

# **Divide and Conquer: A Defense of Functional Localisers**

Rebecca Saxe<sup>1,2</sup>, Matthew Brett<sup>3</sup>, & Nancy Kanwisher<sup>2, 4, 5</sup>

<sup>1</sup>Harvard Society of Fellows

<sup>2</sup>MIT Department of Brain & Cognitive Sciences

<sup>3</sup>MRC Cognition and Brain Sciences Unit, Cambridge, UK

<sup>4</sup>McGovern Institute for Brain Research, MIT

<sup>5</sup>Martinos Center for Biomedical Imaging

submitted to *NeuroImage*, October, 2005

please do not quote without permission

***Address correspondence to:***

Rebecca Saxe  
Department of Brain and Cognitive Science  
MIT 46-4019  
Cambridge MA 02139

saxe@mit.edu

## ***Abstract***

Numerous functionally distinct regions of cortex (e.g., V1, MT, the fusiform face area) can be easily identified in any normal human subject in just a few minutes of fMRI scanning. However the locations of these regions vary across subjects. Investigations of these regions have therefore often used a functional region of interest (fROI) approach in which the region is first identified functionally in each subject individually, before subsequent scans in the same subjects test specific hypotheses concerning that region. This fROI method, which resembled long-established practice in visual neurophysiology, has methodological, statistical, and theoretical advantages over standard alternatives (such as whole-brain analyses of group data): i) because functional properties are more consistently and robustly associated with fROIs than with locations in stereotaxic space, functional hypotheses concerning fROIs are often the most straightforward to frame, motivate, and test, ii) because hypotheses are tested in only a handful of fROIs (instead of in tens of thousands of voxels), advance specification of fROIs provides a massive increase in statistical power over whole-brain analyses, and iii) some fROIs may serve as candidate distinct components of the mind/brain worth investigation as such. Of course fROIs can be productively used in conjunction with other complementary methods. Here we explain the motivation for and advantages of the fROI approach, and we rebut the criticism of this method offered by Friston et al., (in press).

## ***Introduction***

Near Cape Hatteras, North Carolina, millions of cubic meters of warm salt water leave the continental shelf each second, and head north-east across the Atlantic towards Europe. This is the Gulf Stream, the most intensively studied ocean current in the world. One challenge for Gulf Stream oceanography, though, is that neither the position nor the volume of the Stream remains constant from month to month, or from year to year. The location where the current leaves the coastline can vary by as much as a hundred kilometres, north or south (Mariano et al., 2002), and the Stream is particularly hard to study as it approaches Europe and splits into the smaller and more variable branches. As a result of this variability, scientists wishing to take a measurement of Gulf Stream water, or mariners looking for an east-wards bump, cannot simply calculate by longitude and latitude. Instead, they must first determine the path of the Gulf Stream at that moment, by measuring the water temperature or looking at satellite photographs, and then navigate accordingly.

Much as the precise path of the Gulf Stream (measured in longitude and latitude) varies across seasons and years, the precise positions of functionally distinct regions of the brain, such as primary visual cortex (measured in Talairach or MNI coordinates), vary across individuals. So, just as oceanographers must first identify the actual path of the Gulf Stream before using water measurements to make inferences about circulation in the North Atlantic, cognitive neuroscientists often use “localizer” contrasts to functionally identify regions in the brain of a given individual (called “function regions of interest” or fROIs) before using measurements of the activity in those regions to test hypotheses

about the distribution of cognitive functions across the cortex. Here we first sketch the advantages to using functionally defined regions of interest, and then we consider the specific concerns about this practice raised by Friston et al. (in press).

### ***1. Why use fROIs?***

The goal of science is not simply to collect a list of independent facts about the world, but to organize those facts into general patterns, to test hypotheses about those patterns, and to explain them in terms of broader theories. The importance of fROIs concerns all three stages. We consider each in turn.

#### **A. Knowing where we are**

Brain imaging relates function to locations in the brain. Because cognitive neuroscience is not about specific individuals but about people in general, we need some way to specify brain locations that will generalize across individuals. The problem is that the shape of each person's brain is unique, just like the shape of her face. So cognitive neuroscientists face a serious challenge: What counts as the same place in two different brains?

One answer is to consider purely anatomical markers to align similar structures of each subject's brain. For brain regions that can be identified anatomically in vivo, such as the basal ganglia, amygdala or hippocampus, we have an excellent way to register these regions across individuals: locate these structures on anatomical scans within each subject individually, and collect that subject's functional data from the corresponding location. Ongoing efforts to identify cytoarchitectonic markers in individual brains may eventually enable us to extend these methods to the cortex (Amunts et al., 2004, Eickhoff

et al., 2005, Schuchard et al., 2003). Promising results have also been obtained in studies using connectivity patterns to define functional areas (Johansen-Berg et al., 2004). However, for most cortical regions, we currently lack clear anatomical or connectivity markers to define specific cortical areas.

As a result, the most prevalent method of anatomical matching, since the early days of brain imaging, has been to use a standard Cartesian co-ordinate system such as the Talairach & Tournoux stereotaxic system (Talairach et al., 1988). In this system, each individual's data are aligned to a "standard" brain, usually using an automated algorithm to match the subject's anatomical scan to an anatomical template. Increasingly, the co-ordinates reported in the results of individual studies are also used as input to meta-analyses (Duncan, 2000, Paus, 1996).

Nonetheless, the practice of normalising individual brains to a "standard" brain space has important shortcomings. Well-known brain areas (defined functionally, or cyto-architecturally) do not land consistently in the same place in any set of standard anatomical co-ordinates, across individuals. One illustration of this problem is seen in the case of primary visual cortex (area V1), which lies consistently along the calcarine sulcus in all subjects, but which does not land in the same location across all subjects in stereotaxic coordinates (see Figure 1). Another illustration of the problem is seen in Figure 2d, where the selectivity of the FFA is stronger when it is functionally defined individually within each subject than when it is defined from a standard group analysis, because the necessarily imperfect registration across individuals entailed in the group analysis precludes selection of all and only the FFA in each individual, thus blurring the

distinctive functional profile of each region with that of its neighbors (see also Ozcan et al 2005).

More sophisticated strategies for registering individual brains together use reconstructed cortical surfaces that respect sulcal locations. These systems provide better alignment across subjects for retinotopic visual areas (Fischl et al., 1999), but they don't do much better than Talairach co-ordinates for category-selective regions in the temporal lobe (Spiridon et al., 2005). Moreover, post-mortem histology shows that sulcal borders do not reliably coincide with other, well-established anatomical divisions based on cytoarchitecture (Amunts et al., 1999, Amunts et al., 2001).

Thus neither standard stereotaxic registration methods, nor more sophisticated coordinate systems that respect sulcal landmarks, are likely to bring distinct functional regions perfectly into register across subjects. As a result, nearby brain regions with different functional profiles will be averaged together across individuals, reducing both the resolution and sensitivity of subsequent functional analyses (Swallow et al., 2003; see also Figure 2).

The alternative solution is for cognitive neuroscientists to use consistent patterns of functional response profiles to constrain the identification of the “same” brain region across individuals— that is, to use *functional landmarks*. The functional landmark approach requires using the combination of sulcal/gyral divisions (e.g. the fusiform gyrus) and robust functional profiles (e.g. strong preference for faces versus objects) in individual subjects to identify the landmark region, which can then be used to orient research in the whole patch of neighbouring cortex. This practice is already near-universal among cognitive scientists studying, for example, early visual cortex, where the

borders of each subject's retinotopic visual areas are first identified with functional scanning before the main experimental data are collected.

Note that the use of functional landmarks, as described in this section, need not hinge on the reification of fROIs – either as the units of hypotheses about regional functions, or as fundamental theoretical entities in cognitive neuroscience. Functional profiles may be used simply as reliable guides to the “same” location from one brain to another – whether the researcher is interested in the landmark region itself, or rather in a region that is reliably located with respect to the landmark. Once we can identify the “same place” in different brains, we are in a position to combine data across subjects, studies, and labs. Often, though, researchers use fROIs in a theoretically deeper way, to gain a richer understanding of the response profile of each region, in order to test hypotheses about the neural basis of cognitive functions.

## **B. Testing hypotheses**

For the most part, cognitive neuroscientists seek to test hypotheses about the cognitive functions of particular regions of the brain (for an exception see Haxby et al., 2001). As we will argue in this section, this work can proceed most effectively (and can most easily avoid a number of common pitfalls) when researchers specify their hypotheses in advance – including both the profile of functional response that would support the hypothesis, and the specific brain location(s) in which the functional profile should be tested. Although functionally-defined regions of interest are just one of many ways to specify a brain location, they often serve as the regions about which specific functional hypotheses can most sensibly be framed.

First, specifying *in advance* the region(s) in which a hypothesis will be tested increases statistical power by reducing the search space from tens of thousands of voxels to just a handful of ROIs. Greater power allows researchers to investigate with confidence the relatively subtle aspects of a region's response profile, such as those revealed in fMRI adaptation studies (Grill-Spector et al., 2001, Henson et al., 2000); ; see also Figure 3). By contrast, traditional whole-brain analyses produce an explosion of multiple-comparisons, requiring powerful corrections to control false positives (Nichols et al., 2003). Many researchers appear to find the necessary corrections too draconian and it is still common for neuroimaging papers to report uncorrected statistics, which do not provide valid control of false positives. For example, in the last five years, the Brede database of published neuroimaging papers (Nielsen, 2003) records that half (49%) of the 1705 articles that report any brain location p value, *only* report uncorrected p values.

Blunted by the loss of statistical power entailed in multiple comparisons, whole brain analyses are particularly poor tools for establishing the *absence* of an effect of interest in a given region, a real problem because the lack of a difference between two tasks may be critical for testing a cognitive theory (e.g. Jiang, Saxe and Kanwisher, 2004). Pre-defined regions of interest provide a practical, sensitive, and statistically rigorous solution to each this challenge.

Second, using a pre-defined region of interest allows the researcher to test her hypothesis in a statistically unbiased data set. A common alternative practice that we call "voxel-sniffing" proceeds as follows: (1) hypothesise that region X will be recruited differentially for Task A versus Task B, (2) search for the peak voxel in the neighbourhood of X that shows the predicted differential recruitment, and then (3)

display the response of this voxel showing differential recruitment for Task A versus Task B. Since the very same aspect of the data is used both to find the region in which the hypothesis is tested, and show the effect, the resulting data display is biased toward a false positive. Noise or measurement error that contributed to the measured response of the region infects both steps: finding the region and measuring its response. By contrast, when the precise voxels that will make up the ROI are determined in advance, any spurious feature of the data in the localising step of the analysis cannot contribute to a positive result in the independent data used to test hypotheses about that region (see also Liou et al 2005).

Finally, using a pre-established functional definition of the region at stake in a hypothesis helps researchers to avoid spurious claims of contradictions or convergences across subjects and across experiments. The interpretation of one set of neuroimaging data is almost always informed by (and responsive to) a wide previous literature; the challenge is to determine whether the current result reflects recruitment of the very same brain region as in previous reports, or of a nearby region blurred together by averaging.

All of these benefits of pre-defined regions of interest can be garnered using anatomical markers, stereotaxic co-ordinates or functionally-defined regions of interest. Which of these to choose is an empirical/pragmatic question: Which kind of ROI is best suited to the experimental design and the hypothesis being tested? Anatomical ROIs, defined either by stereotaxic co-ordinates or by sulcal and gyral landmarks, may be the only way to define regions of interest in advance for comparisons between healthy subjects and patient populations, to guide hypotheses about brain damage based on functional imaging results in healthy subjects, or to make predictions about healthy brains

based on analyses of selective deficits following lesions (Heberlein et al., 2005, James et al., 2003). Stereotaxic co-ordinates from prior studies may also be used to define regions of interest when researchers need to be especially cautious to avoid a false negative (Jiang et al., 2004, Shuman, 2004).

In many cases, though, a broad consensus now exists in the literature about the locations and properties of particular regions implicated in a contrast of interest, so a pragmatic and efficient way to relate results across studies, and to formulate and test hypotheses about the role of brain regions identified in previous studies, is to use fROIs.

Note that the use of fROIs to pre-specify the region in which a hypothesis will be tested does not require a commitment to the ontological status of that region as a fundamental or homogenous unit. Instead of assuming such a commitment, fROIs can be used to *empirically test* the question of whether a region's response profile reflects a coherent, theoretically interesting functional unit, or not. Some robust and reliable fROIs that can be used as functional landmarks, and that constrain statistical analyses, turn out to have complex and disjunctive, or overly general, response profiles; for example, the lateral frontal and superior parietal regions studied intensively with fROIs by Jiang and collaborators are implicated across a wide range of task contexts (Jiang et al., 2003) and so the authors explicitly argued against the interpretation of these fROIS as theoretical units (Duncan, 2000; another example may be the lateral occipital complex, Kanwisher et al 1997, Malach et al 1995). In other cases, though, fROI analyses do reveal stable, coherent, and theoretically interesting regional response profiles that may help to constrain cognitive and computational theories of the mind. It is to these cases, and these kinds of arguments, that we now turn.

### **C Discovering fundamental Natural Phenomena.**

The third and most theoretically committed reason to use fROI analyses depends on the researcher's philosophy of cognitive neuroscience – that is, on a notion of what kind of hypotheses or theories cognitive neuroscientists should be building and testing with neuroimaging data. We believe that the functional characterisation of specific brain regions is a critical part of this enterprise.

The central tenet of modern cognitive neuroscience is that “different brain areas perform different information processing operations (often referred to as computations) and that an explanation of a cognitive performance involves both decomposing an overall task into component information processing activities and determining what brain area performs each” component (Bechtel, 2002). At its best, cognitive neuroscience can provide neural evidence to distinguish between functional psychological decompositions (or task analyses) of higher cognitive functions. That is, the observed division of labour across brain regions can help decide among alternative psychological theories of a cognitive capacity if the psychological components of the task are reflected in distinct and reliably localised neural components – i.e. patches of cortex. In some cases, fROIs seem to pick out cortical patches of just this kind (Xu, 2005; Yovel et al., 2004).

Note that such an anatomical constraint need not apply to all information processing operations carried out by the brain. First, the question of whether the cortex actually contains *any* functionally distinct components has been hotly debated for much of the history of neuroscience. The influential early pioneer of lesion studies Pierre Flourens denied the existence of any localisation of function, claiming instead that all areas of the cortex served all the same functions, so that any part of the cortex could

accommodate for the loss of any other region. This debate continues, in changing form (Haxby et al., 2001, Spiridon et al., 2002; Section 2B here describes a modern version of this controversy). Second, if some functions *are* localised in the brain, similar computations could in principle be realised differently in each individual brain (called ‘multiple realisability’ in philosophy). Third, in cases when functional properties do align with some anatomical properties consistently across subjects, it is still an empirical question which anatomical marker(s) will constitute the most reliable guide to function: sulcal and gyral landmarks, cytoarchitectonic divisions, and/or patterns of functional connectivity? Finally, even if psychological and neural components of a cognitive capacity are localised and aligned in some regions, the question of how many such neural components/entities there are in the brain (and in the mind) remains open.

For all of these reasons, the status of any given functionally-defined ROI is not something to be assumed from the outset, but is rather contingent on empirical outcomes. Even the observation of robust and distinctive regularities in the response profile of a patch of cortex across individuals may not reflect a neural entity in this theoretical sense, if those regularities do not reflect any single or stable elementary computation. So for each ROI, researchers must assess whether the functional properties of the region are not only stable, but also theoretically coherent and meaningful. Although we make no guesses about how often theoretically-important functions localised to patches of cortex will occur in nature, we believe this standard has been upheld by at least a few widely-used functionally-defined ROIs including MT/V5 (Tootell et al., 1995; the FFA Kanwisher et al., 1997; see Figure 2); the region of the TPJ implicated in understanding

another person's beliefs (Saxe et al., in press); and the AIP which is implicated in grasping (Culham, 2003).

In sum, fROIs are useful for specifying brain locations across subjects, for testing hypotheses concerning the function of specific brain regions, and—occasionally—for investigating candidate separable components of the mind. The use of fROIs is uncontroversial and indeed virtually required in any study of visual cortex involving retinotopic visual areas or MT. We see no reason why the widely accepted benefits of fROIs in research on these regions would not apply equally to other cortical regions. Nonetheless, Friston and colleagues (Friston et al., In press) label the practice a “missed opportunity”, as well as a “potentially restrictive” and even “retrograde” development in experimental design. Next we consider and rebut their main critiques of the fROI method.

## ***2 Common and uncommon misconceptions about fROI: response to Friston et al***

### **A. Telescopic vision**

The most common concern about the use of fROI analyses is that they will obscure the researchers' view of the bigger picture. A small but significant effect observed in the average response of an ROI might reflect the specific engagement of the region of interest, but it could equally reflect (1) a small but consistent effect that is actually occurring across many regions of the brain, or (2) the fringe or tail of a much bigger activation in the same contrast in a neighbouring region, or (3) a different region, sending feedback connections to the region of interest. Friston et al (this issue) express this common concern by arguing that “anatomical specificity cannot be addressed by a

single fROI.” That is, an fROI-constrained analysis could lead the scientist to miss the forest for a tree.

The solution, of course, is to combine fROI analyses with each other, and/or with voxel-based whole brain analyses. None of the advantages of independent, pre-constrained functional ROIs precludes comparing multiple ROIs in the same session, or comparing the results of fROI analyses with whole brain analyses to look for robust activity outside of the ROI. Many researchers use multiple fROIs in combination, for just this reason (see Figure 3). So the worry that the advantages of fROI analyses will lead researchers to focus exclusively on a single ROI seems unsubstantiated by the actual practice in the field. By adding fROIs to our repertoire, cognitive neuroscientists can rigorously study both the forest and the tree.

### **B. Assumptions of homogeneity**

The second basic concern about fROI analyses is that a focus on the response of a pre-defined region will lead researchers into the dangerous – or “untenable” (Friston et al, this issue)– assumption that the neurones that compose the fROI are all homogeneous and equally well described by the average. If fROIs did lead researchers to claim the absolute homogeneity of the neurones under investigation, this would obviously be unfortunate, not to say silly. Once again, though, the solution for fROI analyses is simple: recognise the likelihood of finer grained structure, and treat the question of homogeneity as an empirical one. Also note that this advice applies equally to cognitive neuroscientists who don’t use fROIs, but rely on the response of single voxels, each of which reflects the average activity of hundreds of thousands of individual neurones.

Inhomogeneities, or finer grained structure, within the response of a region or a voxel may arise from any of at least three kinds of sources: (1) sub-groups of neurones doing similar computations over different regions of stimulus ‘space’, as seen for example in V1 (retinotopy) and S1 (somatotopy); or (2) interleaved or tightly packed groups of neurones performing distinct, even unrelated, functions, or (3) sub-groups of neurones performing hierarchically-arranged parts of a single computation, which are too close together to resolve with current technology, especially based on blood-flow.

Researchers using fROIs have been at the forefront of empirical investigations of each of these possibilities for the organisation of responses within regions. For example, several recent studies have used fROIs to explicitly test the information that is contained in the profile of response across the voxels within that fROI. Using classification methods (Haxby et al., 2001, Spiridon et al., 2002) found that although the pattern of response across the voxels within the FFA enabled discrimination of faces from nonfaces, it did not allow discrimination of one nonface category (e.g., bottles) from another nonface category (e.g., shoes), a result that reinforces the selectivity of this region for face processing. Far from assuming homogeneity, this method exploits and directly investigates fine-grained differences in the response profile of different voxels within a fROI (see also Kamitani 2005.) Note further that the identification of fROIs for early visual areas using retinotopy depends precisely on the heterogeneous response profiles of different voxels within an area to stimuli presented to different retinal locations.

In all, practice in the field does not support the concern that using fROIs will lead researchers to assume, rather than to *test*, the finer grain structure of neurones encompassed by a single ROI. But finally, we must concede that all fMRI investigations

sacrifice some degree of fine-grained precision in favour of a broader summary view of neural responses, whether by averaging across a group of voxels, or across even just the thousands of neurones that drive a single voxel. Once cognitive neuroscientists acknowledge this trade-off, we can continue investigate aspects of cognitive function that *are* shared across, and selectively recruit, whole groups of neurones.

### **C. Factorial Designs versus Independent Localisers**

Friston et al., (Friston et al., In press) repeatedly claim that “people who like localisers should like factorial designs even more.” Well, we are people who like localisers, and we are ambivalent about factorial designs. Factorial designs have important advantages, but when used to replace (rather than to complement) independent localizer contrasts, they have significant disadvantages as well.

In a full factorial design, researchers compare the effects of two or more stimulus/task manipulations on the resulting neural response of a region or group of voxels, by presenting the subject with all possible combinations of those stimuli and/or tasks. In the example discussed by Friston et al.,(Friston et al., in press, Kourtzi et al., 2001) compared the effects of stimulus shape (same versus different) and depth (same versus different) on the response of the LOC fROI. Since the authors presented subjects with all four possible combinations of shapes and depths, the experiment used a “full factorial” design. Full factorial designs allow the researcher to simultaneously and rigorously measure both main effects of the factors of interest, and any interactions between these factors. Undeniably, such an experimental design is the best way to test hypotheses about the relative effects of multiple factors on a regional response, especially

when the factors are separable, and possible interactions between the factors are of theoretical interest.

Unfortunately, Friston et al's (Friston et al., in press) argument for replacing localiser contrasts with full factorial design confuses three independent issues: (1) the relatively trivial question of whether the localiser contrasts occur within the same runs, or in separate but interleaved runs within the same scanning session, (2) whether full factorial designs are always the best experimental design for testing hypotheses in cognitive neuroscience; and (3) whether to use the main effects within a full factorial design in place of independent localiser contrasts to identify regions of interest. We consider these three issues in turn.

First, Friston et al.,(Friston et al., in press) have emphasized the fact that fROI experiments often place the localizer experiment in a separate set of runs from the main conditions of interest. Of course this practice is irrelevant for the logic of fROIs, as Friston et al point out in their paper. Whether to put the localiser contrast in the very same runs as the main experiment, or in separate runs, interleaved or counter-balanced with the main experiment (Liou et al 2005) is a just a case-specific question of experimental design. Empirical evidence suggests that the position and selectivity of fROIs is strikingly consistent both within and between scan sessions (Peelen, 2005).

The second point is that full factorial designs are not the only elegant and rigorous experimental designs available to cognitive neuroscientists, and are not perfectly suited to testing all kinds of hypotheses. Other sophisticated experimental designs use parametric gradations of a single factor, or sort events by behavioural response on a trial-by-trial

basis, or use classification methods to investigate the information that is represented in a given area (Cox et al., 2003, Haxby et al., 2001, Kamitani 2005) . This is only to say that the logic of fROIs is compatible with any well-balanced experimental design in the ‘main’ experiment; there is no special relationship between full factorial designs and fROIs, unless the factorial design is being used to replace fROIs as a way to constrain analyses to regions of interest.

The disadvantages of full factorial designs arise only in this final case, when a factorial design is proposed to replace fROIs. Within a full factorial design, orthogonal main effects and interactions are independent of the main effect used to define the regions of interest, and so conform to the logic of fROI analyses. In this sense, a factorial design could simply be used as *one way* of doing fROI analyses. The question then becomes: are factorial designs the best way?

We think not. In a factorial design, though the test of an interaction will be independent from the ROI-definition, the test of the main effect will be biased, since the very same data used to find the region of interest is then used to estimate the magnitude of the main effect. Researchers using fROI analyses therefore use an independent data set – collected during a separate set of conditions – to define the region of interest, and constrain the hypothesis space, from the data that will be used to estimate the magnitude of main effects and interactions in the ‘main’ experiment.

To illustrate by example, Epstein et al. (1998) used a whole-brain analysis to show that a specific area of parahippocampal cortex – the Parahippocampal Place Area (PPA) – responded much more strongly when subjects viewed scenes (including furnished rooms) than when they viewed faces or single objects. They then used

thresholded data from this experiment to define the voxels in the PPA, and performed a second experiment to investigate whether average PPA activation was better explained by the multiple objects in the furnished rooms, or by the spatial layout of the rooms. They therefore included the following conditions: furnished rooms; rooms with the all objects removed, preserving only the spatial layout; and displays of the objects from the furnished rooms rearranged randomly. The inclusion of a replication of the main effect in the second experiment enabled them to show not only that spatial layout alone (i.e., empty rooms) produced a very strong response in the PPA, but further that spatial layout appeared to be alone sufficient to explain the large response to the furnished rooms, because the response to furnished and empty rooms did not differ.

Finally, Friston et al. (in press) imply that there is no cost to using a full-factorial design and then testing the higher level interactions, relative to a separate localiser (or “no loss of statistical efficiency”). On the contrary, simply multiplying the number of conditions in the experiment may produce many conditions that are of no interest to the experimenter, and simultaneously decrease the number of observations that can be conducted for the critical conditions. The result is a loss of power where it counts, and therefore reduced statistical efficiency.

The basic message of this section is: people who like factorial designs should like the combination of factorial designs with an independent localiser contrast even better.

#### **D. Averaging signal in the fROI and other summary measures**

The fROI approach involves first identifying a candidate functional region, and then calculating some overall summary measure of response in that region. This

summary measure is used for two reasons. The first is to decrease the influence of noise that varies between voxels. Of course this is also one of the main motivations for smoothing in traditional whole-brain analyses. The second use of summary measures is to make inferences about the response of the region as a whole, rather than particular voxels within the region. Importantly, the choice of summary measure is not intrinsic to the logic of fROI analyses. The same logical, statistical and theoretical advantages of fROI analyses apply to all these summary measures.

In general, if we know the size and shape of a particular functional area, then the best (most efficient) representation of the signal from this area will be the mean from the voxels within the area<sup>1</sup>. There are three summary measures in frequent use, which we will call *threshold-average*, *peak-smoothed*, and *first eigenvariate*. All three take a weighted mean across voxels within a brain area, but differ in their method for using the localising contrast to estimate how much each voxel contributes to the functional area.

The threshold-average method is the most common approach to region definition for fROI analyses: define the region as the set of all contiguous voxels passing a pre-determined threshold, and then average the response across all voxels in the region. The threshold should be high enough to avoid noise while still detecting signal. The advantage of this approach is that it can adapt flexibly to detect regions with almost any size or shape. Friston et al. (in press) appear to take particular issue with this practice, and strongly prefer the two common alternative methods – the peak-smoothed and first eigenvariate measures.

---

<sup>1</sup> Here we do not consider the effect of spatial autocorrelation, as this is rarely addressed in standard analyses, and applies to all three summary measures.

Peak-smoothed averaging uses smoothed data, and takes the time-course of the voxel showing maximum activity in the localizer. In this case the underlying region response is assumed to be a Gaussian shape with the same size as the smoothing kernel. The choice between the threshold-average and peak-smoothed methods will depend on your model of the underlying regional response. Clearly, if the region of interest is in fact a Gaussian shape of the same size as the smoothing kernel, there will be very little difference between the methods. Indeed, the results of average thresholded and peak smoothed analyses are often highly convergent (Figure 3) (Arthurs et al., 2003). If the region response does not fit this size and shape, though, the peak-smoothed model will be a less efficient model of the region's overall response; for example the PPA is elongated (Epstein et al., 1998), and will therefore not be well matched by a simple Gaussian. Also, because smoothing is usually in three dimensions, rather than on the cortical surface, using smoothing for regions with complex cortical folding may lead to averaging with nearby sulcal banks, grey matter and CSF.

There is a greater theoretical difference between average thresholded and the first eigenvariate measure, which is the second approach suggested by Friston and colleagues (Friston et al., in press). The first eigenvariate results from performing a principal components analysis (PCA) on the region data, where the time points are the observations and the voxels are the variables. The PCA returns a series of components ordered by the proportion of variance each component explains; each component is associated with a vector of weights (one value per voxel) reflecting the contribution of each voxel to that component. We can reconstruct signal proportional to the first eigenvariate by multiplying the (time by voxel) region data by the voxel weights for the first component.

In this case the model of the region shape and size is given by the voxel weights, and the time course is therefore a weighted average. The first eigenvariate reflects the time course of the main component that contributes to a region's response.

In essence the voxel weights from the PCA can be thought of as an attempt to extract the shape and size of the underlying response from within the region activated by the localizer contrast. If the regional response is homogenous, the average-thresholded and first eigenvariate measures will produce very similar results. If the area is not homogenous, we need to treat the first eigenvariate measure with care; the researcher must analyze the spatial weights of the component in order to interpret the results. Because of the spatial weighting, the eigenvariate may refer only to a small portion of the localized region, rather than representing the region's overall response. There is also no guarantee that the eigenvariate will pick out the same part of the localized region for different subjects.

In sum, little practical difference is expected between the results of the different methods of summarizing the response within an fROI. Where differences arise, the standard approach of averaging across voxels above threshold has the advantages of being easy to interpret and appropriate for regions of different shapes and sizes.

#### **E. Why name fROIs?**

Even granting the logical and methodological advantages of fROI analyses, some researchers continue to resist the practice of giving those fROIs *names*. Some worry that reifying discrete regions distorts the actual continuity of the measured BOLD response; for others, names should be eschewed altogether because of the danger of mis-naming

regions before their functions are completely understood. There is merit to each of these concerns, but we believe that the pragmatic advantages of naming fROIs, when such names are treated with care and flexibility, outweigh the disadvantages.

The “activation map” images that commonly accompany brain imaging papers can be misleading to inexperienced readers, by seeming to suggest that the boundaries between “activated” and “un-activated” patches of cortex are unambiguous and sharp. Instead, as most researchers are aware, the apparent sharp boundaries are subject to the choice of threshold applied to the statistical tests that generate the image. What, then, justifies dividing the cortex into regions with boundaries based on this fuzzy, mutable measure of functional profile?

It is an empirical question whether the slopes of a given candidate fROI are shallow or sharp. For some regions, such as the FFA, these boundaries in the measured response are impressively sharp (Spiridon et al., 2005). In other cases it may still be reasonable to assume that the “real” boundaries in the underlying tissue are sharper than the edges in the measured signal, which are distorted and flattened by blood diffusion, partial-voluming, and/or smoothing during the statistical analyses. These regions will have to await technological advances to establish the true grade of their sides.

The broader point, though, is that cognitive neuroscientists need not be shy about reifying fROIs merely because the precise boundary of the region may be hard to determine. Imprecise boundaries are a feature of many scientifically respectable objects, including ocean currents like the Gulf Stream, geographical features like Mt Fuji (see Figure 2), and human body parts like knees and elbows. Instead, both empirical investigation and emerging conventions and consensus are needed to establish the

reference of fROI names: to the region surrounding the peak, and/or to the whole area above a pre-determined altitude.

The second concern is that, in the absence of full knowledge, functional names may be mis-assigned, leading to long-term confusion. Cautionary tales of just such confusion come from the anatomical and functional names assigned to regions of the avian brain, a confusion that generated many papers, much frustration, and eventually whole conferences dedicated to nomenclature ([avianbrain.org/nomen](http://avianbrain.org/nomen)).

So it is important that researchers recognise that a name is a tool, not a conclusion. Naming a region should stimulate empirical research about that region's function, and not inhibit such research. Nevertheless the advantages of being able to talk about the results of fROI analyses outweigh the danger. A name allows researchers to precisely state both the hypotheses and the results of fROI analyses, without constant use of long ellipses (e.g. "the region of the fusiform gyrus recruited significantly more during..."). Also, note that the most common alternative to fROI names - the use of simple sulcal-gyral landmarks (e.g. "the fusiform gyrus", the "temporo-parietal junction") - is both imprecise and actively misleading, since the authors' claims do not apply to the whole anatomical region in question, but only to the subset of that region selected by the fROI.

Of course, fROI names should be used with precision and care. But even if in some cases, the conventional fROI names do not capture the full truth about the function of the patch of cortex that the fROI reflects, we propose that the most powerful way for one scientist to describe this result to another is a sentence that begins with the fROI's name.

## **Conclusion**

In sum, there is much to be gained and little to be lost from the use of fROI analyses, in conjunction with other analysis methods, in many imaging studies. FROI analyses are particularly useful for studies investigating phenomena for which distinctive, well-described, and anatomically restricted functional regions have been described that can be easily and straightforwardly identified in each subject in a few minutes of functional scanning, and where a broad consensus exists in the literature on how these regions are to be functionally localized.

Friston et al., In press note that “the purpose of [their] commentary is to provide a reference for people who do not want to use functional localizers and have to defend themselves against the contrary attitudes of reviewers”, and they cite in their appendix specific reviewer comments such as this one: “I am left wondering about the anatomical relationship between the fusiform region reported here, and the well-known fusiform face area (FFA). Why didn’t the authors use an independent functional localizer for the FFA?” We don’t see an answer to this question in the commentary by Friston et al (Friston et al., in press) and we still think that many studies would benefit from the use of an fROI.

## Acknowledgements

We thank Chris Baker, Giulio Boccaletti, Jim DiCarlo, Paul Downing, Jon Driver, Evelina Fedorenko, Peter Godfrey-Smith, Tania Lombrozo, Mona Spiridon, and Galit Yovel for comments and suggestions.

## Figure Legends

**Figure 1. Aligning V1.** A sagittal slice of the MNI template brain ( $x=+12$ ) overlaid with the probability map for the location of V1, determined posthumously from the cytoarchitecture of ten individuals. Colour scale shows number of brains in which V1 occupies each position, from Blue = 1 subject to Red = 10 subjects. From Wohlschlagel, 2005.

**Figure 2. fROI Case Study: the right FFA.** A small section of one functional slice in one subject showing on the vertical axis the spatial profile of percent signal increase (compared to a fixation baseline) for faces in (a) and for objects in (b). Note that these displays avoid the problem of arbitrary thresholds by showing the mean response at each point in the slice; for comparable data presented on the cortical surface see Figure 7 in Spiridon et al., 2005. Several of the issues concerning fROIs discussed in the text are illustrated here. **1. Why Name fROIs?** Despite the absence of sharp boundaries and the difficulty of defining its precise border, few would quarrel with the use of the name “Mt. Fuji” to refer to the prominent bump in (c). For similar reasons we find it useful to refer to the prominent bump in the response profile to faces shown in (a) as the “fusiform face area” or FFA. **2. Why use fROIs tailored individually to each subject rather than to the group average?** Because the FFA varies in anatomical position across individuals in stereotaxic space, an FFA defined on the basis of group- aligned data will necessarily include regions outside the FFA in some or all subjects, blurring the distinctive functional profile of the FFA and weakening its stimulus selectivity (“Group ROIs” at right), compared to the case in which the FFA is defined for each subject individually based on

that subject's own data ("Individual ROIs" at left). In both cases data were unsmoothed, and the data used to define the ROI were separate from those used to estimate the beta weights of the response of each condition in that ROI. The analysis x category interaction is highly significant, due to a weaker response to faces in the group analysis, relative to the individual analysis, and the reverse trend for all of the non-face categories. We thank Paul Downing and Annie Chan for this analysis.

**Figure 3. fROI Case Study: the right EBA.** Lateral extrastriate cortex can be divided into distinct but neighbouring fROIs, including the Extrastriate Body Area (Bodies > Objects), MT (Moving > Stationary), and the lateral occipital complex (Objects > Textures). All three regions show a robust response to stationary photographs of human bodies but only the right EBA shows a significant enhancement when viewing body-parts displayed from an allocentric, rather than an ego-centric, perspective. These results illustrate the strengths of fROI analyses that we have described. **1. Small effects:** the small but significant preference for allocentric body parts in the rEBA fROI cannot be detected in a whole brain analysis of twelve subjects. **2. Anatomical specificity:** combining multiple fROI analyses allowed the authors to show that the enhanced response to allocentric images was specific to the rEBA. Interaction tests showed that the profile in the rEBA was significantly different from that in right LO and MT. **3. Cross-lab comparisons:** Chan et al. (2004) in Wales and Saxe et al. (2005) in the USA, using different stimuli and different scanners, both found evidence for the same enhancement for allocentric images in the functionally-defined rEBA. The result was replicated even though Saxe, Jamal and Powell reported the average response for all voxels in the fROI,

while Chan, Peelen and Downing reported the average response in a small cube around the peak. (a) 3 slices in a single subject, showing the right EBA (red), LO (blue), and MT (green). (b) Stimuli from Saxe et al, and Chan et al. (c.) Percent signal change to egocentric and allocentric images of body parts in three neighbouring fROIs: right LO, MT, and EBA (Saxe et al). The fourth column shows the replication of the same result from Chan et al (2004).

## Bibliography

- Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K. 1999. Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol* 412(2):319-341.
- Amunts K, Weiss PH, Mohlberg H, Pieperhoff P, Eickhoff S, Gurd JM, Marshall JC, Shah NJ, Fink GR, Zilles K. 2004. Analysis of neural mechanisms underlying verbal fluency in cytoarchitecturally defined stereotaxic space--the roles of Brodmann areas 44 and 45. *Neuroimage* 22(1):42-56.
- Amunts K, Zilles K. 2001. Advances in cytoarchitectonic mapping of the human cerebral cortex. *Neuroimaging Clin N Am* 11(2):151-169.
- Arthurs OJ, Boniface SJ. 2003. What aspect of the fMRI BOLD signal best reflects the underlying electrophysiology in human somatosensory cortex? *Clin Neurophysiol* 114(7):1203-1209.
- Bechtel W. 2002. Aligning multiple research techniques in Cognitive Neuroscience: Why is it Important? *Philosophy of Science* 69:48-58.
- Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19(2):261-270.
- Culham JC, Danckert, S.L., DeSouza, J.F., Gati, J.S., Menon, R.S., Goodale, M.A. 2003. Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp Brain Res* 153(2):180-189.

- Duncan J, Owen , AM. 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23(10):475-483.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, 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(6676):598-601.
- Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1(1):1-47.
- Fischl B, Sereno MI, Tootell RB, Dale AM. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8(4):272-284.
- Friston K, Rotshtein P, Geng J, Sterzer P, Henson R. In press. A critique of functional localizers. *Neuroimage*.
- Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107(1-3):293-321.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425-2430.
- Heberlein AS, Saxe RR. 2005. Dissociation between emotion and personality judgments: Convergent evidence from functional neuroimaging. *Neuroimage*. In Press, Corrected Proof, Available online 13 October 2005,

- (<http://www.sciencedirect.com/science/article/B6WNP-4H9YC6J-1/2/af3152ed1f32b0ebdd3b8eccdd5c84c4>)
- Henson R, Shallice T, Dolan R. 2000. Neuroimaging evidence for dissociable forms of repetition priming. *Science* 287(5456):1269-1272.
- James TW, Culham JC, Humphrey GK, Milner AD, Goodale MA. 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain* 126(11):2763-2475.
- Jiang Y, Kanwisher N. 2003. Common neural mechanisms for response selection and perceptual processing. *J Cogn Neurosci* 15(8):1095-1110.
- Jiang Y, Saxe R, Kanwisher N. 2004. Functional magnetic resonance imaging provides new constraints on theories of the psychological refractory period. *Psychol Sci* 15(6):390-396.
- Johansen-Berg H, Behrens TE, Robson MD, Drobnyak I, Rushworth MF, Brady JM, Smith SM, Higham DJ, Matthews PM. 2004. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc Natl Acad Sci U S A* 101(36):13335-13340.
- Kamitani Y, Tong, F. 2005. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8(5):679-685.
- Kanwisher N, Woods RP, Iacoboni M, Mazziotta JC. 1997. A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience* 9(1):133-142.

- Kanwisher NG, McDermott J, Chun MM. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17(11):4302-4311.
- Kourtzi Z, Kanwisher N. 2001. Representation of perceived object shape by the human lateral occipital complex. *Science* 293(5534):1506-1509.
- Liou M, Su H-R, Lee J-D, Aston JAD, Tsai A, Cheng PE. 2005. A method for generating reproducible evidence in fMRI studies, *NeuroImage*, In Press, Corrected Proof, Available online 14 October 2005 (<http://www.sciencedirect.com/science/article/B6WNP-4HB4DFR-1/2/d381e49e6d3777214bc05092a20c470c>)
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92(18):8135-8139.
- Mariano AJ, Chin TM, Ryan EH, Kovach R, Brown OB. 2002. On Gulf Stream Path Variability. *Journal of Physical Ocean*.
- Nichols T, Hayasaka S. 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res* 12(5):419-446.
- Nielsen FA. The Brede database: a small database for functional neurimaging; 2003 June 19-22, 2003; New York, New York.

- Ozcan M, Baumgartner U, Vucurevic G, Stoeter P, Treede R-D. 2005. Spatial resolution of fMRI in the human parasyylvian cortex: Comparison of somatosensory and auditory activation. *NeuroImage*. 25(3):877-87.
- Paus T. 1996. Location and function of the human frontal eye-field: a selective review. *Neuropsychologia* 34(6):475-483.
- Peelen MV, Downing, P.E. 2005. Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum Brain Mapp* 25(4):402-408.
- Saxe R, Jamal N, Powell L. (2005). My Body or Yours? The Effect of Visual Perspective on Cortical Body Representations. *Cereb Cortex*, in press.
- Saxe R, Powell L. In press. It's the thought that counts: specific brain regions for one component of Theory of Mind. *Psychological Science*.
- Schuchard RA, Cummings RW, Ross D, Watson G. 2003. Comparison of SO perimetry, binocular perimetry, and functional visual field perimetry in patients with macular scotomas. *ARVO Annual Meeting*. p 2778.
- Shuman M, Kanwisher, N. 2004. Numerical magnitude in the human parietal lobe; tests of representational generality and domain specificity. *Neuron* 44(3):557-569.
- Spiridon M, Fischl B, Kanwisher N. 2005. Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum Brain Mapp*.
- Spiridon M, Kanwisher N. 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35(6):1157-1165.

- Swallow KM, Braver TS, Snyder AZ, Speer NK, Zacks JM. 2003. Reliability of functional localization using fMRI. *NEUROIMAGE* 20(3):1561-1577.
- Talairach J, Tournoux P. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Verlag. 122 p.
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 15(4):3215-3230.
- Wohlschlager AM, Specht, K., Lie, C., Mohlberg, H., Wohlschlager, A., Bente, K., Pietrzyk, U., Stocker, 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.
- Xu Y. 2005. Revisiting the Role of the Fusiform Face Area in Visual Expertise. *Cereb Cortex*. 15(8):1234-42
- Yovel G, N. K. 2004. Face perception: domain specific, not process specific. *Neuron* 44(5):747-748.