



Studying the Mind from the Inside Out

DAN LLOYD

Department of Philosophy, Trinity College, Hartford, CT, USA (E-mail: Dan.lloyd@trincoll.edu)

(Accepted: 26 April 2002)

Abstract. Good research requires, among other virtues, (i) methods that yield stable experimental observations without arbitrary (post hoc) assumptions, (ii) logical interpretations of the sources of observations, and (iii) sound inferences to general causal mechanisms explaining experimental results by placing them in larger explanatory contexts. In *The New Phrenology*, William Uttal examines the research tradition of localization, and finds it deficient in all three virtues, whether based on lesion studies or on new technologies for functional brain imaging. In this paper I consider just the arguments concerning brain imaging, especially functional Magnetic Resonance Imaging. I think that Uttal is too harsh in his methodological critique, but correct in his assessment of the conceptual limitations of localist evidence. I propose instead a data-driven test for assessing relative modularity in brain images, and show its use in a secondary analysis of fMRI data from the National fMRI Data Center (www.fmridc.org). Although the analysis is a limited pilot study, it offers additional empirical challenge to localism.

Key words: cognition, functional MRI, functional neuroimaging, modularity

For the last half century science has been poised to confirm one of the last great reductive hypotheses, namely, the identity of “mind” and brain. To several generations of cognitive scientists, the goal has seemed tantalizingly close. And yet, with each passing decade final success proves elusive; optimism finds itself rebuked. A prominent rebuker is William Uttal, and in *The New Phrenology* he critiques what to many is the brightest hope for mind-brain identity, the assignment of mental capacities to localized modules in the brain. Neuropsychology, with a long tradition of lesion studies, is one source for localist thinking, but a transformation in cognitive neuroscience began with the development of functional brain imaging. Positron Emission Tomography and (more recently) functional Magnetic Resonance Imaging have flowered in an incredible array of studies. If sheer numbers – of researchers, research centers, and published studies – suggest a successful research strategy, then functional brain imaging is a very successful enterprise indeed.

Nonetheless, to Uttal the localization strategy is deeply flawed, and his encyclopedic review questions localization from every side. The scientific stakes are very high, so it is essential that critics like Uttal pummel the localists, but it must nonetheless be a fair critique on all fronts. I think in a few respects Uttal’s review overstates the problems with localization, and my first goal is to suggest moderation in those aspects of the critique. But that limited defense of localization will

ultimately only enlarge the difficulties it faces. I will try to characterize the biggest problems, and suggest and demonstrate a way forward that could only be attempted on the foundation of functional brain imaging.

1. Localist Methods

“Localism” denotes, first, a set of research methods, and second, a set of hypotheses that seem to be confirmed by those methods. The methods, roughly, are those of *localization* of an effect in an area of the brain. The hypotheses that may or may not rest on these methods concern *modularity*, the identification of brain areas with particular cognitive functions. In assessing localism, then, it is important to keep the distinction between methods and claims resting on the methods, and it is important to bear in mind the multiplicity at both levels. Uttal’s critique bears on every aspect of localism, but with uneven efficacy. I will critique the critique not primarily to evaluate Uttal’s valuable book, but rather to join Uttal in the effort to assess this important scientific trend.

A research method fails when the observations generated by the method are shown to be inconsistent or arbitrary, or when observations are systematically misinterpreted (introspection is a familiar example). As methods, the two localist strategies of lesion studies and functional brain imaging are independent. Accordingly, problems with one do not necessarily impugn the other. Here, I will restrict my discussion to functional brain imaging, examining some of Uttal’s objections to it as a method.

1.1. LOCALIST OBSERVATIONS

Does functional brain imaging research set arbitrary standards for significance? Uttal points out in one section that any functional image will necessarily have a point of maximum activation, which he implies will make localist hypotheses unfalsifiable (p. 185). In another section, he claims that localist researchers may set thresholds at their convenience (p. 197). The logical point, that there will always be an image maximum, is true, but that in no way implies that the maximum will be significantly different from other image values in any multi-subject study. Nor, in general, are imagers cavalier about significance. Because signal-to-noise in functional brain imaging is generally poor, imagers are fairly obsessed with the statistical validation of their results, with special emphasis on avoiding false alarms. I believe they would be offended at the charge that they adjust their criteria for claiming an effect by arbitrarily setting thresholds for inclusion of more or fewer voxels in the region of interest. In most recent research, data run a thorough gamut of statistical tests, and for extra insurance many researchers set critical values for significance well below the traditional 0.05 probability. The steps for “de-noising” and pre-processing data are drawn from a limited set of standardized options, and any departures from normal practice are discussed at length –

in mathematical terms – in the research reports (Ashburner and Friston, 1997; Bandettini, 1993; Cohen, 1996; Friston, 1996; Lange, 1996; Poline *et al.*, 1997).

Are functional brain imaging observations inconsistent? As a methodological issue, inconsistency would arise if the same stimulus/task conditions gave rise to different observations. Uttal points out, correctly, that functional brain imaging encounters large individual variability. Anatomical variability is commonly corrected by projecting subject brains onto anatomically standard templates, but since this correction serves to identify the same anatomical areas across subjects, it does not undermine functional observations correlated with those areas. Functional variability would undermine the method if intrasubject and intersubject variability was so great that no general observations emerged after averaging within and across subjects. But common areas of significant activation do emerge. The observations are consistent across subjects to that extent at least.

A more puzzling claim is that functional brain imagers neglect to replicate previous results. In this, Uttal follows a website discussion by Harpaz (1999). Only in the narrowest sense of replication, the complete duplication of an experiment, is this criticism warranted. But replication in the sense of coordination with prior studies is especially thorough in imaging studies. Explicit replication of earlier PET results is a common theme of many recent fMRI studies. More important, every study I've seen includes elaborate discussions coordinating current results with many similar studies. Indeed, it would be foolhardy for an imaging researcher to test a hypothesis that had not already been suggested by prior research. Each new study is a motivated variation on a predecessor, and results are invariably compared to previous findings. In this respect, brain imaging advances very much like any science. Pure replication is rarely the sole purpose of any study. Nor, on the other hand, do studies take pure shots in the dark. In the middle is the cautious practice of building on the past to elaborate an emerging picture of the capacity of interest.

To put these points positively, functional brain imaging is a field that exhibits high consistency in methods and observations. If the working methods and assumptions were indeed wrong, the field should exhibit widespread inconsistency and instability in its findings. In the early days of PET and fMRI, its effectiveness for studying higher cognition was unknown. It very well might not have found anything significant in a search for localized contrasts typical of higher functions. Nonetheless, defying the odds, it worked. With so many studies now published, it would be Quixotic to claim that functional brain imaging either doesn't or can't deliver stable observations of relationships between cognition and brain function.

1.2. LOCALIST INTERPRETATION

Another part of Uttal's methodological criticism of localism concerns not the observational consistency of the methods but their interpretation. Here he ascribes several interpretive practices to imaging research, suggesting that imagers would claim the following:

- (i) The loci of activation observed are associated with precisely defined (and independently established) psychological capacities.
- (ii) The activity imaged is a direct reflection of underlying neural activity.
- (iii) The neural activity underlying the image foci is causally sufficient for the execution of the psychological capacity in question.
- (iv) The neural underpinnings of the imaged loci are exclusively dedicated to the capacity in question, and do not contribute to other functions.
- (v) The neural substrate identified through imaging is a single, well-defined region.

Uttal rejects all of these claims, and with good reason. But in general the imaging community would as well.

First, there has indeed been a shared concern to elucidate the relationship between neurodynamics and the blood oxygenation response imaged in fMRI (item ii, above, and p. 195). Very recently Logothetis *et al.* brought a long research arc to fruition by simultaneously recording from neurons with microelectrodes while imaging the same cells with fMRI, and their paper substantiates through theory and experiment the supposition that fMRI detects neurodynamics, albeit indirectly (Logothetis *et al.*, 2001). This once doubtful interpretive assumption seems lately to be resting on firmer foundations than when Uttal completed his manuscript.

A large part of Uttal's critique concerns the status of mental capacities, which he finds historically to be variable, vague, and circularly defined. Quite apart from the issues of localism, this history and Uttal's interpretation of it is fascinating, and I was persuaded that a unique, precise taxonomy of mental capacities can never be established in any psychology based solely on stimuli and behavior. I'm not sure how worried functional imagers should be about this, however. Imagers can finesse this limitation in two ways. First, they can (and do) hedge their claims about the capacities being studied in particular imaging experiments. It is enough to recognize that a particular task activates some aspect of language (or not), or vision, or motor control, etc. These links between the task imaged and the broader capacities are rarely controversial. In making them, imagers are not appealing to an exhaustive taxonomic reality for the capacity, nor claiming that the task fully operationalizes the broader capacity. It is sufficient, in general, to offer the task as an example of a capacity whose exact boundaries are indeterminate. Moreover, imagers should not feel obliged to go further than that. The actual practice of reductive research in science usually passes through intermediate stages where entities to be reduced are vaguely characterized (Wimsatt, 1976). The research clarifies the entities in question at the same time as it reduces them, as the lower levels project new taxonomies and new methods of individuation back into the higher levels.

Uttal's supposition that imagers claim that image loci are causally sufficient for executing the capacity imaged is to some extent a straw man. Even in his list (21–24) of instances of strong localist statements, "causes" often softens to "is associated with." Indeed, the hedged claim of a correlation or "association" of

function and locus is the rule in functional imaging. To illustrate this, I revisited the abstracts of twenty-two imaging papers referenced in (Lloyd, 2000), flagging the language used to express the final conclusions in the abstracts of each. Two papers made strong localist claims in which brain regions were claimed to be sufficient for executing a particular task, or the sole locus of the task (the strong localist claims of Posner and Raichle can also be added to the list (Posner and Raichle, 1994)). Eight papers fell along a spectrum of hedged localist conclusions. At the more strongly localist end of the continuum, some region or other would be reported to be “critically involved,” “engaged in”, or “among a number of components of a system responsible for” some cognitive function. Weaker localist conclusions included verb phrases like “is related to”, “may participate in”, “may reflect”, “may share components with”, “is involved in”, and “may play an important role in”. The remaining papers went no further than to claim that tasks activated certain regions, using verbs like “associated”, “activated”, and “affected.” That is, they made no claims about the causal function of the regions identified. In general, researchers seem to be aware of the logical limits of their claims, although this does not always carry over into their popular presentations. Weaker claims like these do not rule out multifunctionality for particular brain regions. And it is common to find a function associated with several regions of differential activity. In all these respects, the literature of functional imaging usually stops short of the extreme interpretations critiqued in Uttal’s book.

2. From Localism to Modularity

I’ve defended functional brain imagining mainly by showing that the localist methods and interpretations are sound – within rather narrow limits. That is, I’ve claimed that imagers in general don’t employ localization to support strongly modular conclusions, at least not in print. Nonetheless, the *impression* that this research serves strong modularity is widespread. Could imagers justifiably join with the popular press to declare that the trend of all their studies is to support strong modularity? Uttal’s answer is a resounding no, and in this I agree. Strong modularity really would require an exhaustive characterization of the capacity under study, a demonstration that a specific region really was sufficient for executing every aspect of the capacity, and a further demonstration that the region does not contribute to other distinct capacities. Individual studies never reach these ambitious conclusions. But if contemporary imaging research is not supporting strong localization, what does it support? In the remainder of this paper, I will suggest some revisions to the research program governing functional imaging, proposing a new way into the issues of modularity.

Modularity is an empirical hypothesis, but the methods currently employed bear on it only weakly. Consider the usual form of a localist claim about the relationship of a function and an anatomically defined brain region:

Function f is located in region R of the brain.

So stated, any localist hypothesis is completely compatible with claims of distributed processing, since the claim is consistent with the following:

Functions f_1 (and $f_2 \dots f_n$) are located in regions R (and $R_2 \dots R_n$).

Multifunctionality is precisely what distributed processing proposes, where each region contributes to many functions. Conversely, multiregionality for each function equally implies distribution. Accordingly, strong localism demands two further conditions. The strong localist claim must be exclusive (of other functions) and exhaustive (implicating no other regions):

Function f_1 (and no other) is located in region R (and no other).

R, by the way, need not be anatomically contiguous. It could have several components and still be localist, as long as its components are all dedicated to the function. Posner and Raichle's various networks are noncontiguous modules of this sort, subsystems of dedicated function (thus, comments on "networks" in Posner and Raichle, 1994, are not necessarily inconsistent with localism, contrary to Uttal's charge on p. 198).

To test a strongly modular hypothesis, one cannot begin with one function, since any function-based localization leaves open possibilities of many other functions sharing the region. Instead, one must begin with a region of interest, asking what it does. This may at first seem like a hopeless task, since a full characterization of the function of anything requires testing myriad potential functions. But one fortunate side effect of the enormous popularity of functional brain imaging is the wide array of experiments that have already been performed. That is, region-based research can move forward immediately with the large-scale meta-analysis of the many existing imaging studies – I agree with Uttal's call for more of these studies (Cabeza and Nyberg, 2000; Lloyd, 2000). Working from published studies has some limitations, however. For example, contrast images are only comparable if they are based on similar control or baseline comparisons (a condition not met in the Cabeza review. See Lloyd, 2000 for discussion).

Uttal also flags a more serious concern arising from the method of contrastive imaging in general, namely, the assumption of "pure insertion" (Friston *et al.*, 1995). Pure insertion is the working assumption that the processing underlying a task is engaged only while the subject performs the task. For example, in a probe of processes underlying word understanding, a target task of reading words will be contrasted from a baseline task of viewing pseudowords, with the goal of distinguishing the semantic component of reading from the orthographic processing of letters. However, one could argue that the attempt to read pseudowords engages the same semantic processes (unsuccessfully), and even that it is through engaging semantic processing that subjects properly perceive pseudowords as letter strings as opposed to mere marks. The contrast then may be based on a contrast with "semantic failure," a very special cognitive state distinct from the intended baseline. Imaging studies of "higher" cognition generally face a dilemma. Pure insertion only holds when target and baseline tasks are very different (and even

then only approximately), but in those cases the target task will activate many regions. The effort to more narrowly locate modules activated by specific cognitive processing requires finer distinctions between target and baseline tasks, and this in turn undermines the assumption of pure insertion.

This dilemma should inspire a new look at the whole problem. Contrastive imaging needs to be supplemented by some other form of analysis that considers the entire brain prior to filtering with subtractive controls. Moreover, – taking Uttal to heart – this supplemental analysis cannot assume the experimenter’s taxonomy of cognition is anything like the brain’s. That is, we should study modularity from the inside out, considering the activity imaged in the brain first, prior to any interpretation of functional taxonomy. Uttal refers to the “unassailable complexity” of neural processing (pp. 146, 174); it is indeed difficult to meet these new conditions, but I think there is hope. To interpret brain images without either subtraction or a priori distinctions among cognitive processes requires considering whole patterns of neural activation over the entire time span of an experiment. It is, in short, a multivariate problem. Paradoxically, we gain leverage on modularity by thinking globally about its “holistic” manifestations.

To start anew on the issue of modularity, it may be helpful to begin with a few extreme examples – radically simple cases of extremes of modularity and its opposite, distributed processing. Figure 1 shows a stylized brain image, sketched with a mere eleven pixels, and shown at three time points – imagine it as a side view of the cerebrum, with the frontal lobe to the left. Light pixels indicate activation. In all these examples, we will assume that total activity is shown, i.e., these are not subtraction images.



Figure 1. Cartoon brain with strongly localized (modular) function.

In this example, we assume that the dark pixels denote regions that are completely inactive. Strong localization is true for this system over this brief interval, because whatever this brain is doing at any time relies exclusively on a specific subset of regions.

Figure 2 shows the opposite extreme, once again with light pixels representing activation. Note that the actual functions computed in either sketch are irrelevant to the distinction between modular processing and distributed processing. What makes the first figure a sketch of modular processing is that regions active from one time to another do not overlap. In this way, the conditions of exclusivity are met. And, since all the other regions are inactive, the two active pixels exhaust the activity responsible for the processing. In contrast, in the distributed example no

region (other than the set of all) is either exclusively or exhaustively responsible for any activity. Also note that in a time series like Figure 1 patterns of localized activity might repeat at different times. In that case, the “brain” imaged would still be maximally modular, but would display fewer modules in the time series.

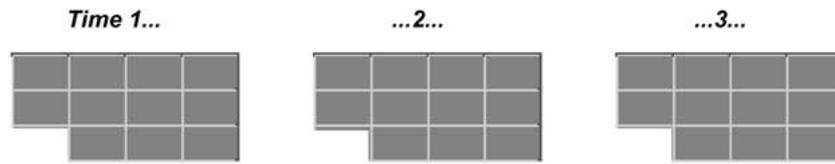


Figure 2. Cartoon brain with fully distributed processing.

In these figures, the distinction between modular and distributed processing is apparent to the naked eye. But of course nothing so simple will ever characterize our nervous system. Easing toward plausibility, Figure 1A shows Figure 1 with noise added to the image.

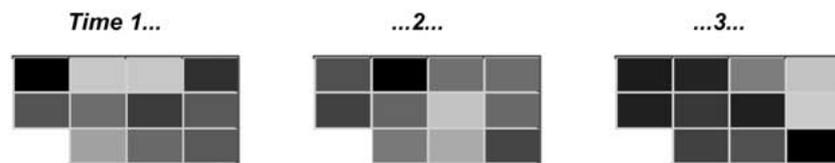


Figure 1A. Modules from Figure 1 in the context of noise.

And Figure 2A illustrates our hypothetical distributed processor with the same amount of noise added.

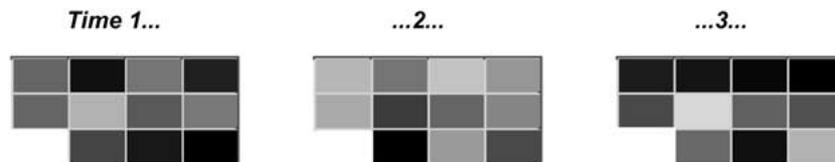


Figure 2A. A fully distributed processor in the context of noise.

The noise added here is simple random variation, but in functional neuroimaging the source of the variation will be partly unknown. It might be environmental, or physiological but unrelated to cognitive processes, or genuinely cognitive (the last possibility is one of the problems of contrastive imaging strategies. How not to throw the baby out with the bathwater?). But the examples suggest that these questions do not need to be resolved, since it remains the case that the system in Figure 1A is more modular than that in Figure 2A. Among the jitters the outlines of the supposed modules can still be seen, as a component of an otherwise random image. That is, the modules can be seen in 1A if one has 2A to compare. The latter figure also shows hotspots of apparently enhanced activation, but again in compar-

ison to 1A it is not as organized. The “brains” can be compared with respect to modularity, but in neither case can we make any absolute statement about whether a system is modular. Not only is the judgment of modularity relative, but it lies on a continuum from highly modular to highly distributed. The examples would fall on that spectrum in the order 1, 1A, 2A, 2. Finally, and most important, where there is a continuum there is a possibility of measurement. “Modularity” in principle is not a conceptual property to be assessed as all or nothing, but rather a specific global property of a system.

The simple cases also suggest how modularity might be measured. Consider again the time series of images presented in Figure 1A. Modules were visible in several dimensions of contrast within the images. In one time slice, the exhaustive contribution of the hypothetical module is visible to the extent that one area of the picture contrasts with others. The module holds center stage in solo performance. Considering the brain globally, this spatial focus is the only way an area could be determined to be sufficient for a particular cognitive function. If any other areas also lit up, we would be unable to deny them a role in processing in parallel to the region in question.

To establish the case that the purported module is exclusively dedicated to a function, we turn from space to time. As in the time series in 1A, the module should be largely isolated in function over time as well. This follows, albeit weakly, from the observation that the module lights up only sporadically in a time series of images. The more intermittent it is, the stronger the case against the module as multi-functional and contributing to many different cognitive processes.

Thus, to measure degree of modularity, one in effect tries to establish how much the system in question resembles Figure 1. Logically, modularity implies that particular regions will be relatively more active in comparison to the rest of the system, measured both at a particular time and over several time points. Accordingly, we seek a global measure that assesses images for intensity peaks and troughs, in contrast to the “flatter” images expected in distributed processing. Several such measures are readily applicable; here I will discuss only the simplest.

One useful measure is the dispersion from the mean for images. It will be characteristic of relatively modular systems that they will feature more regions with extreme values, peaks of intensity rising above planes of relative inactivity. Expressed in relation to the mean intensity of an image, this entails that pixel intensities will tend to collect further from the mean, both above it (the peaks) and below (the troughs). In short, they will be dispersed further from the mean. So, considering all the individual pixel intensities, a simple descriptive statistic of dispersion can test for the presence of contrasts indicative of modularity. Variance and its square root, the standard deviation, are familiar measures of dispersion. They can be directly employed as a global measure of modularity, or one can measure *relative* dispersion by comparing the magnitude of dispersion to the magnitude of the mean. This measure is often called the “coefficient of variation” (CV), and is simply the standard deviation divided by the mean. Where images

have differing mean intensities (as is often the case in functional brain imaging), the coefficient of variation is useful for steadying the frame of reference. By canceling the effects of changing means, CV is not “fooled” by global increases and decreases in activation, but rather shows dispersion alone.

The coefficient of variation can be calculated for any subset of an image or image series. In our toy case, we can create a “CV image” of the entire series by calculating the coefficient for each pixel as it progresses through the series. Figure 3 shows the contrast between the noisy-but-modular brain and the noisy-but-fully-distributed example, Figures 1A and 2A, respectively. All three of the purported modules remain somewhat discernable in the composite.

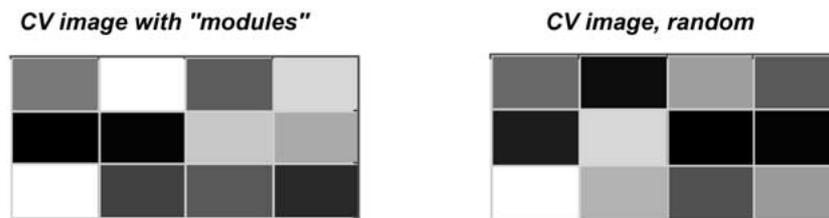


Figure 3. Modularity imaged as coefficient of variation for complete image time series.

A more global analysis, however, is offered by the simple computation of CV for each image in its entirety. Table I, below, lists the global CV for each image of Figures 1A (left column) and 2A (right column).

Table I.

Time	“Modular”	“Random”
...1...	0.63	0.29
...2...	0.47	0.40
...3...	0.52	0.42
MEAN	0.54	0.37

Global measures like the coefficient of variation ultimately preclude pictorial display of brain images. But in another way global CV captures what we are looking for, namely, the distinctive contrasts that result from modularity. It turns out to be a sensitive measure. I simulated a larger version (100 pixels) of Figure 1A incorporating a “module” (intensity peak) that extended over just 2% of the image area. In spatial extent this tiny area is approximately the size of regions identified in localist, contrastive brain imaging studies. This “modular” image was contrasted with a “flat” image, as in Figure 2. Random noise (with a normal distribution) was added to each image to approximate the usual statistics for fMRI. Signal-to-noise, calculated as the ratio of means of the two images, was 2%. The coefficient of

variation in the control image was 0.28; for the modular image it was 0.31. High dispersion values are not unusual, as fMRI signals vary greatly within images. Ten thousand image pairs were generated, and each pair tested to see which image had the higher coefficient of variability. In 96% of cases, the modular image had the higher CV. Thus, in a noisy statistical environment similar to fMRI, the CV test for modularity was generally sufficient to detect a difference.

3. Modularity in a Real System

“Analysis of modularity” is presented here as a tentative method, its further use dependent on further study of its statistical implications. But, should it pass mathematical muster, we will of course want to know if it shows anything about real brains that we can interpret, and thus, ultimately, if it offers insight into cognitive processing by the brain. To turn toward real neuroimaging, just take the eleven pixels in the figures above and multiply by about ten thousand, expand the time series to as many as a thousand images (in each subject), and decrease the observed contrast (“signal to noise”) to about 1%. Perhaps the spectrum of modularity is present across neuroimaging studies, but of course no imaging will make it apparent to our eyes. Simulations suggested that the coefficient of variation captures a measure of modularity. In this section, I will present a brief pilot study based on data archived in the new National fMRI Data Center at Dartmouth (www.fmridc.org).

The study, by Postle *et al.* (2000), aimed to discriminate brain activity associated with working memory from that associated with saccadic behavior. Subjects engaged in four quite different tasks during their scan sessions: “Forward Memory,” in which subjects viewed a sequence of illuminated squares in different screen locations, recalled the sequence through a delay, and pressed a button in response to a probe about location of one of the squares; “Manipulate Memory,” in which subjects viewed a similar sequence of squares, but were instructed to mentally reorder the squares in space, and respond to a probe about the rearranged squares; “Guided Saccade,” in which subjects tracked a sequence of highlighted targets, but performed no other task; and “Free Saccade,” in which subjects made unguided horizontal saccades. Each trial of 34 seconds also included delay periods where subjects fixated on a central marker, and 17 seconds of rest with no instructions. Thus, the experiment imaged a variety of cognitive states in each subject, making it an interesting candidate for a search for evidence of modularity.

As with many other studies of functional brain imaging, the experiments of Postle *et al.* (2000) consisted of many repetitions of the four tasks, giving each subject twenty-four opportunities to attempt each one. Their brains were imaged every two seconds, collecting three-dimensional images (also known as “volumes”), each with about 16,000 image elements or “voxels.” For this pilot analysis, I measured the coefficients of variations for all the volumes recorded for three of Postle’s five subjects. The small sample, along with many other limitations,

imply that the exercise should not be regarded as more than an illustration of issues and opportunities that may arise as we look through this alternative window into the brain.

Figure 4 shows the evolution of the coefficients of variation over time in two subjects.

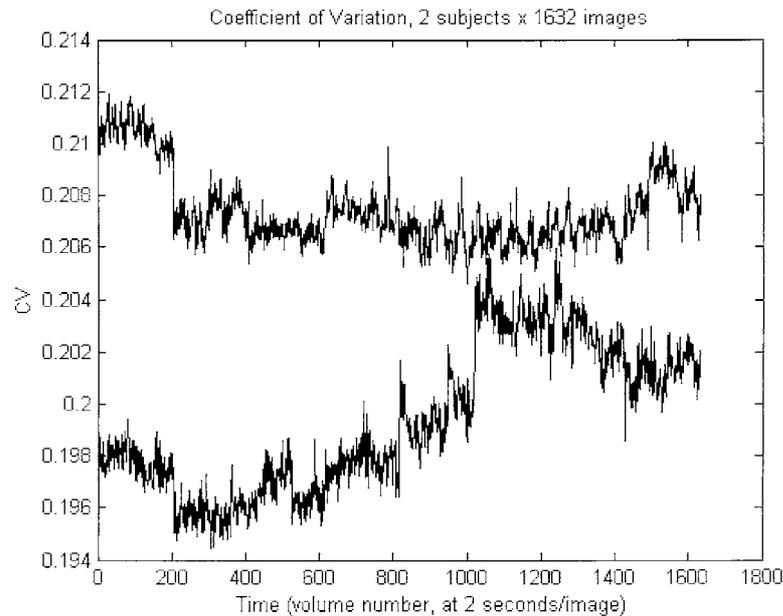


Figure 4. Variation in two subjects over the entire experiment in Postle *et al.* (2000).

One first observes that both subject brains vary within a limited range. Unlike the simulations discussed above, there is no random control, so the global effect of about 0.2 cannot be understood as an absolute measure of either modularity or its absence. What can be interpreted, however, are the contrasts in CV between images. In these two subjects, those contrasts are small, approximately 0.05 with occasional jumps. But though small the effect may be significant. But it will require further analysis to establish a basis for either accepting or rejecting hypotheses on these data.

The next step, accordingly, is to compare the changing CV within particular tasks. Figures 4 and 5 jointly suggest that the variability lies near a few plateau values; closer inspection shows that level jumps occur at run boundaries, where the experiment pauses for a few minutes to allow the scanner to download data to disk. So, prior to comparing CV for particular tasks, I recalculated each time series of image CVs as z scores. In this way, the large scale shifts shown in Figures 4 and 5 are removed, in the hope of allowing smaller effects to appear.

Figure 6 shows CV time series for each task in the three subjects. All four results display significant differences over time (as assessed with 1-way ANOVA. Free

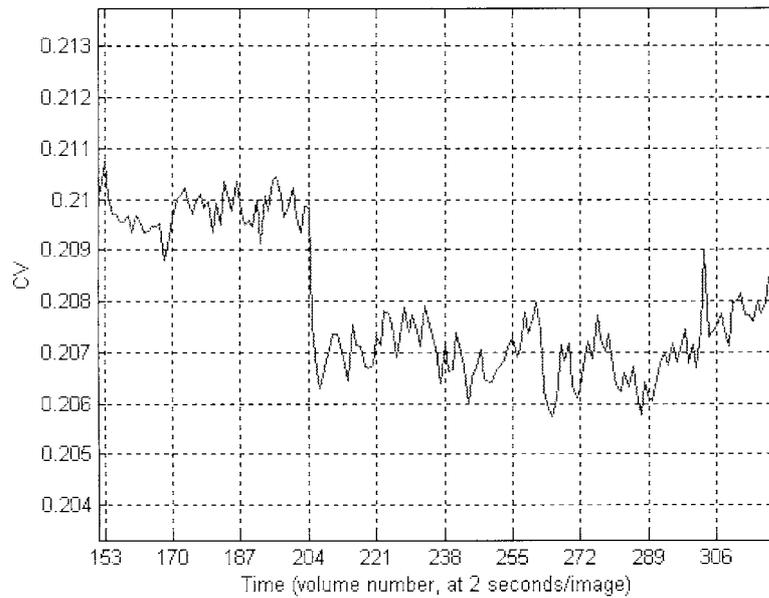


Figure 5. Close view of Figure 4.

Saccade: $F = 28.6$, $p < 0.0001$; Guided Saccade: $F = 3.12$, $p < 0.0001$; Forward Memory: $F = 6.21$, $p < 0.0001$; Manipulate Memory: $F = 6.84$, $p < 0.0001$). CV analysis has detected a small but significant variation, and similarities across types of tasks can be observed in the graphs.

Each panel in Figure 6 follows the participants through a complicated 34-second routine. Their initial instructions (“Inst”) alerted them to the task to come, letting them know if it was a memory task, a “no memory” task (i.e., guided saccade), or the “free saccade” task. In the last case, subjects began saccadic movement immediately and continued while a blank screen was displayed during the Stimulus interval (“Stim”). For the other three tasks, the Stimulus was a sequence of highlighted squares in various locations. Then appeared a second set of instructions. In the two Saccade conditions, subjects were instructed to Fixate. In the two memory conditions, subjects were instructed either to remember the order of highlighted squares (Forward Memory) or to mentally rearrange the order (Manipulate Memory). All subjects then fixated on a central marker for seven seconds. Then the memory subjects were probed about the location of a particular element in the remembered or rearranged sequence, and answered yes/no through a button press. The saccade subjects saw the same probe array, but did nothing. Finally, all subjects finished with seventeen seconds of rest (*Note*: The hemodynamic response requires 4–6 seconds to reach its maximum; all the onsets in the task sequence have been shifted four seconds in Figure 6 to accommodate this lag (Aguirre *et al.*, 1997, 1998).

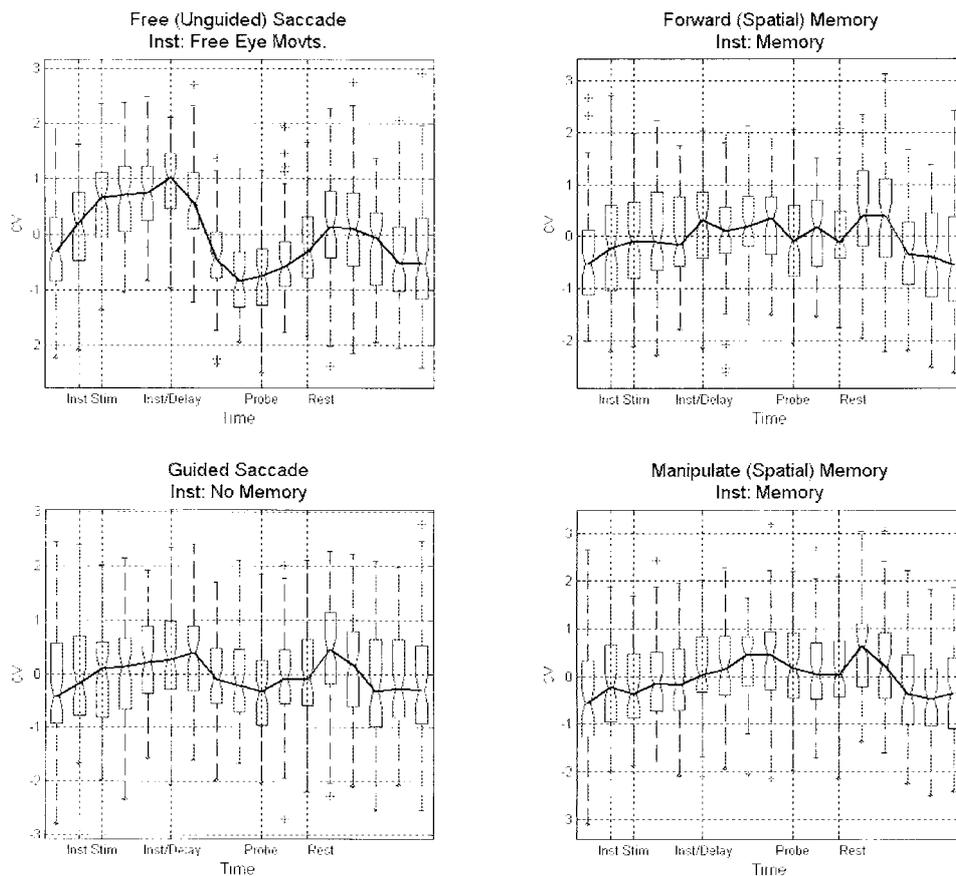


Figure 6. Coefficients of variation measured over time in repeated tasks in three subjects in Postle *et al.* The graph shows the median CV z score flanked by boxes indicating quartiles above and below the median. Plusses are outliers. Marked time points are: Inst: Initial Instructions, as stated in titles on each panel. Stim: Stimulus. In the Free Saccade condition, the stimulus was a blank screen and saccades began immediately. In the other conditions, the stimulus was a sequence of squares appearing in different locations. Inst/Delay: Instructions and Delay. For the two saccade conditions, subjects were further instructed to fixate. For the memory conditions, subjects were instructed to remember the order of positions occupied in the stimulus sequence (“Forward Memory”), or to mentally rearrange the order along a single dimension (“Manipulate Memory”). During Delay, all subjects fixated a central cross. In both Memory conditions, the spatial location memory was tested at “Probe,” with match/mismatch button response. In Saccade conditions, probe array was also presented but with no behavioral response. Uncontrolled rest followed for all subjects.

With four conditions and the complicated time-line of each task, there could be many comparisons available. I will focus on just a few of them. First, the two memory conditions and the two saccade conditions seem similar to each other. In the saccade conditions, CV peaks early. This is most marked in the Free Saccade condition, where CV peaks in a time course reflecting the earlier start of the saccade

task. In contrast, CV peaks during the no-response delay in the two Memory tasks. Other differences do not seem to make a difference in these series. In the saccade tasks, it did not matter that in the Guided Saccade condition subjects tracked a stimulus identical to that in the memory conditions. And in the two memory conditions, it did not matter whether subjects simply retained the order of highlighted squares, as opposed to mentally manipulating that order. Between the conditions, it did not seem to matter whether subjects actually initiated a behavior response. Finally, rather strikingly, a second CV peak occurred in all subjects during the uncontrolled rest period following all tasks.

In the previous section, I suggested that the coefficient of variation provided a broad measure of image contrast, and that image contrast in turn would appear as an effect of modularity. Figure 6 shows contrast peaks. If we assume for the sake of discussion that observations like this also emerge from a larger subject pool, how might they bear on the issue of modularity? In one way, these results might encourage a modest modularity. As intended, the results rest on entire images rather than subtraction images, and so any activation that appears in the image can affect CV. As a result, the observations are data-driven results that do not rest on assumptions about which cognitive processes are distinct or additive. That is helpful in finessing the fallacies of method that otherwise hinder localism.

But it is not easy to find support for any particular modules based on these results. The expected sensory and motor modules do not appear. Saccadic movements are a constant in all four tasks, but only seem prominent in the saccade tasks. The visual stimulus is distinct and relevant in three of four tasks, but does not appear as a peak. The other behavior of button-pressing also lacks any CV manifestations. These various inconsistencies are surprising in that the sensory and motor modules were the first and most robust to be established in functional brain imaging, appearing more sharply than any of the many cognitive processes studied subsequently (even Uttal concedes that the senses and specific motor responses are modular). Whatever else one might have expected, in a whole-brain data-driven exploration like this, these peripheral models would have been expected to appear. But they don't.

Expected modules do not appear. Conversely, unexpected modules do appear – all four tasks display “strong” modularity during the uncontrolled rest period, after the task is finished. If CV is a measure of modularity, then some module seems to engage after the task is dispatched. Alternatively, the peak need not represent the same module in every case. Perhaps subjects turn immediately to their mental To Do lists once their prompted chore is finished. But nonetheless in the observations shown here following every task in both subjects, the peak was observed in almost every case. Not even the P300 EEG response is as steady as this.

Taken together, the surprises in this analysis offer empirical support for Uttal's defense of distributed processing (that is, once again, if they were to extend to other subjects and many more experiments). The unexpected results contradict two assumptions of modularity that found many studies: first, it contradicts the

assumption that the conscious rest state engages no particular cognitive process, and so can be presumed to be a diffuse state of activation that can be removed to highlight activity of interest. Second, and closely related, these observations contradict the assumption that sensory and motor states are subsumed by modules, and these too are stable across tasks, and therefore also candidates for removal in order to isolate particular processes.

4. Back to the Drawing Board

Data, data, everywhere. Powerful computers and extraordinary multivariate sensors afford science a wealth of information, far more than any human could describe or explain. Yet the basic impulse of science remains the same: to explain a concrete, specific phenomenon in terms of other concrete, specific phenomena. Cognitive neuroscientists, like all scientists (and philosophers), are constrained by limits of language, communication, and ultimately conception, and naturally seek Newtonian elegance at the foundations of cognition. Localism and modularity are optimistic attempts to achieve simplicity and systematicity, but to approach this goal localist research must filter the complex data on which it rests. I disagree with Uttal about the observational validity of functional brain imaging. It is certainly showing something real about cognition. But I agree that localist methods cannot answer some fundamental questions without begging them. Accordingly, cognitive neuroscience needs new strategies for data analysis. I expect progress in this field, whereas Uttal may not. With powerful computers also comes the capacity to revisit the richness of existing data, and cast the net again and again. Here, I've taken a first look at one simple measure based on an analysis of modularity, but there are many others. It is much too early to declare the end of functional brain imaging. Indeed, the adventure has only begun.

References

- Aguirre, G., Zarahn, E., and D'Esposito, M., 1997: Empirical analysis of BOLD fMRI statistics, *Neuroimage* **5**, 199–212.
- Aguirre, G., Zarahn, E., and D'Esposito, M., 1998: The variability of human, BOLD hemodynamic responses, *Neuroimage* **8**, 360–369.
- Ashburner, J., and Friston, K., 1997: Spatial transformation of images, in R.S. Frackowiak, K. Friston, C.D. Frith, R. Dolan and J. Mazziotta (eds), *Human Brain Function*, Academic Press, San Diego.
- Bandettini, P., 1993: Processing strategies for time-course data sets in functional MRI of the human brain, *Magnetic Resonance in Medicine* **30**, 161–173.
- Cabeza, R., and Nyberg, L., 2000: Imaging Cognition II: An empirical review of PET studies with normal subjects, *Journal of Cognitive Neuroscience* **12**, 1–47.
- Cohen, M.S., 1996: Rapid MRI and functional applications, in A. Toga and J. Mazziotta (eds), *Brain Mapping: The Methods*, Academic Press, San Diego.
- Friston, K., 1996: Statistical parametric mapping and other analyses of functional imaging data, in A. Toga and J. Mazziotta (eds), *Brain Mapping: The Methods*, Academic Press, San Diego.

- Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., and Turner, R., 1995: Analysis of fMRI time-series revisited, *Neuroimage* **2**, 45–53.
- Harpaz, Y., 1999: Replicability of cognitive imaging of the cerebral cortex by PET and fMRI: A Survey of recent literature. <http://www.yehouda.com/replicability.html>.
- Lange, N., 1996: Statistical approaches to human brain mapping by functional magnetic resonance imaging, *Statistics in Medicine* **15**, 389–428.
- Lloyd, D., 2000: Terra cognita: From functional neuroimaging to the map of the mind, *Brain and Mind* **1**, 93–116.
- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A., 2001: Neurophysiological investigation of the basis of the fMRI signal, *Nature* **412**, 150–157.
- Poline, J.B., Holmes, A.P., Worsley, K., and Friston, K., 1997: Making statistical inferences, in R.S. Frackowiak, K. Friston, C.D. Frith, R. Dolan and J. Mazziotta (eds), *Human Brain Function*, Academic Press, San Diego.
- Posner, M., and Raichle, M., 1994: *Images of Mind*, W.H. Freeman, New York.
- Postle, B.R., Berger, J.S., Taich, A.M., and D'Esposito, M., 2000: Activity in human frontal cortex associated with spatial working memory and saccadic behavior, *Journal of Cognitive Neuroscience* **12**(2), 2–14.
- Wimsatt, W., 1976: Reductionism, levels of organization, and the mind-body problem, in G. Globus, G. Maxwell and I. Savodnik (eds), *Brain and Consciousness*, Plenum, New York.

