment with two contrasts/conditions *A* and *B*, acquired across multiple

---

[9] The general application of the subject-specific fROI methodology includes both (i) cases where the response is estimated for the same conditions that are used to define the ROIs (as considered here; for example, we can use the faces>objects contrast in the even runs to define the ROIs and then extract the response to these two conditions from the odd runs in order to estimate the magnitude of the responses; see the section 'Some practical considerations' for a discussion of how cross-validation can be used in such cases to maximize the power of these analyses), and (ii) cases where the response is estimated for a new condition or set of conditions (as discussed in the context of functional resolution, later in this section); for example, we can use the faces>objects contrast to define face-sensitive regions and then examine the response of these regions to body parts). In either case, either independent subsets of the data or orthogonal contrasts should be used for defining the ROIs vs. for estimating the responses (e.g., Kriegeskorte et al., 2009). Examples of these different applications will be shown in the simulations in the section 'Simulation examples'.
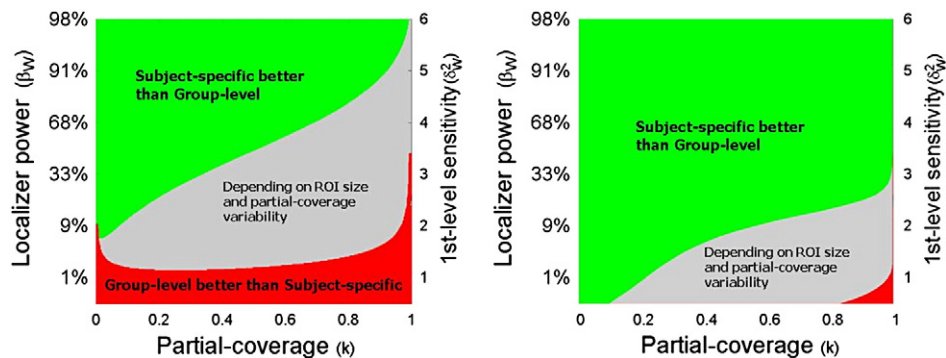
Fig. 4. A comparison of multi-subject sensitivity (left) and functional resolution (right) in subject-specific fROI analyses vs. group-level fixed-ROI analyses. The plot displays the relative advantages of the subject-specific fROI method compared to the group-level fixed-ROI method, for a range of single-subject design sensitivities (y-axis) and average partial-coverage values (x-axis). Both methods are constrained to use the same limited amount of functional data. The "quality" of the data is represented by the 1st-level design sensitivity (y-axis, where higher values correspond to stronger effects, less noise, and/or longer acquisitions). For any given partial-coverage value, the subject-specific fROI methodology will result in increased multi-subject sensitivity and functional resolution compared to the fixed-ROI methodology as long as the single-subject design is sufficiently high-powered (green area).

runs for each subject (6 runs), and spanning the entire set of functional acquisitions. The fixed-ROI method will estimate the average effect of the *A* condition within the ROI using the entire dataset for each subject (6 runs), while the subject-specific fROI method will estimate the average effects within the fROI using a cross-validation approach (it will use 5 runs to localize the key region in each subject individually, and the remaining run to quantify the magnitude of response in that region, repeating this procedure iteratively to cover all sensible partitions of the data and averaging across the magnitudes so obtained; see the section 'Orthogonality of the localizer contrast and the contrast of interest' for details). When addressing questions about functional resolution, the fixed-ROI method will use a conjunction of the contrasts *A* and *B* to evaluate the selectivity of the response in the ROI, while the subject-specific fROI method will examine the response to condition *B* when using condition *A* as the localizer contrast.

In Fig. 4 we illustrate the relative sensitivity and functional resolution advantages of the subject-specific fROI method over the group-level fixed-ROI method, for a range of average partial-coverage values, which are inversely related to the relative extent of inter-subject variability (x-axis), and single-subject design sensitivities (y-axis; these values characterize the "quality" of the functional data, with higher values representing stronger effects, less noise, and/or longer acquisitions). The plot illustrates the conditions that result in improved multi-subject sensitivity (left) and functional resolution (right) for the subject-specific fROI method (green area), or the group-level fixed-ROI method (red area).[10]

The subject-specific fROI method outperforms the fixed-ROI method across a range of different (realistic) conditions. For example, if the average extent of activation within an ROI (partial-coverage) is below 80%, the subject-specific method will outperform the fixed-ROI method in both sensitivity and functional resolution for any ROI size and any level of partial-coverage variability as long as the single-subject design sensitivity is above 4.4 (or, equivalently, when the localizer contrast power is above 80%). In contrast, the group-level fixed-ROI method will outperform the subject-specific fROI method only for extremely low-powered single-subject designs (with single-subject design sensitivity below 1.37, or, equivalently, with the localizer contrast power

below 2%[11]). For intermediate values of the localizer contrast power (between 2% and 80%), the relative advantages of one method over the other will depend on the ROI size and the level of partial-coverage variability, with larger ROIs and higher inter-subject variability leading to relative advantages for the subject-specific fROI method. In practice, it is not unreasonable to expect power levels of 90% or above for individual voxel-level analyses. For example, single-subject design sensitivity values explored by Desmond and Glover (2002) suggest that with $n = 100$ independent timepoints per condition, single-subject design sensitivity values range from 4 to 100, with average values around 25, which are well above the displayed range of single-subject sensitivity values in our analyses (from 0 to 6). This suggests that for a well-powered single-subject design (with sensitivity of 6 or above), the subject-specific fROI method will outperform the group-level fixed-ROI method in all cases except when the activation extent in every subject encompasses the entire ROI and the inter-subject overlap is perfect (average partial-coverage greater than 0.99), not a realistic possibility under any amount of inter-subject variability in the loci and extent of activation (Hellier et al., 2003).

Standard voxel-based analyses

As we noted above, because the use of functional localizers is typically framed in the context of ROI analyses, their advantages are often interpreted too narrowly (i.e., in terms of the relative merits of ROI- vs. whole-brain voxel-based analyses). We here discuss how the most widely used group analysis method, multi-subject voxel-based analyses, can be considered a special case of fixed-ROI analyses. In particular, voxel-based analyses can be thought of as fixed-ROI analyses with ROI size of 1 voxel (see Appendix C). In this case, partial-coverage values for each subject are dichotomous, equaling 1 if the subject shows activation in the relevant voxel, or 0 if the subject does not show activation in the voxel. The average partial-coverage value represents in this case the level of inter-subject overlap (i.e., the proportion of subjects for whom the voxel is active), and partial-coverage variability is determined

---

[10] To derive the power and false-discovery rates associated with each sensitivity level, we assumed that individual subjects' functional localizer masks were defined using an uncorrected voxel-level p<.001 false-positive level. We considered ROI sizes between 1 voxel and 100 resels (approximately 10% of the brain volume), partial-coverage values between 0 and $k(1-k)$, $n=100$ independent time-points per condition, and (for functional resolution comparisons) average between-subject variability estimates from Desmond and Glover (2002).

[11] Single-subject design sensitivities are expressed as signal-to-noise ratios, and they represent the overall quality of the functional data for each subject. Localizer power is expressed as a true-positive rate, and it represents the proportion of voxels that appear in the fROI localizer among those voxels that are truly active for this contrast (a measure related to the single-subject design sensitivity, but also to the false-positive threshold used in the localizer as well as to the proportion of sessions worth of data used for the localizer contrast).
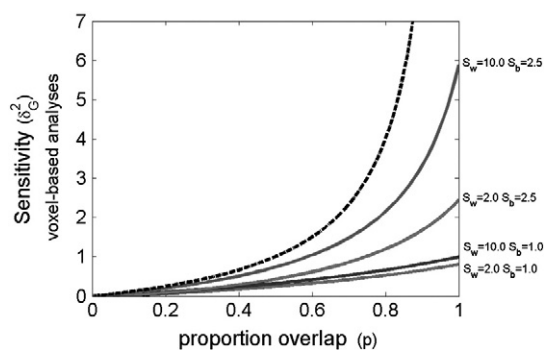
Fig. 5. Sensitivity of voxel-based analyses as a function of spatial variability in the loci of activations and as a function of within- and between-subject sensitivities. The sensitivity of multi-subject voxel-based analyses (y-axis) is plotted for different levels of inter-subject variability in the loci of activation, i.e., different proportion overlap values (x-axis) for a normal range of within- and between-subject sensitivity levels (Desmond and Glover, 2002) (solid lines). The dotted line shows the maximal power achievable by *any* (arbitrarily high) set of within- and between-subject sensitivities (that depend on how many scans we perform for each subject, and how strong and consistent the underlying effect is in the subjects where it is present), highlighting the fact that partial across-subject overlap imposes a concrete and severe limit on the maximal power achievable in group-level voxel-based analyses, and the size of this detrimental effect depends on the level of overlap.

fully by the average partial-coverage value (so it no longer plays a role independent from the partial-coverage average measure).

Fig. 5 illustrates the detrimental effect of reduced inter-subject overlap on the sensitivity of multi-subject voxel-based analyses. The values of the within- and between-subject sensitivities used in this plot (different solid lines) represent low-sensitivity and high-sensitivity values adapted from Desmond and Glover (2002). The dotted line represents the theoretical bound on the voxel-level sensitivity that can be achieved for *any* combination of within- and between-subject sensitivities. For reference, a sensitivity value of 1 corresponds approximately to a 90% power level in a study with 25 subjects with a false-positive level of p<.001. This bound means, for example, that in a study with 25 subjects, and when using an uncorrected voxel-level p-value of .001, we will only be able to detect with a power greater than 80% voxels that show 43% overlap or above across subjects, *irrespective of how strong the effect is* (no matter how many scans we perform for each subject, how strong the underlying effect is in the subjects where it is present, or how consistent this effect is across this reduced set of subjects—in the limit, zero between-subject variability). In summary, group-level voxel-based analyses are extremely sensitive to the degree of inter-subject overlap of activations.

The voxel-based sensitivity analyses presented here considered the effect of limited inter-subject overlap on any given voxel. A standard practice in the field is to use a certain amount of spatial smoothing (up to 12 mm FWHM) in order to increase the effective amount of overlap of activations across subjects (Scouten et al., 2006). With respect to the sensitivity of the resulting multi-subject analyses, smoothing is equivalent to increasing the ROI size in fixed-ROI analyses (see Appendix C for a more formal correspondence). As in the fixed-ROI case, increasing the amount of spatial smoothing only offers a partial solution to the presence of inter-subject variability in the loci of activation, and the sensitivity of group-level voxel-based analyses will still decrease with increasing amounts of inter-subject variability even in the case of optimally-sized smoothing kernels. In addition to these sensitivity limitations, there are two main caveats with multi-subject voxel-based analyses that use smoothing as a way to combat the effect of inter-subject variability in the loci of activation. (These are directly parallel to the caveats discussed for group-level fixed-ROI analyses.) First, smoothing potentially aggregates across both

active and inactive voxels so the effect sizes can be underestimated compared to those within the true loci of activation.[12] And second, smoothing effectively decreases the functional resolution of these analyses. In this case, the decrease in functional resolution can be characterized as an effective decrease in the spatial resolution of the multi-subject analyses. In other words, spatial smoothing limits the maximal spatial frequency of the effects that are detectable at the group level below the spatial resolution achievable by the single-subject analyses. Since the main goal of this paper is to discuss the merits of subject-specific localizers in the presence of inter-subject variability in the loci of activation, we will now discuss how the subject-specific approach – traditionally only applied to ROI-based analyses – can also be applied to voxel-based analyses. In particular, building on the analogy between smoothed voxel-level analyses and fixed-ROI analyses, we will now discuss a voxel-based analog of the subject-specific fROI method.

Subject-specific localizers in the context of whole-brain voxel-based analyses

This section introduces an improved method of smoothing for voxel-based analyses that uses functional localizers to obtain all of the gains typically associated with spatially smoothing the functional data (increased sensitivity and increased robustness to inter-subject variability in the loci of activation), without suffering from its detrimental effects (reduced functional resolution).

The motivation behind this method is straightforward: we saw in the context of the fixed-ROI method that increasing the size of the ROI to encompass the extent of expected inter-subject variability (matched filter theorem) is at best a partial solution to the problem of inter-subject variability in the loci of activation, and that a better solution is to increase the ROI size while limiting the within-ROI aggregation to functionally homogeneous voxels (using a functional localizer to identify functionally homogeneous areas). A direct application of this method to voxel-based analyses is to constrain smoothing at each voxel so that it operates only over functionally homogeneous neighboring voxels (identifying these voxels using an orthogonal functional localizer contrast, see Fig. 6). The result would be a subject-specific nonstationary smoothing operation informed by functional localizers (we call this method *subject-specific voxel-based analyses*, due to its similarities with the subject-specific fROI method), and its implementation, properties, and interpretation, are relatively simple to characterize, which we do next.

Implementation

First, a particular contrast between conditions is chosen to serve as a localizer, effectively masking the entire brain to include only voxels that pass a certain threshold level of activation for that contrast (in independent data). Then, the data from each voxel in the brain (whether included in the mask or not) are smoothed based on a given kernel (e.g., a Gaussian) that operates only on the voxels in that mask. Thus, each voxel is smoothed only with nearby voxels that have the functional property selected in the "localizer" contrast; any other nearby (and perhaps spatially interleaved) voxels that do not show this property are given zero weight in the smoothing kernel. This procedure enables an effect to be detected even if it is present in a small proportion of voxels in a local region, and even if other, spatially interspersed voxels show the opposite pattern of activation, because those other voxels will be excluded from the smoothing operation. However, those other voxels showing the opposite pattern can also be detected, if they show this

---

[12] As discussed in the case of fixed-ROI analyses, the notion of bias is relative to a particular model of neural activation. The analyses in this paper attempt to dissociate the extent of activation and the strength of the neural response. It is only with respect to this dissociated measure of the neural response strength (robust to small variations in the exact location and size of the activation near each voxel) that the effect size estimation of voxel-based analyses can be considered biased.
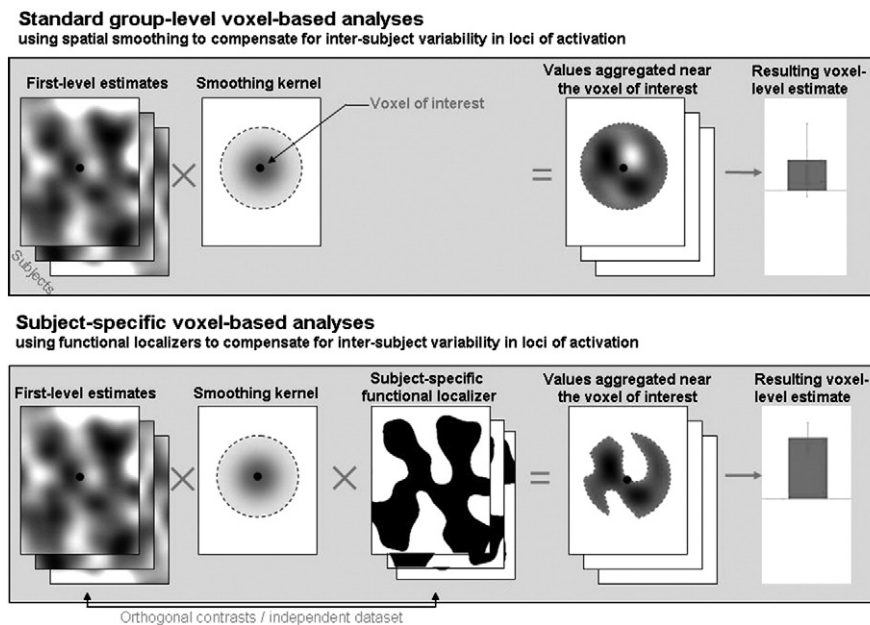
Fig. 6. A schematic illustration of the use of functional localizers in the context of voxel-based analyses. Standard voxel-level analyses (top) that use smoothing as a way to compensate for inter-subject variability in the loci of activation aggregate, for each voxel, the BOLD data (or single-subject estimates) across a surrounding area defined by the smoothing kernel support. The application of functional localizers in the context of voxel-based analyses (bottom) limits this aggregation to only those surrounding voxels within the subject-specific functional localizer mask (obtained from an orthogonal contrast or independent dataset). In the presence of inter-subject variability in the loci of activation this approach offers higher sensitivity and functional resolution, and a reduction of bias, in the resulting voxel-level estimates.

pattern replicably, in a parallel analysis of the same data that starts with a different (perhaps opposite) masking contrast. This method of selectively smoothing each voxel only with voxels that have been selected for a certain functional property, and excluding voxels that do not show this property, enables us to detect each of two distinct (and perhaps opposite) activation patterns that may arise in spatially interleaved voxels, and that would cancel each other in a standard (not-functionally-constrained) smoothing procedure. This analysis can thus reveal, via two different functional masks, that each of two different (and perhaps even opposite) contrasts can be present *in the same voxel*, which simply means that both patterns of functional response are reliably present in voxels located in the vicinity of that voxel. Importantly, this selective-smoothing procedure is not circular, since the actual response magnitudes and statistical activation maps are based on data independent of the functional data used in the smoothing mask, as in the subject-specific fROI analyses (see Fig. 7).

These benefits of selectively smoothing over functionally similar voxels in first-level individual-subject analyses, as just described, translate into benefits for second-level multi-subject analyses. Here again, separate analyses are conducted for each functional contrast/mask that is used to constrain the smoothing. Thus, all the individual subject data that have been smoothed based on a certain localizer contrast mask (as described above) would enter into one group analysis, and all the same data that have been smoothed as constrained by a different functional contrast, would enter into a different group analysis. In this fashion, the same voxel in the group space could show a significant effect of two different (and perhaps opposite) contrasts, again meaning that some voxels in the vicinity of that voxel reliably show one effect, and other voxels in its vicinity reliably show the other/opposite effect. Here too, of course, the actual data that go into the group analysis are independent of those used to constrain the smoothing. Crucially, by performing separate group analyses in which the smoothing of the data was constrained by two different functional contrasts, this method can detect two different effects in the group analysis even if there is no alignment

of the specific voxels showing those effects in the common space. In this sense, this method improves both sensitivity and functional resolution (see Appendix D for details), although, as with any smoothing procedure, there is a loss of spatial resolution.

Properties

Subject-specific voxel-based analyses share the qualitative and quantitative properties with the subject-specific fROI method, as they can be characterized as multiple small fROIs centered at every voxel and with size equal to the smoothing kernel support. In addition, since the traditional smoothing operation for voxel-based analyses can be similarly characterized as multiple small fixed-ROIs (equally centered at every voxel and with size equal to the smoothing kernel support), all of the properties and comparisons in the section 'Subject-specific functional ROI analyses' equally apply to the comparison between smoothed voxel-based analyses and subject-specific voxel-based analyses. In particular, for any size of the smoothing kernel, the subject-specific voxel-based analyses will show increased sensitivity, increased functional resolution, and reduced bias, compared to the standard (smoothed) voxel-based analyses, under the general conditions discussed in the section 'Subject-specific functional ROI analyses' (Fig. 4, see Appendix D for further details).

Interpretation

Subject-specific voxel-based analyses allow researchers to examine questions about the presence of an effect *at or near* each voxel, while (i) permitting variability in the precise location of the effect across subjects (unlike voxel-based analyses without spatial smoothing), and (ii) maintaining high functional resolution (i.e., the ability to discriminate functionally different effects within the support of the smoothing kernel), unlike voxel-based analyses with spatial smoothing, which aggregate across all of the voxels within the smoothing kernel support. To illustrate the importance of this latter point, consider the case of ocular dominance columns in V1. Let's assume for a moment, as a thought experiment, that the spatial resolution of
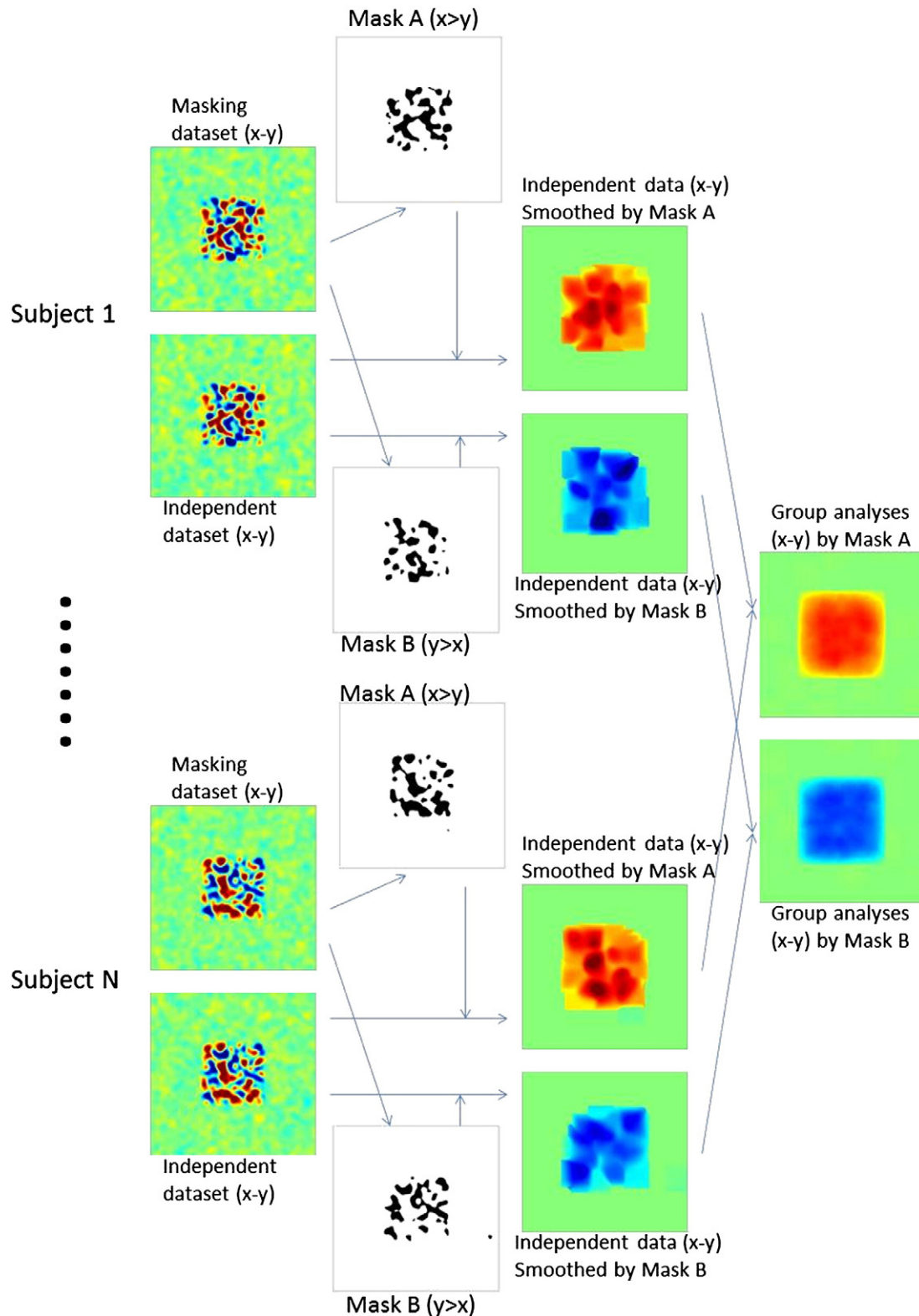
Fig. 7. Diagram describing subject-specific voxel-level analysis steps, in the presence of a distributed representation of two opposite effects. Each subject's data is divided into a *masking dataset* and a second *independent dataset* where statistical inferences will be evaluated. The masking dataset is used to derive subject-specific functional localizers for the two contrasts of interest (mask *A* for a x>y contrast, and mask *B* for the opposite y>x contrast). The independent dataset for each subject is then smoothed using two separate procedures, which operate only over voxels in mask *A* or *B*, respectively. Aggregating the resulting maps across multiple subjects allows the discovery of opposite distributed effects within the same area.

single-subject functional data was high enough to differentiate responses to right- vs. left-eye stimuli. Of course, due to their very fine spatial scale, ocular dominance maps cannot be expected to line

up across subjects, and spatial smoothing cannot be expected to help improve this alignment. If we were to perform a standard voxel-based paired *t*-test looking for left- vs. right-dominant responses in V1

across subjects, we would conclude that no voxels in V1 exhibit ocular dominance because of the lack of an appropriate inter-subject match of these responses (independent of the level of spatial smoothing). In contrast, a subject-specific voxel-based analysis using a sufficiently large smoothing kernel allows the researcher to explore the differential left- vs. right-dominant responses *at* or *near* each voxel while maintaining the functional resolution to discriminate these different responses within the support of the smoothing kernel, as discussed above. The diagram in Fig. 7 illustrates these analyses (with *x* and *y* standing for the responses to left- and right-eye stimuli, respectively). For example, by using a left > right ($x > y$) localizer contrast when exploring the non-orthogonal responses to left- and right-eye stimuli, the researcher would correctly identify V1 as showing left- vs. right-dominant responses at or near each voxel within V1 despite responses across different subjects not spatially overlapping (because only voxels with a left > right response for each subject would be aggregated around each voxel in these analyses). In addition, these analyses would also be able to determine that the left- and right-dominant areas within V1 are spatially non-overlapping (e.g., by using a $x > y$ localizer contrast to identify voxels that respond more strongly to left-eye stimuli than to right-eye stimuli and then examining the response of those voxels to the right-eye stimuli – *y* condition – or vice versa). In this thought experiment standard voxel-based analyses could not address any of these questions because the assumption of overlapping responses across subjects would not be met.

Simulation examples

In order to exemplify the behavior of the different analysis methods discussed in this paper in the presence of inter-subject variability in the loci of activation, we present a series of simulations. Fig. 8 illustrates the nature of the simulated data. The data were designed so that each subject had nearby but fully distinct (non-overlapping) areas that respond to stimulus *A* and stimulus *B*. Within the area of interest and for each

subject there are voxels that respond to *either A* or *B*, but there are no voxels that respond to *both* of the stimuli. In addition, the loci of activation (a sphere with a radius of 10 voxels) were varied across subjects (randomly distributed following a normal distribution with a standard deviation of 10 voxels). The strength of the responses to each stimulus type was also randomly distributed across subjects (following a normal distribution with a mean of 1% and a standard deviation of 0.25% BOLD signal change). Single-subject estimates of each subject's response were simulated by adding a random and spatially independent noise component with the mean 0% and a standard deviation of 0.25% BOLD signal change to each subject's modeled response.

The main goal of this simulated experiment was to investigate the individual and joint responses to the two stimuli as a function of the type of analysis. In particular, the research questions this experiment addresses are: a) whether the BOLD responses within the area of interest are modulated by any of the two stimuli (yes); b) whether the BOLD responses within the area of interest are differentially modulated by the two stimuli (yes); and c) whether the possible BOLD responses to these two different stimuli originate from the same locations/neural substrates (no). The bottom panel of Fig. 8 shows the results of a standard voxel-based analysis if there was no inter-subject variability in the loci of activation (i.e., if the activations were perfectly aligned across subjects). In the absence of inter-subject variability, these analyses would correctly identify: a) areas responding to *A* and/or *B* separately (first two plots); b) areas responding more strongly to *A* than *B*, and areas responding more strongly to *B* than *A* (next two plots); and c) no areas responding to both stimuli (A|B and B|A tests, last two plots). The simulations below illustrate how these conclusions would be affected by the presence of inter-subject variability in the loci of activation when using group-level vs. subject-specific analyses.

Four methods of analysis, discussed in the section 'Standard group-level fixed-ROI analyses' up to the section 'Subject-specific localizers in the context of whole-brain voxel-based analyses', respectively, were implemented: (1) a fixed-ROI approach, using a fixed (subject-
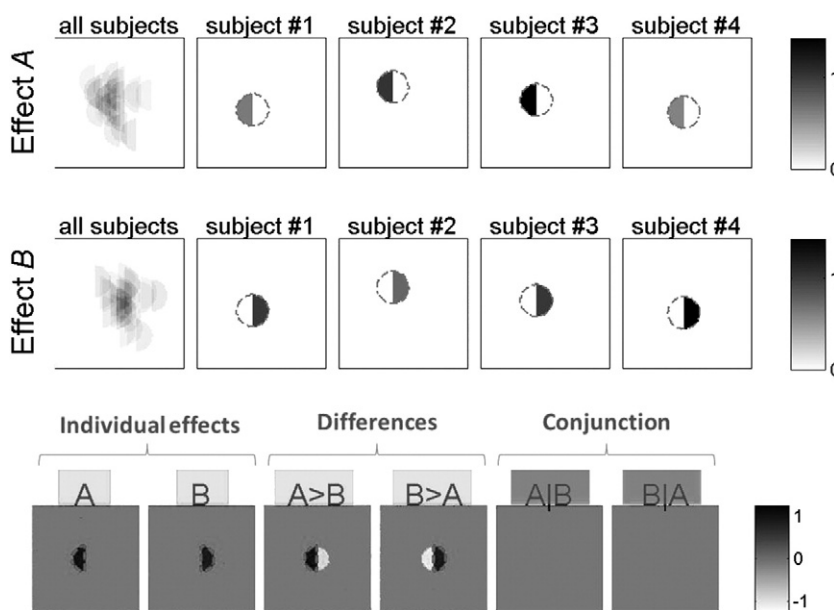

Fig. 8. Spatial distribution of simulated data. Within an area of 100 × 100 voxels, the entire activation of interest for each subject (n = 25) was assumed to lie within a sphere with a radius of 10 voxels. The left half of this sphere, for each subject, responded to stimulus *A* and the right half responded to stimulus *B*. The location of the sphere of activation varied randomly for each subject (with a standard deviation of 10 voxels). Top: the leftmost plots display the overlap among all of the subjects' response to each stimulus type. The four plots to the right show examples of the response strength and location of voxels activated by each condition in four sample subjects (added noise not shown). The strength of the BOLD responses (average 1% BOLD signal change) is color coded. Note that the responses to *A* and *B* are fully spatially separated for each subject (there is no individual voxel that responds to both *A* and *B*). Bottom: if there was no inter-subject variability in the loci of activation (i.e., if the spheres of activation perfectly aligned across subjects) a voxel-wise analysis would correctly indicate that: a) there are some areas that respond to *A* and/or *B*, and the effect sizes of the responses are approximately 1% BOLD signal change for each condition (*individual effects* plots; dotted grey line represents significant areas); b) these same areas respond differentially to the two stimuli (*differences* plots A > B and B > A); and c) there are no areas that respond jointly to both stimuli (*conjunction* plots A|B and B|A). Contrast inferences that are expected to be answered positively for at least some analysis unit are marked light grey (A, B, A > B, and B > A), while those that are expected to be answered negatively for all analysis units are marked dark grey (A|B, and B|A).

independent) spherical Gaussian ROI with 60 voxels FWHM and centered at the average locus of activation; (2) a subject-specific fROI approach constrained to the entire 100×100 voxel area; (3) a smoothed voxel-based approach using a 12 mm FWHM smoothing filter; and (4) a subject-specific voxel-based approach using the same 12 mm FWHM smoothing kernel. For each of these methods, we performed six group-level analyses addressing the three research questions, with the first two examining the main effects of A and B, the next two examining the differential (one-sided) effects $A > B$ and $B > A$, and the last two examining the conjunction of A and B in the form of the conditional effects $A|B$ and $B|A$. For the standard group-level methods, the conditional $A|B$ was computed by masking the estimated group-level response to A with the supra-threshold voxels ($p < .001$) of the estimated group-level response to B. For the subject-specific methods, the conditional $A|B$ was computed by estimating the effect of A when using B as the localizer contrast. Localizer contrasts were simulated using an independent and equally distributed dataset and supra-threshold voxels were computed using an FDR-corrected p-value of .05 (uncorrected single-subject p-values were estimated using a Gaussian approximation to the T-statistics appropriate for high degrees of freedom).

The results of the ROI-based analyses are shown in Fig. 9. Significant responses to each of the stimuli, A and B, are found ($p < .0001$) at the ROI level in both the fixed-ROI and the subject-specific fROI methods. However, the estimated effect size within the ROI is severely underestimated by the fixed-ROI method (estimated average effect sizes A: 0.05, B: 0.05 %BOLD signal, cf. modeled population with mean 1 %BOLD signal, the average true effects in the simulated sample were A: 1.02, B: 0.91). The amount of underestimation in the subject-specific fROI method is considerably lower (estimated average effect sizes A: 0.96, B: 0.85). When considering the $A > B$ and $B > A$ contrasts, aimed at detecting areas that respond differentially to the *two* stimuli, the fixed-ROI method fails to find areas with a significant difference ($p > .37$), while the subject-specific fROI method finds strong differential responses to A and B ($p < .0001$). As stated above, the simulated data contained voxels, for each subject, that responded *only* to either A or B. Because the fixed-ROI method averages across voxels that respond only to A and

voxels that respond only to B in approximately equal proportions, it has little *sensitivity* to find these differential responses. When considering the $A|B$ and $B|A$ contrasts, aimed at detecting areas that respond to *both* stimuli A and B, the fixed-ROI method finds strong evidence ($p < .0001$) for these types of responses, while the subject-specific fROI method finds no evidence ($p > .13$). (Again note that in the simulated data there were *no* voxels in any of the individual subjects that responded to both A and B.) Because the fixed-ROI method averages across some voxels that respond to A and some other voxels that respond to B, even if these voxels are never truly the same for any given subject, it has little *functional resolution* to discriminate against common responses to A and B.

The voxel-based analyses (Fig. 10) show the same pattern as the ROI-based analyses when comparing the standard group-level and the subject-specific methods. The areas that respond to A and B are successfully identified by both methods (dotted lines in the plots show significant areas at an uncorrected $p < .001$ level). The standard group-level method severely underestimates the effect sizes (average effect within significant voxels A: 0.17, B: 0.12 %BOLD signal change, over areas of 517 and 595 voxels, respectively), while the subject-specific method is much more accurate in estimating effect sizes (A: 1.08, B: 0.98 %BOLD signal change, over areas of 805 and 848 voxels, respectively). Similar to the fixed-ROI method, group-level voxel-based analyses show low sensitivity to the $A > B$ and $B > A$ contrasts (no significant voxels at $p < .001$) and low functional resolution indicated by false positive results in the $A|B$ and $B|A$ contrasts (the conjunction analyses show a cluster of 241 voxels that respond to both A and B). In contrast, the subject-specific method results in high sensitivity to the $A > B$ and $B > A$ contrasts (785 and 850 supra-threshold voxels, respectively), and high functional resolution indicated by true negative results in the $A|B$ and $B|A$ contrasts (no significant voxels at $p < .001$).

Regarding the sensitivity of the different methods, these simulations corroborate the theoretical analyses presented in the section 'Standard group-level fixed-ROI analyses' up to the section 'Subject-specific localizers in the context of whole-brain voxel-based analyses' and complement previous empirical demonstrations (e.g., Fedorenko et al.,
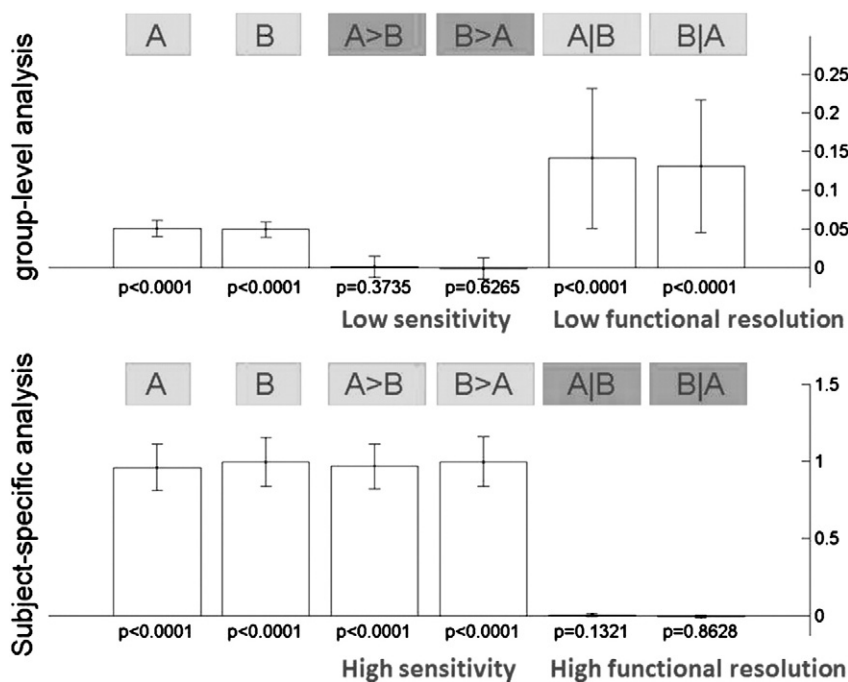

Fig. 9. Simulation results: ROI-based methods. Group-level fixed-ROI analyses (top) compared to subject-specific fROI analyses (bottom) on the same simulated dataset. Top: in the presence of inter-subject variability in the loci of activation group-level ROI-based results fail to find significant activation differences between the two conditions A and B ($A > B$ or $B > A$, $p > .37$), and incorrectly find significant responses to the conjunction of A and B ($p < .0001$). Bottom: subject-specific fROI analyses result in the correct inferences in all cases ($p < .0001$ for $A > B$ and $B > A$ tests, and $p > .13$ for $A|B$ and $B|A$ tests). Significant effects are marked light grey and non-significant effects are marked dark grey (compare to expected behavior in the absence of inter-subject variability, Fig. 10 bottom plot).
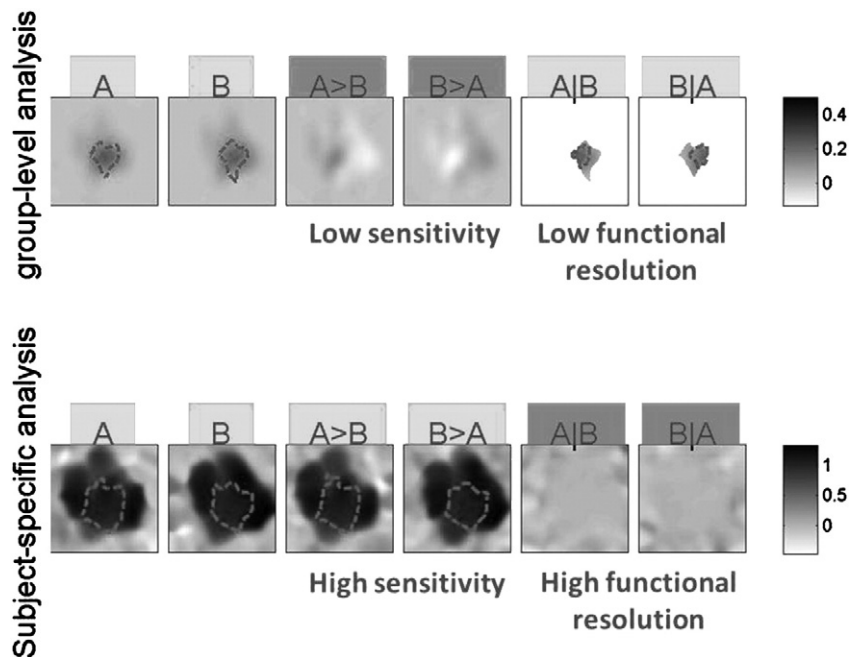
Fig. 10. Simulation results: voxel-based methods. Group-level voxel-based analyses with spatial smoothing (top) compared to subject-specific voxel-based analyses (bottom) on the same simulated dataset. Top: in the presence of inter-subject variability in the loci of activation group-level voxel-based results fail to find significant activation differences between the two conditions *A* and *B* ($A > B$ or $B > A$ hypotheses, no significant voxels using a height threshold level of p<.001), and incorrectly find significant responses to the conjunction of *A* and *B* ($A|B$ and $B|A$ hypotheses result in a cluster of 241 voxels at the same threshold level). Bottom: subject-specific voxel-based analyses result in the correct inferences in all cases (large clusters at a height threshold of p<.001 for the $A > B$ and $B > A$ hypotheses, and no significant voxels at this same threshold level for the $A|B$ and $B|A$ hypotheses). Significant effects are marked light grey and non-significant effects are marked dark grey (compare to expected behavior in the absence of inter-subject variability, Fig. 10 bottom plot).

2010, 2012; Saxe et al., 2006), showing that inter-subject variability in the loci of activation can dramatically decrease the sensitivity of standard group-level analyses (both ROI- and voxel-based). Furthermore, these simulations highlight that standard group-level methods consistently underestimate effect sizes and the selectivity of the areas under investigation. Consequently, relying exclusively on the results of group-level analyses (ROI- or voxel-based) may lead researchers to erroneously conclude that some region responds to both *A* and *B* in a non-differentiable way. In contrast, subject-specific analyses (both ROI- and voxel-based) would correctly indicate that there are voxels that respond only to stimulus *A*, others that respond only to stimulus *B*, and no voxels that respond to both. In summary, the subject-specific analysis methods offer an effective solution to the potential sensitivity, bias, and functional resolution issues present in the group-level fixed-ROI and voxel-based analyses, while offering robust inferences in the presence of inter-subject variability in the loci of activation.

Some practical considerations

Orthogonality of the localizer contrast and the contrast of interest

Subject-specific analyses require the localizer contrast and the contrast of interest to be orthogonal (e.g., Kriegeskorte et al., 2009; Vul and Kanwisher, 2010). In some cases, one might want to use the same contrast both as the localizer and as the contrast of interest (e.g., using *faces > objects* localizer contrast to define the ROIs and then estimating the size of the *faces-objects* effect in these areas). In other cases one might want to use different contrasts that may or may not be conceptually orthogonal (e.g., using the *faces > objects* contrast to define face-sensitive regions and then examining the response of these regions to *body parts*). In order to guarantee orthogonality, a typical solution involves partitioning the dataset and defining the localizer contrast and the contrast of interest using non-overlapping partitions (e.g., a

localizer contrast defined in the even runs, and a contrast of interest defined in the odd runs). However, a more powerful approach involves M-fold cross-validation. For example, in the context of *M* runs, one could define a contrast of interest A_1 that uses estimates only from the first run, associated with a localizer contrast B_1 that only uses estimates from the rest of the runs (2 to *M*). This could be repeated for each of the runs resulting in *M* pairs of localizer contrasts and associated contrasts of interest ($A_1 B_1$ to $A_M B_M$). One can then obtain a run-specific estimate from each of these pairs and average the resulting estimates to obtain the subject-specific estimates. This approach maintains the full within-subject sensitivity of the original data (cf. Neyman–Pearson lemma), while at the same time providing high power for the localizer contrasts (approaching the entire dataset power as the number of partitions increases).

Thus, all of the benefits of functional localizers as detailed in this paper can be obtained even for studies that were run without an explicit localizer contrast (see e.g., Friston et al., 2006, for arguments against using a separate localizer session). Researchers may simply use most of the data in any given contrast to localize the key region in each subject individually, and then use the remaining left-out data to quantify the magnitude of response in that region (repeating this iteratively to cover all sensible partitions of the data, and averaging across the magnitudes so obtained). In this fashion, it is possible to obtain all of the benefits of a functional localizer retroactively in the analysis of any experiment, even if an explicit localizer contrast was never run in the first place.

Population-level inferences in subject-specific analyses

The general method of multi-subject analyses that use functional localizers is the following: a) define a localizer contrast and its associated voxel-level threshold; b) define a contrast of interest; c) define the units of analysis (a set of a priori ROIs for ROI-based analyses,

or the extent of the "smoothing" kernel for voxel-based analyses); and d) perform multi-subject analyses of the contrast of interest, spatially constrained by (a) and (c). This procedure performs population-level inferences about the strength of the effect (b) spatially constrained to the subject-specific conjunction of (a) and (c). These inferences generalize to the subpopulation for which this conjunction is not empty within each analysis unit (ROIs or voxels). False-positive control of these inferences is afforded by resulting multi-subject analysis p-values. The threshold used to define the subject-specific voxels of interest from the localizer contrast (α_W) does not influence the presence of false positives in the population-level inferences, but it affects the size of the population that these inferences generalize to. In particular, highly conservative thresholds may lead to only a small percentage of subjects in the study sample showing *any* supra-threshold voxels within the unit of analysis (within a ROI or near a voxel). In this case, any inferences from this subset of the study sample will generalize to the same proportion of the population (e.g., if only 50% of the subjects show *some* localizer voxels within the a priori ROI, then any observed effect within this subsample will generalize to only 50% of the population). In practice and depending on the particular application, researchers might want to use a strongly corrected localizer threshold (low α_W values) or a more liberal one (high α_W values, see Duncan and Devlin, 2011, for a useful discussion). In the extreme liberal case of choosing a α_W threshold value of p = 1 (100% false positive level), the localizer contrast will include all voxels, and the subject-specific method will perform exactly like the standard group-level methods (fixed-ROI or voxel-based analyses with spatial smoothing). In this way, subject-specific methods generalize the corresponding standard (group-level) methods, simply adding one additional level of control (choice of α_W level on a separate localizer contrast) and performing exactly like these standard methods in the limiting minimal-specificity case ($\alpha_W = 1$), which disregards the localizer contrast information. The option of defining a separate localizer contrast and contrast of interest allows for greater flexibility in the research questions that can be addressed with this method.

After performing subject-specific analyses, a recommended practice would be to explore the subject-specific localizer masks (a) and compute the proportion overlap across subjects over the original units of analyses (c). This procedure enables the researcher to characterize the amount and nature of inter-subject variability in the loci of activation, and the proportion of subjects showing an effect at the level of the chosen unit of analysis (ROIs or voxels). Statistically significant results that show low inter-subject overlap might lack practical significance (as they relate to only a small proportion of the population), or they might indicate larger inter-subject variability in the loci of activation than anticipated (if this is the case, then using larger ROIs or smoothing kernels may be necessary).

The units of analysis (a priori ROIs for ROI-based analyses, or the extent of the "smoothing" kernel for voxel-based analyses) can be viewed as priors characterizing the expected spatial distribution of the responses to the localizer contrast across subjects. In practice, subject-specific analyses are considerably more resilient to ROI sizes (or smoothing kernel sizes) that extend beyond the true variability in the loci of activation across subjects than to those that fall short of the true variability (see Fig. 3, right). Depending on the particular application, researchers might use liberal priors (e.g., larger "smoothing" kernels) in order to maximize sensitivity, for example when exploring novel effects or areas, or more conservative ones in order to improve the specificity of the analyses, for example when studying areas that have been already characterized in previous studies.

Tools for performing subject-specific analyses

An SPM toolbox for performing all of the subject-specific analyses described in this paper is available at http://www.nitrc.org/projects/spm_ss. The toolbox is implemented as an alternative second-level analysis procedure, and it can be used on any existing set of first-level analysis results.

The toolbox implements both ROI- and voxel-based subject-specific analyses, performs automatic cross-validation across runs when the localizer contrast and the contrast of interest are not orthogonal, and implements both restricted maximum likelihood as well as ordinary least squares estimation of population-level effects and multivariate hypothesis testing for mixed within- and between-subject designs (see Appendix F for sample scripts and description of the analysis parameters).

Discussion and conclusions

In this study we quantified the conditions under which analyses that use functional data to constrain the correspondence of voxels or regions across subjects outperform standard stereotaxically registered group-level analyses in terms of their expected sensitivity (to detect significant activations) and functional resolution (to discriminate between two different functional activations). We demonstrated that in the presence of inter-subject variability in the loci of activations, analyses that take into account differences in the precise locations of functional activations across subjects have higher sensitivity and functional resolution, as well as a reduction in bias (i.e., more accurate estimation of effect sizes[13]), compared to traditional group analyses, even when the same total amount of data is used for each.

Our analyses have three important implications. *First,* from a methodological perspective, the merits of subject-specific functional localizers are independent of the debate about ROI- vs. voxel-based analyses (e.g., see the debate between Friston et al., 2006, and Saxe et al., 2006). In particular, to understand the advantages of subject-specific functional localizers, it is not relevant whether the units of analysis are ROIs or voxels. Both ROI- and voxel-based analyses (assuming some spatial smoothing) use some form of aggregation across voxels, even if they differ in the size of these aggregation regions. Fixed-ROI analyses aggregate across *all* voxels within a (typically large) ROI. Standard voxel-based analyses aggregate across *all* voxels in the vicinity (the support of a typically small spatial smoothing kernel) of each voxel. Irrespective of the size of the aggregation region, functional localizers allow a more selective aggregation: only across a subset of those voxels (within an ROI, or within a smoothing kernel support) that show supra-threshold responses in a functional localizer contrast. In this way, aggregation occurs over a subset of voxels that show more homogeneous functional responses, compared to methods that perform aggregation based only on stereotaxic coordinates without functional information. Such selective aggregation increases the functional resolution of the resulting inferences, since we can now choose to aggregate over different subsets of what was previously considered a homogeneous region. It could be argued that this increase in functional resolution comes at the cost of reduced sensitivity, since some portion of the available functional data must be used to compute the functional localizer contrasts. However, we showed – using theoretical analyses and simulations – that this intuition is, in general, not supported. On the contrary, sensitivity of multi-subject analyses not only does not decrease but, in most cases, increases when the aggregation across voxels is constrained by functional localizers (for the same amount of functional data). We argue that this benefit occurs because in standard group-level analyses sensitivity is reduced by the lower functional homogeneity across the aggregated voxels (e.g., in the presence of inter-subject variability in the loci of activation). Finally, in addition to increases in functional resolution and sensitivity, we showed that functional localizers result in a more accurate estimation of the effect sizes of functional responses.

---

[13] In this manuscript we consider a simple model of neural activation that considers a localized response, with unknown strength, position, size and shape, where all of these parameters vary across subjects (see Appendix A for details). The reported biases reflect a mismatch between the effect size measures from different analysis types and the modeled mean strength of (localized) activation.

The theoretical analyses and simulations presented here constitute a strong case for the use of functional localizers in neuroimaging studies, especially when investigating questions of functional selectivity (relative responses to multiple stimuli/conditions within the same area). Standard group-level methods and conjunction analyses are not particularly well suited to answer questions about functional selectivity, and they can easily miss important functional dissociations, especially in cases of adjacent but functionally distinct regions (as shown in the simulations in the section 'Simulation examples'). In contrast, subject-specific analyses allow researchers to explore the response to one stimulus/condition spatially constrained to those areas that show significant responses to a second stimulus/condition *within each subject*, resulting in a level of functional resolution only limited by the spatial resolution of the subject-level functional data (and most importantly not limited by the assumption of stereotaxic alignment of functional activations).

To illustrate this point with an example from the literature, consider the controversy currently surrounding language-sensitive regions and the questions related to the degree of functional specialization of those regions (similar controversies exist in other domains). A marked discrepancy exists between studies of patients with focal lesions that have revealed cases of highly specialized linguistic deficits and neuroimaging studies that have argued for overlap between language and a number of non-linguistic cognitive processes, such as arithmetic processing (e.g., Dehaene et al., 1999), musical processing (e.g., Levitin and Menon, 2003), general working memory (e.g., Owen et al., 2005), aspects of action representation (e.g., Binkofski et al., 2000), or general processing of hierarchically structured representations (e.g., Koechlin and Jubault, 2006) One possible reason for this discrepancy is that selectivity of language-sensitive cortex is underestimated in neuroimaging studies because they almost exclusively rely on group-level methods and in some cases don't even directly compare a language task to a non-language task, relying on approximating the locations of language activations in the common space (see Fedorenko et al., 2010, 2012, for additional discussion). This possibility is especially worrisome because regions engaged in similar kinds of computations (as claimed for aspects of language and some of the non-linguistic processes listed above) are plausibly located near each other. Therefore, in order to conclusively determine the extent of overlap between linguistic processing and candidate non-language processes, methods should be used that have been shown to be the most sensitive/selective, such as the methods that use subject-specific functional masks. Exclusive reliance on the data from group-based analyses may lead to misguided theorizing and incorrect models of the functional architecture of the language system, and human cognition more generally. Indeed, we have recently shown for the case of high-level language processing that fROI-based methods reveal much greater functional specificity than was apparent in prior work using traditional methods (Fedorenko et al., 2011).

Second, the analyses presented here point to an important practical advantage of the subject-specific methodology. In particular, improvements in the quality of individual subjects' data (due to better scanner acquisition sequences, larger field strengths, or improvements in physiological noise reduction methods) that result in increased single-subject sensitivity will directly lead to an increase in multi-subject *sensitivity* for the subject-specific analyses (either ROI- or whole-brain voxel-based). In contrast, these improvements will have at best a small effect on the sensitivity of the stereotaxically registered multi-subject analyses, because the power of such analyses will be limited by the amount of inter-subject variability (i.e., by the average coverage of the ROI volume in ROI-based analyses, or by the proportion of subjects showing an effect at any given voxel in voxel-based analyses). If we further consider the fact that increases in the spatial resolution of single-subject analyses can directly lead to an increase in multi-subject *functional resolution* for the subject-specific analyses (but not for the stereotaxically registered group analyses), we can conclude that subject-specific functional localizers

are optimally suited for taking advantage of current and future technical developments in functional neuroimaging.

The limited effect that reductions in within-subject noise have on group-level statistics has been reported before (Desmond and Glover, 2002; Scouten et al., 2006). Unfortunately, many neuroimaging studies (i) still rely primarily on traditional group-based methods, rather than using high-powered within-subject designs, and (ii) often don't explore the locations of activations in individual subjects (Devlin and Poldrack, 2007). This reliance on group-based analyses has limited the potential impact of technical developments aimed at reducing within-subject noise or increasing within-subject resolution (e.g., Hutton et al., 2011; Logothetis, 2008; Preibisch et al., 2003; Stringer et al., 2011; Triantafyllou et al., 2011). The subject-specific methods discussed here can take full advantage of the improvements in SNR (increasing sensitivity) and within-subject resolution (increasing functional resolution), which would directly translate into similar improvements in multi-subject analyses. Furthermore, although current methods do not yet enable us to clearly relate fMRI activations to the cytoarchitectonic properties of the underlying cortical regions in individual brains, increases in the resolution of anatomical MRI may make such analyses possible in the foreseeable future (see e.g., Barbier et al., 2002; Walters et al., 2003; Bridge et al., 2005; Eickhoff et al., 2005; Bridge and Clare, 2006; Carmichael et al., 2006; Sigalovsky et al., 2006; Walters et al, 2007; Hinds et al., 2009; Turner et al., 2008, for promising strides in this direction). Paradigms that elicit robust functional activations at the individual-subject level will be of particular use in this enterprise.

The issue of native vs. common space, which we raised in the Introduction, is worth revisiting briefly, as it is relevant to high-resolution fMRI studies. As discussed in the Introduction, subject-pecific fROI analyses can be performed directly in the subject's native anatomical space (and at the resolution at which the data were acquired), or they can be performed on the normalized data (as long as an approximate inter-subject coregistration is possible). In either of those cases subject-specific analysis methods will take advantage of the increases in spatial resolution, which will translate directly into the increases in functional resolution. In contrast, standard group analyses performed in the common space would hardly benefit at all from the increased spatial resolution. Thus, high-resolution studies are necessarily restricted to analyses of single subjects' activation patterns, typically focusing on some region(s) of interest. The voxel-based subject-specific analysis presented here allows performing whole-brain analyses, which nevertheless retain the benefits of high spatial resolution.

And third, note that an explicit "localizer contrast" is not needed to benefit from the advantages of the subject-specific methods described in this paper, thus obviating concerns about the use of separate localizer sessions raised by Friston et al. (2006). In particular, Friston et al. (2006) discussed the advantages (increased sensitivity and better control of contextual factors) of using an orthogonal localizer contrast within a factorial design compared to a split session design (e.g., one localizer session and one main-experiment session). The cross-validation approach to multi-session studies – described in the section 'Orthogonality of the localizer contrast and the contrast of interest' – permits the application of subject-specific analyses irrespective of the chosen experimental design. Provided that the subject-level designs are reasonably high-powered and that multiple functional runs have been acquired for each subject, an iterated leave-one-out procedure across runs allows researchers to analyze any data collected from a single contrast. For example, all but one of the functional runs may be used to localize the region of interest in each subject individually, and the remaining left-out run may be used to quantify the magnitude of response in that region (repeating this iteratively to cover all sensible partitions of the data). This method allows the researcher to benefit from the increased sensitivity and a more accurate estimation of effect sizes of subject-specific analyses when considering a single contrast of interest (for both ROI- and voxel-based analyses). In addition, in the

presence of multiple contrasts of interest, researchers may also investigate the selectivity (relative strength) of the responses to these contrasts using the subject-specific methods described in this paper (e.g., exploring the response to one stimulus/condition spatially constrained to those areas that show significant responses to a second stimulus/condition *within each subject*), benefitting from the increased functional resolution of these analyses, even if none of these contrasts had been originally thought/designed as an explicit "localizer contrast". Because this procedure does not require orthogonality between the contrasts of interest, researchers have the flexibility of choosing a context-dependent localizer (e.g., in a factorial design) or a context-independent one (e.g., a separate localizer-specific task), depending on their specific goals, no longer constrained by the limitations of the analysis methods.

We hope that the analyses presented in this paper lead to increased interest in multi-subject studies that utilize high-powered within-subject designs and in further exploration and interpretation of the patterns of activation at the individual-subject level. These practices can help lead to a synergistic relationship between physicists who work to increase the power of single-subject analyses and scientists who could use that power to obtain a clearer picture of the functional architecture of human cognition. Of course, in deciding on an analysis method, some of the caveats raised previously with respect to some of the methods discussed here (e.g., ROI-based methods that use either subject-specific or fixed ROIs; Friston et al., 2006; Friston and Henson, 2006) should be kept in mind. For example, in some cases, focusing on a particular region or set of regions may lead to some degree of scientific "myopia", where researchers do not consider as deeply as they should i) what happens in the rest of the brain, outside of their ROIs, and/or ii) how the effects observed in the ROI(s) may be affected by activity in other parts of the brain. This approach may thus unwittingly constrain the hypotheses entertained about the ROI(s) and/or obscure important generalizations. ROI-based analyses should therefore always be supplemented with whole-brain voxel-based analyses, and the subject-specific version of such an analysis presented here allows researchers to reap the benefits typically associated with subject-specific fROI analyses (higher sensitivity, higher functional resolution, and more accurate estimation of effect sizes).

One way to reduce the differences in sensitivity and functional resolution between group-level vs. subject-specific methods is to develop better co-registration methods. As discussed in the Introduction, due to inter-subject variability in the locations of cytoarchitectonic areas with respect to sulci and gyri, even the most advanced methods – which attempt to align folding patterns across subjects (Fischl et al., 1999) – will be limited in their ability to align functional regions across brains (e.g., Frost and Goebel, 2012). Some recent work (e.g., Chen et al., 2012; Sabuncu et al., 2010; Yeo et al., 2010) has begun using functional activations to complement anatomical landmarks in the co-registration process. This line of work seems promising for the development of better methods for aligning brains in a common space. However, in order for these methods to be successful, powerful within-subject designs are still critical. Furthermore, we believe that increases in spatial resolution will ultimately challenge the premise of a one-to-one mapping between activation patterns across subjects (e.g., ocular dominance patterns in V1 cannot be expected to meaningfully match across subjects). In this way, we foresee that multi-subject analysis methods will be increasingly faced with the issue of between-subject variability irrespective of future improvements in inter-subject co-registration methods. The alternate methods sketched here provide a solution to this problem, by exploiting the functional activation patterns present in each subject to inform the ways that data are aggregated across subjects.

Acknowledgments

Appendix A. Bias and sensitivity in fixed-ROI analyses

Within a fixed-ROI, of size N_0 voxels, we will assume that the true activation locus for each subject encompass only a proportion of the voxels in the ROI (N_i voxels, for subject i),[14] and that the strength of activation within these active voxels is constant for each subject (purely homogenous area) and it can be characterized across-subjects by a random variable μ_i with mean μ and (between-subject) variability σ_B^2. In addition we will assume that measurements at each voxel are affected by a certain degree of estimation error characterized by a random variable ε with mean zero and (within-subject) variability σ_W^2. We will define the random variable k_i (with mean k and variance σ_k^2) as the ratio between N_i and N_0, characterizing the partial coverage of each subject-specific activation relative to the group-level ROI. For example, if in a given subject only half of the voxels within the ROI are truly active in the relevant contrast, k_i for this subject would equal 0.5. The average BOLD response x_i within this fixed-ROI for each subject can then be characterized as:

$$x_i = k_i \cdot \mu_i + N_0^{-1/2} \cdot \varepsilon_i.$$

The average (across n subjects) of the BOLD response within this fixed-ROI will then be a random variable with mean

$$\mu_G = k \cdot \mu$$

and variance

$$\sigma_G^2 = n^{-1} \cdot \left[\left(k^2 + \sigma_k^2 \right) \cdot \sigma_B^2 + \mu^2 \cdot \sigma_k^2 + N_0^{-1} \cdot \sigma_W^2 \right].$$

First, we could note that, as an estimator of the population-level mean μ of the strength of the activation, the above measure provides a biased estimate, underestimating μ by a factor proportional to the average partial-coverage measure k. Second, we should note that the variability is not only affected by the within- and between- subjects terms σ_W^2 and σ_B^2 but that variability increases with factors proportional to the average and variability in partial coverage measures (k and σ_k^2, respectively). Overall, the sensitivity of group-level fixed-ROI-based analyses, expressed as the signal-to-noise power ratio[15] δ^2, would be:

$$\delta_G^2 \equiv \frac{\mu_G^2}{\sigma_G^2} = \frac{n}{N_0^{-1} \cdot k^{-2} \cdot \delta_W^{-2} + \left(1 + \delta_k^{-2} \right) \cdot \delta_B^{-2} + \delta_k^{-2}} \tag{A1}$$

where δ_k^2 $(= k^2 / \sigma_k^2)$, δ_W^2 $(= \mu^2 / \sigma_W^2)$, and δ_B^2 $(= \mu^2 / \sigma_B^2)$, are the signal-to-noise power ratios of the partial-coverage, within-subject, and between-subject effects, respectively. In order to better understand the different factors affecting the sensitivity of fixed-ROI

---

[14] While referring to voxels for simplicity, more accurately volume units should be expressed in resels (resolution elements) to accommodate the spatial covariance of the BOLD measurements.

[15] The power ratio δ^2 measure is directly related to the true-positive rate β of the analyses. For example, for a one-sided T-test:

$$\beta = T(\delta - T_0, n - 1)$$
$$T_0 \equiv T^{-1}(1 - \alpha, n - 1)$$

where $T(t, dof)$ represents the cumulative distribution function of the t-statistic, α is the pre-specified false-positive rate and T^0 the corresponding t-statistics cutoff value.

analyses it is useful to consider some sensitivity bounds derived from considering the extreme cases for each of these effects.

First, if we consider the extreme case of perfect inter-subject overlap, with all of the voxels within the ROI being part of the true loci of activation for all subjects, the sensitivity reduces to the familiar form combining the within- and between-subjects sensitivities, with the additional (compared to voxel-based analyses) beneficial reduction in the within-subject variability resulting from averaging across multiple voxels/resels:

$$k \leq 1 \Rightarrow \delta_G^2 \leq \frac{n}{N_0^{-1} \cdot \delta_W^{-2} + \delta_B^{-2}}$$

In contrast, in the presence of partial coverage (where the true locus of activation for each subject only covers a portion of the ROI), the sensitivity of the multi-subject analyses is detrimentally affected not only by the reduction in average coverage (increase in k^{-2}) affecting the within-subject term (which can be partially compensated by improvements in single-subject sensitivity). It is also affected by the increase in the inter-subject variability in partial coverage values (characterized by the term δ_k). The sensitivity of fixed-ROI analyses, in the presence of inter-subject variability in the extent of activation within the ROI ($k<1$), will then always be smaller than the above extreme-case, only reaching this maximum sensitivity value in the case of perfect inter-subject overlap ($k=1$).

Second, if we consider the extreme case of perfect within-subject sensitivity ($\delta_w = \infty$, e.g., if we obtain an arbitrarily large number of scans for each subject), we see that the sensitivity of the group-level ROI analyses will be bounded by a value that depends not only on the between-subject variability but also on the distribution of partial coverage values:

$$\delta_W \leq \infty \Rightarrow \delta_G^2 \leq \frac{n}{\delta_B^{-2} + \delta_k^{-2} \cdot (1 + \delta_B^{-2})}$$

This bounds means, for example, that in a study with 25 subjects, and when using an uncorrected ROI-level p-value of 0.01, for an effect that, without inter-subject variability in partial coverage would be detectable at a power greater than 80%, the presence of inter-subject variability in the loci of activation (e.g., uniformly distributed loci between 1 and N_0 voxels across subjects) would reduce the power of these analyses to only 28%.

Third, it is also interesting to note that the sensitivity of fixed-ROI analyses is still bounded irrespective of *both* the within- and between-subject variabilities/sensitivities by a value that only depends on the distribution of partial coverage values:

$$\delta_W \leq \infty, \delta_B \leq \infty \Rightarrow \delta_G^2 \leq n \cdot \delta_k^2.$$

This illustrates that, even if we consider an arbitrarily strong and consistent effect, variability in its precise location or in its spatial extent across subjects can limit the sensitivity and consequently the power to find this effect using fixed-ROIs.

Appendix B. Bias and sensitivity in subject-specific ROI analyses

Functional localizers can be expected to detect a proportion β_W (the power of the independent within-subject analyses) of the true loci of activation (N_i voxels for each subject), while additionally including a proportion α_W of false-positives. Averaging

within these voxels eliminates the detrimental effect due to the variability in the sizes of the activation loci, as the average effect size within the active voxels is no longer modulated by the size of the loci of activation. This procedure also increases the functional resolution of the resulting analyses, as we are only considering those voxels that show a significant effect for each subject, and reduces the possible bias in effect size estimates. In particular, the average BOLD response x_i within this subject-specific ROI for each subject can then be characterized as:

$$x_i = (1 - \alpha_W) \cdot \mu_i + \left((1 - \alpha_W)^{-1} \cdot \beta_W \cdot N_i \right)^{-1/2} \cdot \varepsilon_i$$

where β_W and $1 - \alpha_W$ are the sensitivity and specificity,[16] respectively, of the within-subject localizer contrast. The weighted average[17] (across n subjects) of the BOLD response within this subject-specific ROI will have a mean:

$$\mu_G = (1 - \alpha_W) \cdot \mu$$

and variance:

$$\sigma_G^2 = n^{-1} \cdot \left[\left(1 + \delta_k^{-2} \right) \cdot (1 - \alpha_W)^2 \cdot \sigma_B^2 + N_0^{-1} \cdot k^{-1} \cdot (1 - \alpha_W) \cdot \beta^{-1} \cdot \sigma_W^2 \right].$$

First, we could note that, as an estimator of the population-level mean μ of the strength of the activation, the above measure provides a reduced amount of bias, underestimating by a factor proportional to the specificity of the localizer contrast $(1 - \alpha_W)$, compared to the underestimation by a factor proportional to the average partial-coverage measure (k) in the case of fixed-ROI analyses. This is important because, while k represents the coverage within the ROI and this is typically a property of the data beyond the control of the analyst, $1 - \alpha_W$ represent the specificity of the within-subject analyses, which we can potentially make arbitrarily large (for example, by simply obtaining more acquisitions for the localizer contrast, or by taking advantage of improvements in scanning acquisition techniques). Second, we should note that the variability is no longer affected by any term that is independent of the within- and between-subjects terms σ_W^2 and σ_B^2, which we will see has important implications in the robustness of these analyses to inter-subject variability in the locus and extent of activations.

---

[16] α_W is expressed as false discovery rate. In these derivations we are disregarding the effect of the variability of actual false discovery rates across subjects since that variability has a small and bounded effect when conservative FDR-controlled localizers are used. For more liberal localizer thresholds it may be necessary to take into account this additional source of variance for a more accurate representation of the sensitivity of these analyses.

[17] We are considering first a simple estimator that uses whitening to compensate for differences in within-subject variances. In particular, this estimator takes the form of a weighted average of the subject-specific signals x_i using as weights the number of observed supra-threshold voxels in the localizer contrast for each subject. This estimator is not the maximum likelihood estimator (which would consider the relative contributions of both the within- and between-subject variances to determine the optimal whitening weights, see the discussion at the end of this appendix). We chose to discuss first this simple but suboptimal estimator instead because: a) it results in the theoretical sensitivity equations that are most directly and easily comparable to the fixed-ROI case; and b) it can be considered a limiting case of the more sensitive maximum likelihood estimator discussed below.

Overall, the sensitivity of subject-specific fROI analyses, expressed as the signal-to-noise power ratio δ^2, would be:

$$\delta_G^2 \equiv \frac{\mu_G^2}{\sigma_G^2} = \frac{n}{N_0^{-1} \cdot k^{-1} \cdot (1-\alpha_w)^{-1} \cdot \beta_W^{-1} \cdot \delta_W^{-2} + (1 + \delta_k^{-2}) \cdot \delta_B^{-2}}. \tag{B1}$$

Comparing this equation to the sensitivity of the fixed-ROI analyses in the previous section (Eq. (A1)), we can directly derive the conditions for which the subject-specific analyses will result in improved sensitivity:

$$\delta_{SS}^2 > \delta_{fixed}^2 \Leftrightarrow (1-\alpha_w) \cdot \beta_W \cdot \left(1 + \delta_W^2 \cdot N_0 \cdot \sigma_k^2\right) > k \tag{B2}$$

It can be seen that, for any arbitrary distribution of partial-coverage values (characterized by k and σ_k^2 values), and as long as the between-subject variability in partial coverage values σ_k^2 is different from zero, we can always find a minimal single-subject power (characterized by δ_W^2 and β_W that will result in improved sensitivity when using the subject-specific fROI method compared to the fixed ROI method.[18]

If instead of the simple weighted average estimator discussed above we use a restricted maximum likelihood estimation (reML) approach, we can define a set of more general whitening factors of the form:

$$w_i = \left(r + N_i^{-1}\right)^{-1/2}.$$

The constant r (different for each ROI) is obtained in the process of maximum likelihood estimation, and it represents the ratio of the between- to within-subjects variance. The reML approach reduces to a line search maximization of the residual likelihood function:

$$L = -\log\left(\sigma_\varepsilon^2\right) + \frac{1}{n}\sum_{i=1}^{n} \log(w_i)$$

where $\sigma_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n} w_i \cdot (\varepsilon_i - \bar{\varepsilon})^2$ and $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} w_i \cdot \varepsilon_i$ represents the sample weighted variance and mean, respectively, of the residuals ε_i obtained from a first pass of the second-level general linear model. The sensitivity of subject-specific fROI analyses using this reML estimator would be:

$$\delta_G^2 \equiv \frac{\mu_G^2}{\sigma_G^2} = \sum_{i=1}^{n} \frac{1}{N_0^{-1} \cdot k_i^{-1} \cdot (1-\alpha_w)^{-1} \cdot \beta_W^{-1} \cdot \delta_W^{-2} + \delta_B^{-2}} \tag{B3}$$

which is always equal or greater than the sensitivity of the simplified case considered in Eq. (B1) (equality achieved in the case of constant partial-coverage values or infinite within-subject variance). Also note that, if we consider the extreme case of perfect within-subject

sensitivity ($\delta_w = \infty$, e.g., if we obtain an arbitrarily large number of scans for each subject), the sensitivity of the reML estimator would be:

$$\delta_W \leq \infty \Rightarrow \delta_G^2 \leq \frac{n}{\delta_B^{-2}}$$

where the detrimental effect of limited partial-coverage values has been completely removed (cf. compare this to the equivalent extreme case in Appendix A).

Appendix C. Bias and sensitivity in (unsmoothed) voxel-level analyses

Standard voxel-level analyses assume a homogeneous subject population at the level of each voxel. Here we will consider how power in the group-level analyses is affected by a relaxation of this homogeneity assumption in the presence of reduced inter-subject overlap. In particular we will consider here the possibility that due to inter-subject variability in the precise location of the functional activations, for any given voxel, only a proportion p of the subjects present a true effect (while the rest present a null effect). These initial analyses consider a voxel in isolation where the true activation can be considered to be either present or absent for any given subject. We characterize this effect by the random variable p_i, taking the values 1 with probability p or 0 with probability $1-p$. In this case, the BOLD response x_i at a given voxel for each subject can be characterized as:

$$x_i = p_i \cdot \mu_i + \varepsilon_i.$$

The average (across n subjects) of the BOLD response at a given voxel will then be a random variable with mean

$$\mu_G = p \cdot \mu$$

and variance

$$\sigma_G^2 = n^{-1} \cdot \left[p \cdot \sigma_B^2 + \mu^2 \cdot p \cdot (1-p) + \sigma_W^2\right].$$

First, we could note that, as an estimator of the population-level mean μ of the strength of the activation, the above measure provides a biased estimate, underestimating μ by a factor proportional to the proportion overlap p. Second, we should note that the variability is not only affected by the within- and between-subjects terms σ_W^2 and σ_B^2 but that variability increases with factors proportional to the proportion overlap measure p. Overall, the sensitivity of group-level voxel-based analyses, expressed as the signal-to-noise power ratio δ^2, would be:

$$\delta_G^2 \equiv \frac{\mu_G^2}{\sigma_G^2} = \frac{n \cdot p^2}{\delta_W^{-2} + p \cdot \delta_B^{-2} + p \cdot (1-p)}. \tag{C1}$$

At this point it might be worth noting that, unsurprisingly, this equation can be considered a special case of Eq. (A1) that expresses the sensitivity of fixed-ROI methods. If we consider $N_0 = 1$ (one voxel), and assume that the random variable N_i (number of active voxels) follows a Bernoulli distribution with parameter p, Eq. (C1) directly follows from Eq. (A1). It is interesting to consider in this context the interpretation of some of the limitations found in the context of fixed-ROI methods.

First, if there is a complete inter-subject overlap ($p = 1$, all of the subjects show an effect at this voxel), the signal-to-noise ratio reduces

---

[18] Eq. (B2) assumes that the within-subject signal-to-noise power ratio δ_W^2 is the same for the two methods. For comparisons that consider the same amount of functional data available for the two methods this assumption is only valid if the localizer contrasts have been defined using cross-validation. For methods that use a simpler partitioning of the data (e.g. one session used to define the localizer contrast, and the rest of the sessions used to define the other effects/contrasts of interest) the left-hand side of Eq. (B2) should include one additional multiplicative $(1-\gamma)$ factor, where γ characterizes the proportion of the data used to obtain the localizer. Similarly in this case the localizer power β_w will be equal to $\beta = T(\sqrt{\gamma} \cdot \delta_w - T_0, n-1)$. Note, nevertheless, that the above proposition still holds in this more liberal scenario.

to the familiar form (the combination of within- and between-subjects sensitivities):

$$\delta_G^2 = \frac{n}{\delta_W^{-2} + \delta_B^{-2}}.$$

In contrast, in the presence of reduced inter-subject overlap, and irrespective of the within- and between-subjects sensitivities, the sensitivity of the group-level analyses is always bounded by a maximum value that depends on the amount of inter-subject overlap p:

$$\delta_G^2 \leq \frac{n \cdot p}{1 - p}$$

Importantly, this bound means that, for any false-positive level chosen, and any desired power-level, there will be a minimal amount of inter-subject overlap necessary to achieve this power level, no matter how strong the original signal is in each individual subject. The minimal amount of inter-subject overlap would be:

$$p_{\min} = \frac{\delta_0^2}{n + \delta_0^2}$$

where $\delta_0 \equiv T^{-1}(\beta, n-1) + T^{-1}(1-\alpha, n-1)$, β is the desired power level, α is the false-positive level used, and n is the number of subjects. For example, in a study with 25 subjects, and when using an uncorrected voxel-level p-value of 0.001, we will only be able to detect with a power greater than 80% voxels that shows 43% overlap or above across subjects, *irrespective of how strong the effect is* (no matter how many scans we perform for each subject, how strong the underlying effect is in the subjects where it is present, or how consistent this effect is across this reduced set of subjects—in the limit, zero between-subject variability).

The voxel-based sensitivity analyses presented here considered the effect of limited inter-subject overlap on any given voxel. A standard practice in the field is to use a certain amount of spatial smoothing (up to 12 mm FWHM) in order to increase the effective amount of overlap of activations across subjects (Scouten et al., 2006). Yet smoothing introduces a new source of variability in the analyses due to the possible variability in the extent of true activations within the smoothing kernel support. In addition, smoothing can decrease the functional resolution of the analyses and result in underestimation of effect sizes as a result of averaging within each subject across some voxels that do and some voxels that do not show a true effect. The reasoning here is similar to the reasoning used in the context of fixed-ROI methods. Indeed, spatially smoothed voxel-based analyses can also be conceptually treated as a general form of the fixed-ROI analyses considered here, where the analyses are performed for every possible ROI (one ROI centered at each voxel, and with the size of the ROI determined by the size of the spatial smoothing kernel). As such, these analyses enjoy the same general advantages, and suffer from the same caveats, as the fixed-ROI analyses considered in the Introduction. In practice, Eq. (A1) can be directly applied to study the power of voxel-based analyses in the presence of spatial smoothing if we simply extend this to "soft" ROIs by considering N_0 to represent the effective size of the smoothing kernel (e.g., using the Welch–Satterthwaite equation):

$$N_0 = \frac{\left(\int h(x)dx\right)^2}{\int h(x)^2 dx} = \left(4\pi\sigma^2\right)^{3/2} \approx 3.4115 \cdot FWHM^3$$

where N_0 and *FWHM* are both expressed in voxels here, and if we extend k to represent a weighted-average measure of partial

coverage, computed as the average of the presence/absence of an effect p_i at the surrounding voxels and using $h(x)$ as a probability measure:

$$k_i = \frac{\int h(x) \cdot p_i(x)dx}{\int h(x)dx}.$$

In practice and when considering Eq. (A1) in the context of smoothed voxel-level analyses, N_0 can typically be assumed to take values close to one resel, and the random variable k_i can be thought to represent the proportion of truly active voxels within the support of the smoothing kernel for each subject.

Appendix D. Subject-specific voxel-level analyses

These analyses can be specified simply by a transformation of the subject-specific contrast of interest volumes $\beta_i(x)$ (volume of estimated single-subject effect of interest for subject i of the form:

$$\tilde{\beta}_i(x) = \frac{[\beta_i(x) \cdot T_i(x)] \otimes h(x)}{T_i(x) \otimes h(x)}$$

where $\otimes$ represents the convolution operator, $T_i(x)$ represents a thresholded subject-specific localizer-contrast volume (taking the value 1 at supra-threshold voxels, and 0 otherwise), and $h(x)$ represents the ROI-defining spatial "smoothing" kernel. Note that if $T_i(x) = 1$ for all voxels x, this transformation reduces to the familiar spatial smoothing strategy[19] using the smoothing kernel $h(x)$:

$$\tilde{\beta}_i(x) = \frac{\beta_i(x) \otimes h(x)}{\int h(x)dx}$$

In contrast, when $T_i(x)$ represents a subject-specific thresholded localizer contrast volume, the transformed volumes $\tilde{\beta}_i(x)$ represent the weighted average among supra-threshold voxels *for each subject* and within an ROI centered at each voxel and with size determined by the support of $h(x)$. In this way, group-level analyses performed on the transformed volumes effectively implement the subject-specific analysis methodology in the context of voxel-based analyses. As in the case of subject-specific fROIs and in order to compensate for heteroscedasticity we can use a voxel-specific whitening factor of the form:

$$w_i(x) = [T_i(x) \otimes h(x)]^{1/2}$$

which results in the weighted average estimator discussed in the context of sensitivity analyses (with weights $w_i(x)^2$), or more generally we can use reML estimation leading to a voxel-specific whitening factor of the form:

$$w_i(x) \equiv \left(r(x) + [T_i(x) \otimes h(x)]^{-1}\right)^{-1/2}$$

where $r(x)$ represents the ratio of between- to within-subjects variance implicitly estimated over the MLE procedure (see Appendix B). Unlike the traditional group-level voxel-based analyses, the effective degrees of freedom of the subject-specific voxel-based analyses will be nonstationary. Methods that do not assume spatially stationary degrees of freedom (e.g., permutation tests; Hayasaka et al., 2004)

---

[19] Smoothing the beta volumes is equivalent to smoothing the original BOLD time-series, for any linear first-level analyses.

should be preferred when performing cluster- or set-level inferences from these voxel-based analyses.

In the same way that Eq. (A1) could be directly applied to study the sensitivity of voxel-based analyses in the presence of spatial smoothing, Eq. (B1) can be directly applied to study the sensitivity of the presented subject-specific voxel-based analyses by simply considering N_0 to represent the effective size of the smoothing kernel $h(x)$ and k to represent a weighted measure of partial coverage. In this way, the advantages of functional localizers can be shown to extend to both ROI- and voxel-based analyses. In particular, for any given smoothing kernel size, using subject-specific voxel-based analyses will result in increased sensitivity, functional resolution, and reduced bias, compared to the standard voxel-based analyses in the presence of smoothing, under the general conditions discussed in the section 'Methods and results'.

Appendix E. Functional resolution in fixed-ROI and subject-specific fROI analyses

Functional resolution refers to the ability to detect an effect A while discriminating against a spatially adjacent but non-overlapping second effect B. It can be generally characterized as the joint likelihood of detecting effect A while not detecting effect B:

$$S = \beta_A \cdot \left(1 - \beta_{B|A}\right)$$

where β_A represents the sensitivity (power) to detect the effect A, and $\beta_{B|A}$ represents the a posteriori sensitivity to detect the effect B given A. In the case where effect B represents a null effect, this measure of functional resolution equals the product of the true-positive rate times the true-negative rate. This measure is also known as the AUC (area under the curve), representing the area under the receiver operating characteristic (ROC) curve, and characterizing the *discrimination* ability of a statistical test.

In order to address the impact of inter-subject variability in the loci of activation on the functional resolution of the different analysis approaches, we will consider a given true loci of activation of interest (an area that responds selectively to effect A) surrounded by areas that show a markedly different response (areas that respond selectively to effect B). We will assume that the researcher defines a ROI around the expected location of the true loci of activation (where the effect A alone should be present), but due to inter-subject variability in the exact location/shape of this locus of activation, the effect B might also be present within the ROI for a small proportion of voxels (due to "leakage" from nearby areas where the effect B instead of A is present). We will investigate the ability of the different analysis approaches to identify the effect A while discriminating against the effect B in the presence of inter-subject variability in the loci of activation.

In particular, following the definitions in Appendix A we will consider ROI-based analyses with a given ROI of size N_0, where the loci of activation of interest within this ROI covers N_i voxels for each subject i (with k_i as the ratio between N_i and N_0). These voxels respond to the effect A and not B (the strength of activation to the condition A within these voxels is characterized by a random variable with mean μ and within- and between-subjects variability σ_W^2 and σ_B^2, while for the condition B it is characterized by a random variable with zero-mean and within-subject variability σ_W^2). The rest of the voxels ($N_0 - N_i$) respond to the effect B and not A (opposite pattern than before, same mean and variability values for simplicity). In addition the effects A and B are assumed to be independent and all tests are assumed one-sided.

Fixed-ROI analyses

Fixed-ROI analyses exploring selective responses to A can be implemented through a group-level conjunction of A and B (exploring the responses to B over those ROIs that show a significant response to A). The functional resolution of fixed-ROI analyses can then be expressed as:

$$S = \beta_A \cdot (1 - \beta_B) = T\left(\sqrt{\delta_G^A} - t\right) \cdot \left(1 - T\left(\sqrt{\delta_G^B} - t\right)\right)$$

where T represents the cumulative distribution of the T-statistic, t represents a t-value threshold (inverse cumulative distribution of the T-statistic at the chosen false positive level), δ_G^A represents the sensitivity (power-ratio) of the fixed-ROI analyses when considering the effect A (with partial coverage values k_i) in isolation:

$$\delta_G^A = \frac{n}{N_0^{-1} \cdot k^{-2} \cdot \delta_W^{-2} + (1 + \delta_B^{-2}) \cdot \delta_k^{-2} + \delta_B^{-2}}$$

and δ_G^B represents the sensitivity of the fixed-ROI analyses when considering the effect B in isolation. In this case, a proportion $1 - k$ of the ROI voxels will suffer from "leakage" of the effect B from nearby areas, and δ_G^B will take the value:

$$\delta_G^B = \frac{n}{N_0^{-1} \cdot (1-k)^{-2} \cdot \delta_W^{-2} + (1 + \delta_B^{-2}) \cdot \delta_k^{-2} \cdot (1-k)^{-2} \cdot k^2 + \delta_B^{-2}}$$

Subject-specific fROI analyses

Subject-specific fROI analyses exploring selective responses to A can be implemented through a subject-specific conjunction of A and B: within each ROI and for each subject, exploring the responses to B over those voxels that show a significant response to A (in other words B is the contrast of interest while A is the localizer contrast). The functional resolution of subject-specific fROI analyses can then be expressed as:

$$S = \beta_A \cdot \left(1 - \beta_{B|A}\right) = T\left(\sqrt{\delta_G^A} - t\right) \cdot \left(1 - T\left(\sqrt{\delta_G^{B|A}} - t\right)\right)$$

where δ_G^A represents the sensitivity (power-ratio) of the subject-specific fROI analysis when exploring the effect A in isolation:

$$\delta_G^A = \sum_{i=1}^{n} \frac{1}{N_0^{-1} \cdot k_i^{-1} \cdot (1 - \alpha_w)^{-1} \cdot \beta_W^{-1} \cdot \delta_W^{-2} + \delta_B^{-2}}$$

$$\geq \frac{n}{N_0^{-1} \cdot k^{-1} \cdot (1 - \alpha_w)^{-1} \cdot \beta_W^{-1} \cdot \delta_W^{-2} + (1 + \delta_k^{-2}) \cdot \delta_B^{-2}}$$

and $\delta_G^{B|A}$ represent the sensitivity (power-ratio) of the subject-specific fROI analyses when exploring the effect B while using the effect A as a functional localizer. In this case, only a small proportion (controlled by the false-positive level of the localizer contrast) of the voxels analyzed within each ROI will suffer from "leakage" of the effect B from nearby areas. The average response x_i to the effect B within the subject-specific ROI can then be characterized as:

$$x_i = \alpha_W \cdot \mu_i + \left((1 - \alpha_W)^{-1} \cdot \beta_W \cdot N_i\right)^{-1/2} \cdot \varepsilon_i$$

and $\delta_G^{B|A}$ will take the value:

$$\delta_G^{B|A} = \sum_{i=1}^{n} \frac{1}{N_0^{-1} \cdot k_i^{-1} \cdot \alpha_W^{-2} \cdot (1 - \alpha_W) \cdot \beta_W^{-1} \cdot \delta_W^{-2} + \delta_B^{-2}} \geq \frac{n}{N_0^{-1} \cdot k^{-1} \cdot \alpha_W^{-2} \cdot (1 - \alpha_W) \cdot \beta_W^{-1} \cdot \delta_W^{-2} + (1 + \delta_k^{-2}) \cdot \delta_B^{-2}}$$

Appendix F. Sample spm_ss toolbox script and details of the analysis parameters[20]

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SAMPLE SCRIPT[20] %

experiments=struct(...
    'name', 'exptname',...
    'pwd1', '/path/to/the/data/directory/',...
    'pwd2', 'firstlevel_modelname',...
    'data',{{'subj_01','subj_02','subj_03','subj_04','subj_05'}});

spmfiles={};
for nsub=1:length(experiments.data),

spmfiles{nsub}=fullfile(experiments.pwd1,experiments.data{nsub},experiments.pwd2
,'SPM.mat');
end

% Basic analysis parameters
ss=struct(...
    'swd', '/path/to/the/output/directory/',...
    'type', 'mROI',...
    'files_spm',{spmfiles},...
    'Localizer_contrasts',{{'A-B'}},...
    'Localizer_thr_type', 'FDR',...
    'Localizer_thr_p', .05,...
    'EffectOfInterest_contrasts',{{'A-B'}},...
    'model', 1,...
    'estimation', 'OLS',...
    'ExplicitMasking', none,...
    'ask', 'missing');

% Additional parameters for type=mROI
%     'ManualROIs', '/path/to/the/ROI/directory/ROIs.img',…
%     'overlap_thr_roi', .5,...

% Additional parameters for type=voxel
%     'smooth', 6,...

% Additional parameters for type=GcSS
%     'smooth', 6,...
%     'overlap_thr_voxel', .1,...
%     'overlap_thr_roi', .5,...

ss=spm_ss_design(ss);
ss=spm_ss_estimate(ss);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Explanation of the analysis parameters

type type of subject-specific analysis ("mROI": ROI-based (with pre-existing ROI volumes); "voxel": whole-brain voxel-based; "GcSS": whole-brain ROI-based (with ROI volumes being automatically defined based on the activation patterns observed across subjects)).

The *mROI analysis* is appropriate when the researcher already has some ROI volume(s) that they would like to use. These ROI volumes could, for example, be anatomical regions, functional parcels created from the GSS (formerly GcSS) analysis (see below), or ROIs defined based on peak coordinates/activated regions from other studies.

The *voxel analysis* and the *GcSS* (a.k.a. *GSS*) *analysis* are both appropriate when the researcher does not have spatial priors on where to expect to observe activations. And both of these analyses are alternatives to the traditional random-effects voxel-level analysis.

The *voxel analysis* is equivalent to an mROI analysis that considers small spherical ROIs centered at each voxel. This analysis asks *where* in the brain are activations spatially consistent across subjects, except that in contrast to the standard analysis the subject-specific method softens the assumption of perfect inter-subject coregistration (by testing hypotheses *at* or *near* a coordinate in normalized space, instead of exactly *at* a given coordinate in normalized space).

The *GcSS* (a.k.a. *GSS*) *analysis* also asks *where* in the brain are activations spatially consistent across subjects, except that in contrast to the standard analysis the GSS method does not require voxel-level inter-subject overlap, and hypotheses are tested at the level of empirically-obtained spatial parcels. Parcels generated in one study can be used in subsequent studies to constrain activations (e.g., we generated parcels for brain regions sensitive to high-level linguistic processing in Fedorenko et al., 2010, and now use those parcels in all subsequent studies).

Localizer_contrasts specify the localizer contrast name(s).

Localizer_thr_type this parameter determines the correction for multiple comparisons for the first-level voxel-based analysis used to define the localizer mask. It can be set to "none" (uncorrected), "FDR" (false discovery rate) or "FWE" (family-wise error), similar to the options built into SPM. In addition to these standard options, this parameter can also be set to "automatic" or "percentile". The "automatic" option represents some of our ongoing research. The analyses in this manuscript suggested that sensitivity of the subject-specific analyses was related to the "area under the curve" product $\beta_W(1-\alpha_W)$, where α_W and β_W represent the localizer false positive threshold (expressed as a false-discovery rate) and power, respectively. The "automatic" option finds the "optimal" false discovery rate threshold that empirically maximizes this product for each study. Last, the "percentile" option is a simpler option that, instead of selecting voxels based on their *absolute* voxel-level statistics, uses only their *relative* values, selecting a fixed percentage of voxels within each ROI as the localizer mask for each subject.

Localizer_thr_p localizer threshold value, in units determined by the value of the above "Localizer_thr_type" option (e.g., if choosing "FDR", this value represents the false discovery rate threshold value that should be used to create the localizer masks).

EffectOfInterest_contrasts specify the effect-of-interest contrast name(s).

model select and define the second-level analysis model and between-subject contrast. Currently implemented automatic models[21] include:

- 1—one-sample *t*-test (e.g., for single-group analyses);
- 2—two-sample *t*-test (e.g., for analyses comparing two subject groups);
- 3—multiple regression (e.g., for ANOVA or correlation analyses; e.g., Henson and Penny, 2003).

estimation select the estimation method. Currently implemented methods include:
- OLS (ordinary least squares): straight averaging;
- ReML (restricted maximum likelihood): weighted averaging.

[NB: the weighted averaging option is implementing a linear mixed model (or "mixed-effects" analysis, which is typically considered more robust than a pure "random-effects" analysis, particularly for first-level data that can be expected to show varying estimation error across subjects). The actual weight values are derived as a function of the number of voxels active for each subject, and the function relating the number of active voxels and the weight values is specific to each ROI and is estimated via a maximum likelihood fit, as the optimal form of this function depends on the within- and between-subject variances, which are not known a priori and thus must be estimated. The weights are reported in the *data*.csv file (under a subheading "Weights"). Note also that using the ReML

^{footnote}

---

[21] Other general linear models can also be manually defined entering the desired design matrix and contrast vectors, as in the SPM package.

estimation option will lead to degrees of freedom being non-integer numbers.]

ExplicitMasking (optional) this option allows using an explicit mask image so that the analyses are limited to voxels within the mask. You can specify here an image file (for example the thresholded group-level results image), or leave the field empty if you do not want to use any explicit mask. If you use a mask file, the ROI analyses will be restricted to the intersection of the subject-specific masks, the ROIs volume, and the explicit mask.

ask this parameter can be set to "none" (in which case any missing information is assumed to take default values), to "missing" (in which case any missing information will be asked about), or "all" (in which case each parameter will be asked about).

ManualROIs (for "mROI" analyses only) here you can specify the path to the ROI volume(s). This should be an analyze or nifti image containing natural numbers, each number referring to a single ROI (e.g., a volume containing 1's for voxels within ROI#1, 2's for voxels within ROI#2, and 0's otherwise).

overlap_thr_roi the ROI-level overlap threshold limits what areas (for voxel-based analyses) or ROIs (for ROI-based analyses) are considered in the results of the subject-specific analyses. This threshold (the default value is .5) represents the minimum proportion of subjects that need to show any significant localizer effect within an ROI (for ROI-based analyses) or within a sphere around a voxel (for voxel-based analyses) in order for this ROI or voxel to be considered in the display of the analysis results. Depending on the localizer threshold, there might be some subjects that do not show any localizer effects within an ROI (or near a voxel). The current parameter allows researchers to disregard ROIs/voxels where only a small subset of the population shows the effect), by choosing the minimum proportion of the population that they wish their inferences to apply to. When interested in effects that are widely prevalent in the population, choose a relatively large ROI-level overlap threshold (e.g., .8–.9). For a more exploratory approach or when interested in effects that may exist in only a small subset of the population, choose a relatively small low ROI-level overlap threshold (e.g., .1).

smooth (for "voxel" and "GcSS" analyses only) select the level of smoothing. For analyses of type "voxel", this parameter defines the extent of the "smoothing" kernel (i.e., a soft region of interest around each voxel). For analyses of type "GcSS", this parameter defines the smoothing kernel used to smooth the estimated inter-subject overlap map to improve the parcellation (in particular, to prevent over-parcellation; see e.g., Fedorenko et al., 2010).

overlap_thr_voxel (for "GcSS" analyses only) this parameter defines the overlap threshold (i.e., minimum proportion of subjects showing an effect in each voxel) when empirically estimating the parcels of interest. This parameter constrains the extent of these parcels. For example, you can eliminate from the overlap

map voxels where fewer than .10 of the subjects show activation by setting the threshold to .10. [NB: researchers tend to overestimate the extent of voxel-level inter-subject overlap. In our experience, the overlap typically does not exceed ~.70, and for most voxels it is considerably lower. As a result, we recommend using unthresholded, or liberally thresholded, overlap maps.]

References

Aguirre, G.K., Detre, J.A., 2012. The development and future of perfusion fMRI for dynamic imaging of human brain activity. NeuroImage 62 (2), 1279–1285.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. J. Comp. Neurol. 412 (2), 319–341.

Barbier, E.L., Marrett, S., Danek, A., Vortmeyer, A., van Gelderen, P., Duyn, J., Bandettini, P., Grafman, J., Koretsky, A.P., 2002. Imaging cortical anatomy by high-resolution MR at 3.0 T: detection of the stripe of Gennari in visual area 17. Magn. Reson. Med. 48, 735–738.

Binkofski, F., Amunts, K., Stephan, K.M., Posse, S., Schormann, T., Freund, H.J., et al., 2000. Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. Hum. Brain Mapp. 11 (4), 273–285.

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

Bridge, H., Clare, S., 2006. High-resolution MRI: in vivo histology? Philos. Trans. R. Soc. Lond. B Biol. Sci. 361, 137–146.

Bridge, H., Clare, S., Jenkinson, M., Jezzard, P., Parker, A.J., Matthews, P.M., 2005. Independent anatomical and functional measures of the V1/V2 boundary in human visual cortex. J. Vis. 5, 93–102.

Brodmann, K., 1909. Vergleichende lokalisationslehre der grosshirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues. Barth, Leipzig.

Carmichael, D.W., Thomas, D.L., De Vita, E., Fernández-Seara, M.A., Chhina, N., Cooper, M., Sunderland, C., Randell, C., Turner, R., Ordidge, R.J., 2006. Improving whole brain structural MRI at 4.7 Tesla using 4 irregularly shaped receiver coils. NeuroImage 32, 1176–1184.

Chen, G., Fedorenko, E., Kanwisher, N., Golland, P., 2012. Deformation-invariant sparse coding for modeling spatial variability of functional patterns in the brain. Proc. neural information processing systems (NIPS) workshop on machine learning and interpretation in neuroimaging.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., Tsivkin, S., 1999. Sources of mathematical thinking: behavioral and brain-imaging evidence. Science 284, 970–974.

Derrfuss, J., Vogt, V.L., Fiebach, C.J., von Cramon, D.Y., Tittgemeyer, M., 2012. Functional organization of the left inferior precentral sulcus: dissociating the inferior frontal eye field and the inferior frontal junction. NeuroImage 59, 3829–3837.

Desmond, J.E., Glover, G.H., 2002. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. J. Neurosci. Methods 118, 115–128.

Devlin, J., Poldrack, R.A., 2007. In praise of tedious anatomy. NeuroImage 37, 1033–1041.

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

Duncan, K.J.K., Devlin, J.T., 2011. Improving the reliability of functional localizers. NeuroImage 57 (3), 1022–1030.

Duvernoy, H.M., 1991. The Human Brain. Springer-Verlag, New York.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25 (4), 1325–1335.

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

Fedorenko, E., Kanwisher, N., 2009. Neuroimaging of language: why hasn't a clearer picture emerged? Lang. Ling. Compass 3 (4), 839–865.

Fedorenko, E., Kanwisher, N., 2011. Functionally localizing language-sensitive regions in individual subjects with fMRI: a reply to Grodzinsky's critique of Fedorenko & Kanwisher (2009). Lang. Ling. Compass 5 (2), 78–94.

Fedorenko, E., Hsieh, P.-J., Nieto-Castanon, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. A new method for fMRI investigations of language: defining ROIs functionally in individual subjects. J. Neurophysiol. 104, 1177–1194.

Fedorenko, E., Behr, M., Kanwisher, N., 2011. Functional specificity for high-level linguistic processing in the human brain. PNAS 108 (39), 16428–16433.

Fedorenko, E., Nieto-Castanon, A., Kanwisher, K., 2012. Syntactic processing in the human brain: what we know, what we don't know, and a suggestion for how to proceed. Brain Lang. 120, 187–207.

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999. High-resolution inter subject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T.T., Mohlberg, H., Amunts, K., Zilles, K., 2008. Cortical folding patterns and predicting cytoarchitecture. Cereb. Cortex 18 (8), 1973–1980.

Friston, K.J., Henson, R.N., 2006. Commentary on: divide & conquer: a defense of functional localizers. NeuroImage 30, 1097–1099.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994. Assessing the significance of focal activations using their spatial extent. Hum. Brain Mapp. 1, 214–220.

Friston, K.J., Rotshtein, P., Geng, J.J., Sterzer, P., Henson, R.N., 2006. A critique of functional localizers. NeuroImage 30 (4), 1077–1087.

Frost, M.A., Goebel, R., 2012. Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. NeuroImage 59 (2), 1369–1381.

Geschwind, N., Levitsky, W., 1968. Human brain: left–right asymmetries in temporal speech region. Science 161, 186–187.

Grosbras, M.-H., Lobel, E., Van de Moortel, P.-F., Le Bihan, D., Berthoz, A., 1999. An anatomical landmark for the supplementary eye field revealed with fMRI. Cereb. Cortex 9, 705–711.

Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., Nichols, T.E., 2004. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage 22 (2), 676–687.

Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Le Goualher, G., Collins, D.L., et al., 2003. Retrospective evaluation of inter-subject brain registration. IEEE Trans. Med. Imaging 22 (9), 1120–1130.

Henson, R.N.A., Penny, W.D., 2003. ANOVAs and SPM. Technical report, Wellcome Department of Imaging and Neuroscience.

Hickok, G., Okada, K., Serences, J., 2009. Area Spt in the human planum temporale supports sensory–motor interaction for speech processing. J. Neurophysiol. 101, 2725–2732.

Hinds, O., Polimeni, J.R., Rajendran, N., Balasubramanian, M., Amunts, K., Zilles, K., Schwartz, E.L., Fischl, B., Triantafyllou, C., 2009. Locating the functional and anatomical boundaries of human primary visual cortex. NeuroImage 46 (4), 915–922.

Holmes, A.P., Friston, K.J., 1998. Generalisability, random effects and population inference. NeuroImage 7, S754.

Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., Bernarding, J., Weiskopf, N., 2011. The impact of physiological noise correction on fMRI at 7T. NeuroImage 57 (1–4), 101–112.

Iwamura, Y., Tanaka, M., Sakamoto, M., Hikosaka, O., 1983. Converging patterns of finger representation and complex response properties of neurons in area 1 of the first somatosensory cortex of the conscious monkey. Exp. Brain Res. 51 (3), 327–337.

January, D., Trueswell, J.C., Thompson-Schill, S.L., 2009. Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. J. Cogn. Neurosci. 21 (12), 2334–2344.

Juch, H., Zimine, I., Seghier, M.L., Lazeyras, F., Fasel, J.H., 2005. Anatomical variability of the lateral front lobe surface: implication for intersubject variability in language neuroimaging. NeuroImage 24, 504–514.

Julian, J., Fedorenko, E., Webster, J., Kanwisher, N., 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. NeuroImage 60, 2357–2364.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17 (11), 4302–4311.

Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. Neuron 50, 963–974.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience — the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540.

Levitin, D.J., Menon, V., 2003. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. NeuroImage 20, 2142–2152.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.

Malach, R., Reppas, J.B., Benson, R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B.H., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. U. S. A. 92, 8135–8139.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage 19, 1233–1239.

Matelli, M., Luppino, G., Rizzolatti, G., 1991. Architecture of superior and mesial area 6 and the adjacent cingulate cortex in the macaque monkey. J. Comp. Neurol. 311 (4), 445–462.

McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9 (5), 605–610.

Miki, A., Raz, J., van Erp, T.G., Liu, C.S., 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.

Mikl, M., Marecek, R., Hlustik, P., Pavlicova, M., Drastich, A., Chlebus, P., Brazdil, M., Krupa, P., 2008. Effects of spatial smoothing on fMRI group inferences. Magn. Reson. Imaging 26, 490–503.

Miller, M.B., Van Horn, J.D., Wolford, G.L., Handy, T.C., Valsangkar-Smyth, M., Inati, S., Grafton, S., Gazzaniga, M.S., 2002. Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. J. Cogn. Neurosci. 14 (8), 1200–1214.

Nieto-Castañón, A., Ghosh, S.S., Tourville, J.A., Guenther, F.H., 2003. Region of interest based analysis of functional imaging data. NeuroImage 19 (4), 1303–1316.

Ono, M., Kubik, S., Abernathey, C.D., 1990. Atlas of the Cerebral Sulci. Thieme Verlag, Stuttgart.

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-Back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25 (1), 46–59.

Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S., 2007. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8, 91.

Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2011. Handbook of fMRI Data Analysis. Cambridge University Press.

Preibisch, C., Pilatus, U., Bunke, J., Hoogenraad, F., Zanella, F., Lanfermann, H., 2003. Functional MRI using sensitivity-encoded echo planar imaging (SENSE-EPI). NeuroImage 19 (2), 412–421.

Rajkowska, G., Goldman-Rakic, P.S., 1995a. Cytoarchitectonic definition of prefrontal areas in the normal human cortex: I. Quantitative criteria for distinguishing areas 9 and 46. Cereb. Cortex 4, 307–322.

Rajkowska, G., Goldman-Rakic, P.S., 1995b. Cytoarchitectonic definition of prefrontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46. Cereb. Cortex 4, 322–337.

Rozzi, S., Ferrari, P.F., Bonini, L., Rizzolatti, G., Fogassi, L., 2008. Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. Eur. J. Neurosci. 28 (8), 1569–1588.

Sabuncu, M.R., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V., 2010. Function-based inter-subject alignment of the cortical anatomy. Cereb. Cortex 20 (1), 130–140.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". NeuroImage 19 (4), 1835–1842.

Saxe, R., Powell, L., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. Psychol. Sci. 17 (8), 692–699.

Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. NeuroImage 30 (4), 1088–1096 (discussion 1097–9).

Scouten, A., Papademetris, X., Constable, R.T., 2006. Spatial resolution, signal-to-noise ratio, and smoothing in multi-subject functional MRI studies. NeuroImage 30 (3), 787–793.

Shuman, M., Kanwisher, N., 2004. Numerical magnitude in the human parietal lobe; tests of representational generality and domain specificity. Neuron 44 (3), 557–569.

Sigalovsky, I.S., Fischl, B., Melcher, J.R., 2006. Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. NeuroImage 32, 1524–1537.

Stringer, E.A., Chen, L.M., Friedman, R.M., Gatenby, C., Gore, J.C., 2011. Differentiation of somatosensory cortices by high-resolution fMRI at 7 T. NeuroImage 54, 1012–1020.

Swallow, K.M., Braver, T.S., Snyder, A.Z., Speer, N.K., Zachs, J.M., 2003. Reliability of functional localization using fMRI. NeuroImage 20, 1561–1577.

Thirion, B., Pinel, P., Tucholka, A., Roche, A., Ciuciu, P., Mangin, J.-F., Poline, J.-B., 2007. Structural analysis of fMRI data revisited: improving the sensitivity and reliability of fMRI group studies. IEEE Trans. Med. Imaging 26 (9), 1256–1269.

Thirion, B., Varoquaux, G., Poline, J.-B., 2010a. Accurate definition of brain regions position through the functional landmark approach. Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention.

Thirion, B., Tucholka, A., Poline, J.-B., 2010b. Parcellation schemes and statistical tests to detect active regions on the cortical surface. Proceedings of the 19th International Conference on Computational Statistics, pp. 565–572. Paris, France.

Tomaiuolo, F., MacDonald, J.D., Caramanos, Z., Posner, G., Chiavaras, M., Evans, A.C., Petrides, M., 1999. Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. Eur. J. Neurosci. 11, 3033–3046.

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

Tootell, R.B.H., Reppas, J.B., Dale, A.M., Look, R.B., Sereno, M.I., Brady, T.J., Rosen, B.R., 1995b. Functional MRI evidence for a visual motion after effect in human cortical area MT/V5. Nature 375, 139–141.

Tootell, R.B.H., Hadjikhani, N.K., Vanduffel, W., Liu, A.K., Mendola, J.D., Sereno, M.I., Dale, A.M., 1998. Functional analysis of primary visual cortex (V1) in humans. Proc. Natl. Acad. Sci. U. S. A. 95, 811–817.

Triantafyllou, C., Polimeni, J.R., Wald, L.L., 2011. Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. NeuroImage 55, 597–606.

Turin, G.L., 1960. An introduction to matched filters. IRE Trans. Inf. Theory 6 (3), 311–329.

Turner, R., Oros-Peusquens, A.M., Romanzetti, S., Zilles, K., Shah, N.J., 2008. Optimised in vivo visualisation of cortical structures in the human brain at 3 T using IR-TSE. Magn. Reson. Imaging 26, 935–942.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–289.

Vul, E., Kanwisher, N., 2010. Begging the question: the non-independence error in fMRI data analysis. Foundations and Philosophy for Neuroimaging. .

Walters, N.B., Egan, G.F., Kril, J.J., Kean, J., Waley, P., Jenkinson, M., Watson, D.G., 2003. *In vivo* identification of human cortical areas using high-resolution MRI: an approach to cerebral structure–function correlation. PNAS 100 (5), 2981–2986.

Walters, N.B., Eickhoff, S.B., Schleicher, A., Zilles, K., Amunts, K., Egan, G.F., Watson, J.D., 2007. Observer-independent analysis of high-resolution MR images of the human cerebral cortex: in vivo delineation of cortical areas. Hum. Brain Mapp. 28 (1), 1–8.

White, T., O'Leary, D., Magnotta, V., Arndt, S., Flaum, M., Andreasen, N.C., 2001. Anatomic and functional variability: the effects of filter size in group fMRI data analysis. NeuroImage 13, 577–588.

Wohlschläger, A.M., Specht, K., Lie, C., Mohlberg, H., Wohlschläger, A., Bente, K., Pietrzyk, U., Stöcker, T., Zilles, K., Amunts, K., Fink, G.R., 2005. Linking retinotopic fMRI mapping and anatomical probability maps of human occipital areas V1 and V2. NeuroImage 26 (1), 73–82.

Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Holt, D., Amunts, K., Zilles, K., Golland, P., Fischl, B., 2010. Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex. IEEE Trans. Med. Imaging 29 (7), 1424–1441.

Zilles, K., Schleicher, A., Langemann, C., Amunts, K., Morosan, P., Palomero-Gallagher, N., Mohlberg, H., Bürgel, U., Steinmetz, H., Schlaug, G., Roland, P.E., 1997. Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. Hum. Brain Mapp. 5 (4), 218–221.