© DIGITAL STOCK

Using Small Numbers of Subjects in fMRI-Based Research

Searching for Universal Truths in Human Brain Function and Organization

BY ROBERT L. SAVOY

Most discussions of brain activation functional magnetic resonance imaging (fMRI) data start with the analysis of images from a single subject but move rather quickly to the issues associated with group comparisons and drawing population conclusions. There are many technical issues associated with those group comparisons that do not arise, or are very different, for data from a single subject. In addition, there are contexts (most notably clinical evaluations) where studying an individual subject is essential. This article reviews a number of issues related to the special features and potentials associated with doing functional brain imaging experiments with one or a small number of subjects. It concludes with the speculation that, ironically, it may someday turn out that the information from a few brains, thoroughly studied, will reveal more about the universal aspects of human brain function and organization than the current torrent of studies from large collections of brains.

Background

Psychological tests are normally designed for individuals. For both scientific and clinical reasons, it makes sense to try to extract and verify information about individuals rather than groups. In the ideal setting, information from enough individuals is then used to draw conclusions about groups.

However, there are a number of contexts in which it makes more sense to speak about group averages than about individual measurements. Sometimes this is because we wish to ignore structural variability in the brains of individuals who compose the population data; sometimes this is because we wish to ignore functional variability in the organization of the brains of individuals (independent of any structural differences). In addition, sometimes there are situations where the measurements of small effects in an individual do not reach statistical significance, but those measurements *are* statistically significant when combined with effects in many subjects.

In the context of functional brain imaging, there has always been a sort of tension about what is the right kind of question to be asking. For the most part, current brain mapping researchers address questions about the *average* brain: seeking the aspects of brain function and brain organization that are common across individuals. After reminding the reader of several important aspects of this kind of study, the remainder of the article will focus on contexts and examples where studying individual brains has been serendipitously useful or clinically essential. The article will conclude with a speculative proposal for a study to be conducted with an extremely small number of subjects, or even a single individual subject, in the distant hope that such studies might point the way to integrating the torrent of data coming from functional brain imaging.

Averaging Across Subjects and Its Special Statistical Issues

Consider, for example, the simple case of having someone tap their finger on a table in time to a metronome that is supplying a regular auditory cue (the beat). In some clinical contexts, this task is then followed by stopping the metronome and having the subject continue tapping at the same frequency as best they can. When these tasks are performed while a subject is having their brain activity measured using fMRI (or other technologies), a constellation of active brain areas are seen. What are the most sensible questions to ask about these brain activations?

One question is, "What areas of the brain are active across all individuals who are engaged in this task?" For many researchers, this seems to be the only question of interest. The implicit suggestion is that the differences shown in brain activity between subjects is likely to be generated by uninteresting noise in the procedure. Who cares if a little piece of cortex happens to be active in Subject 1 but not Subject 2 when they do this task? Maybe this variation in activity is because while performing the tasks, Subject 1 was anticipating the fun they might have at an upcoming party while Subject 2 was worrying about being late for an upcoming appointment. These uncontrolled variables will change brain activations in idiosyncratic ways that are of no interest in the context of measuring finger tapping to a metronome. It is best to get them to cancel out, just as one does with any source of uncontrolled noise, by collecting data from many subjects and averaging the results. In the case of the aforementioned task, the important observation is that the primary motor cortex, premotor cortex, cerebellum, and auditory cortex are all active in ways that change across conditions but are consistent across subjects.

0739-5175/06/$20.00©2006IEEE

It may someday turn out that the information from a few brains, thoroughly studied, will reveal more about the universal aspects of human brain function and organization than the current torrent of studies from large collections of brains.

The preceding description is what underlies most current research reports using fMRI. Indeed, the importance of keeping the distinction straight between comparing groups (random effects analysis) and looking at effects in individual subjects (fixed-effects analysis) is emphasized in many discussions of fMRI data analysis. One particularly clear and useful example is [1], in which the question "How many subjects constitute a study?" is considered. The authors remind us that the relevant measure of variance (which determines our confidence about differences between measurements of brain states) is not the same when we are making many measurements of a single individual as it is when making measurements across two populations of individuals. A simple example helps to make this idea clear.

Suppose investigators have a device for measuring the height of people from a distance of 100 m. It is a fairly accurate device, giving investigators a reading of, say, 179 cm when they measure me. When the investigators measure my wife, they get 173 cm. Can they tell if I am taller than my wife? The answer is "no" until they know something about the precision of the measurement. If they take a 1,000 separate (independent) measurements of me, and the mean reading is 179 with a variance of .5 cm, and a 1,000 measurements of my wife with a mean reading of 173 and a variance of .5 cm, then they can have very high confidence that I am taller than my wife. What about the question: "Are men taller than women?" Whether they have one reading or 10,000 readings of me and my wife, they really only have one piece of information about the differences in height between men and women. Rather than taking 1,000 measurements to obtain a precise estimate of the height of each of us, it would be better to have noisy measurements of 1,000 different men and of 1,000 different women to address the latter question, because it is the variance in the population rather than (or, more accurately, in addition to) the variance in the measurement of a single subject that is important for assessing the statistical significance of the differences.

This example reminds us of one reason for the initial excitement about fMRI. Being able to take thousands of measurements of a given brain and calculating astronomically small p-values in t-tests from those measurements across different brain states gave an exhilarating (albeit sometimes misleading) sense of power for the technique. This initial enthusiasm was eventually tempered by the realization that magnetic resonance (MR) images of blood oxygenation collected less than 6 s apart could not be considered independent measures of a given brain state-temporal autocorrelation was not zero between such measurements. It was also tempered by the realization, outlined above, that highly significant differences between measured states in one brain should not be interpreted as a quantitative assessment of the accuracy available for measuring differences between groups of brains. Nevertheless, with proper care regarding the statistical issues associated with autocorrelation (and effective degrees of freedom), it was true that repeated measurements of these small effects greatly increased the power to detect subtle differences, even in individual brains, between activation patterns associated with different experimental tasks. But as elaborated in the following, many of those differences were then averaged away by combining data from different brains.

The preceding is simply a review of some aspects of statistics, as applied to measuring differences in brain states. In the world of functional brain mapping, there are additional considerations. The most important of these comes about because the measuring tools are relatively insensitive and the objects that they are measuring (human brains) are structurally quite variable between people. These two things interact in the analysis of fMRI data, just as they did in earlier studies using positron emission tomography (PET).

In early studies of cognitive processing using PET, the measured differences between brain states in a given individual were sometimes small. Data were collected across many individuals in order to gain signal-to-noise by averaging. But what should be averaged? Because the brains of different individuals were structurally different, averaging could wash out the effects. That is, the activation site in one brain might not line up with the activation site in another, so when the two were averaged by overlapping the raw brain images, a net loss in signal-to-noise could result. A breakthrough of sorts came with the realization that there was an existing procedure ([2] developed in the context of subcortical neurosurgery) to transform individual brains so that they would line up better with each other. Simply put, it was a procedure to place a given brain in a standard orientation and then piecewise linearly scale it to fit into a standard rectangular box, based on a few easily identified landmarks of the brain. More sophisticated computational techniques for aligning brains across individuals have since been developed (e.g., [3]), and the effectiveness of getting constructive interference between activations across the brains of different subjects has been improved. Nevertheless, there remain fundamental problems (as indicated in some of the following discussions) with drawing conclusions based on averaged brain activations.

Single Subjects in Groups: Keeping Your Eyes Open

Even when a study is, by design, going to be a group analysis, at least one very experienced reviewer of fMRI-based research [4] strongly encourages the presentation of data for individual subjects as well. There are many ways that an averaged picture can be misleading. The additional space required for one figure that shows all the subjects' results, rather than just the carefully selected typical result, is not too large, and the comfort generated in the reader (or reviewer) is large.

In the case of one study, this general rule had particularly fortuitous consequences. In 1996, Michael Chee spent six months at our laboratory designing both structural and fMRI studies. The project on which he worked with my colleagues and me involved the comparison of language activation via visual presentation of words versus auditory presentation of words. For this study [5], all subjects were screened to be strongly right-handed, using the Edinburgh handedness scale; and all the subjects *were* right-handed. But it is known from the lesion literature that, even among right-handers, there is a small percentage of people who have right-lateralized (rather than the much more common left-lateralized) brain specialization for language processing. When the data were analyzed for the eight subjects in this experiment, seven showed the expected clearly left-lateralized activations during language processing tasks involving a semantic judgment—deciding whether given nouns were abstract (words like *peace*, *love*, and *understanding*) or concrete (words like *book*, *dog*, and *table*)—when that task was compared with a nonsemantic task—deciding whether visually presented words were in uppercase or lowercase letters or deciding whether spoken words had one or more syllables (see Figure 1). However, one subject showed, just as unambiguously, right-lateralized brain activity during the language processing tasks. This subject came back to the laboratory and agreed to be tested again on both the study task and on several other, more standard language processing tasks in the magnet. In all experimental conditions, he showed consistent right-lateralization for language, despite being strongly right-handed [6]. The point here is not that right-lateralizing right-handers exist (that was known from earlier studies). The point is that by attending to the data from this individual subject and treating it separately from the group average, more accurate and cleaner summary results were obtained.

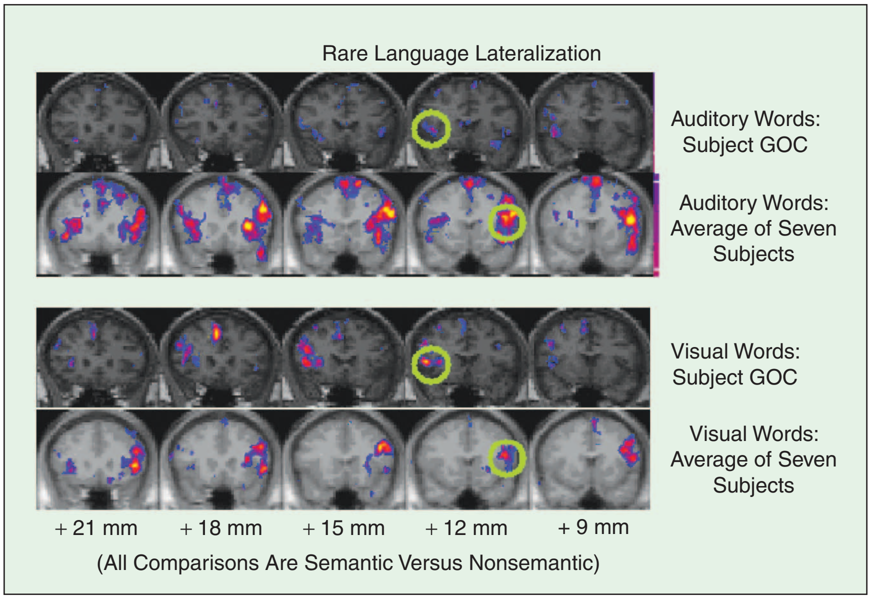

Fig. 1. Mirror-image reversals of lateralization in the data for subject GOC as compared with the average of seven other subjects (as ones circled in green).

More importantly, there are ways that averaging data can lead to the obscuring of important variability that is not due to an isolated case (as in the previous example) but is due to systematic differences across groups. An important review of imaging research in the schizophrenia literature [7] pointed out several ways in which group averaging can be misleading, quite apart from any difficulties with aligning the different brains with one another. While several sources of error were discussed in [7], the one most relevant to the present discussion is the idea that a patient population may show more variability in the location in the brain where a given activity is to be found. In particular, the author of [7] argues that some studies conclude that schizophrenic subjects show less activity in dorsolateral prefrontal cortex (DLPFC) because their activations are more variable in spatial location within the DLPFC. When averages are collected across the groups (healthy versus schizophrenics), the greater consistency in spatial location of brain activations in healthy subjects resulted in greater average activation for the DLPFC. In contrast, when indices of activation were derived from individual subjects rather than from group-averaged data, the schizophrenia group actually showed significantly greater activation within the prefrontal cortex (PFC) than healthy subjects. Again, attending to the data from individual subjects was crucial to understanding some discrepancies in the literature.

The last example in this section comes from an experiment that had a subtle design flaw, which apparently affected the data for only a fraction of the subjects [8]. A visual cue was presented regularly, every 16 s, and subjects were required to make a simple motor response when the visual cue appeared. For each subject, the hemodynamic response associated with this motor act, averaged across many trials, was calculated. Because the topic of the study was individual variation in fMRI responses, the data for each of the 32 individual subjects was shown—there was no averaging across subjects. This important study of individual differences in fMRI data documented the substantial variability in the absolute size of the hemodynamic response, which was not a surprise to experienced investigators. But a number of readers noticed something unusual in the time course of the measured hemodynamic response for a few of the subjects. There was a surprising variation in the onset time of the hemodynamic response: for some subjects, the MR signal began to rise immediately after the visual cue was presented, instead of having a delay of about 2 s, as is normally seen in this context. This had never before been reported.

Subsequent to the publication of the above study, the investigators came to realize that there might be a problem with the regularity of stimulus presentation: some subjects' brains could be anticipating the regular occurrence of the visual cue, and neural activity in the motor areas being studied could appear earlier than normal. The authors tested this hypothesis by repeating the experiment with a variable intertrial

Current brain mapping researchers address questions about the average brain, seeking the aspects of brain function and brain organization that are common across individuals.

interval, and, indeed, the short-delay rise in subjects' hemodynamic response disappeared. But the point, in the context of this article, is that the short-delay rise phenomenon was not seen in most of the subjects in the original study. For whatever reason, only a few of the subjects' brains generated this anticipatory response. (For a related discussion and other cautionary tales, see [9].) If the average of all the subjects had been the only thing reported, it would have been much more difficult to detect the problem.

Even averaging across trials for an individual subject can be a danger. In the preceding study, if the data for the individual trials associated with this task had been plotted (instead of just the average), the problem might well have been discovered much earlier. Plotting the time course data for each individual trial for these subjects (time shifted back to the onset time of the stimulus) might have revealed that the initial rise was moving to earlier times as the subjects' brains learned the temporal contingencies and began to anticipate the cue.

The general message is simply that any kind of averaging runs the risk of hiding important variations in the data. The perennial admonition to "stay close to your data" is appropriate any time averaging is done.

Single Subjects: Special Cases in Research

While no one makes it a policy to study only one subject per investigation, there have been some interesting reports that relied, in important ways, on the finding of exceptional subjects. Some of these subjects were found accidentally (as in [5], above), while others were actively sought [10]. Sometimes the subject was exceptional because of the physical structure of their brain [10], sometimes because of the operational activity of their brain during a routine task [8], sometimes because of the unusual functional organization of the brain [6], and sometimes because the clinical symptoms of a neurological disorder were exceptionally convenient for study [11], as has been the case in serendipitous clinical reports throughout history. (Indeed, the first reports of hemodynamic changes associated with neural activity in human brains preceded the advent of MRI by almost a century [12], [13].) A few of these examples are described more fully in the following.

Brodmann Area 17 in the human brain is known as primary visual cortex, and it follows the banks of the Calcarine fissure in the occipital lobe of the brain. The organization of the neurons in this part of the primate brain is known to include ocular dominance columns—columns of neurons whose input in a given column is all from one eye or the other or from a fixed, weighted average of inputs from both eyes. This much was known from the direct recording of single cells in primate and other mammalian brains. Early attempts to document this phenomenon using fMRI in humans were hampered by subject movement. The effect could be seen, but it was not easy to reproduce it in a single individual across multiple imaging sessions. One group of investigators—in addition to using high field magnets and all the tricks of MRI, physiological monitoring, and head stabilization and head-position monitoring they could—sought to give themselves one more advantage in addressing this challenge. They used conventional structural MRI to screen subjects in order to find a few who had exceptionally straight Calcarine fissures. Why did this help? The idea was that, when the folds that made up this fissure happened to be particularly straight, it was possible to predict what the ocular dominance columns should look like in a variety of different imaging planes, and, more importantly, it was possible to line up the imaging planes so that all of the relevant tissue would be contributing to the given image, thus improving the signal to noise. The result was an elegant and convincing demonstration of the same columnar organization in individual subjects across multiple imaging sessions [10].

My personal favorite example of a special individual subject comes from a study on the visual auras that sometimes precede migraine headaches [11]. For a substantial fraction of the people who get migraine headaches, the headache is preceded by a period of time during which there are noticeable (to the subject) distortions of the visual field. Often, there is an appearance of ephemeral crosshatching around a region of the visual field within which perception is lost. This region of loss in the visual field (scotoma) expands and moves peripherally over time. Approximately 20 min after the beginning of the aura, normal vision has returned, but usually the headache is about to start. It is known that these phenomena are somehow related to vaso-constriction and vaso-dilation near the affected portion of the brain. Since fMRI is designed to look at changes in blood flow, blood volume, and blood oxygenation, it seemed like a natural tool for exploring the visual auras associated with migraine headaches.

One group [11] solicited subjects by posting a bulletin in their building. After describing the procedures that would be used and obtaining informed consent, the subjects were requested to remain alert for the onset of the visual auras that preceded their migraine headaches as they went about their normal business. If a subject noticed the onset of an aura, they were to come immediately to the MRI room for a brain scan. Any ongoing study in the MRI would be interrupted (there was a sign warning other investigators of this possibility), and the subject with the aura was set up as quickly as possible and scanned. This whole process, however, took many minutes, so the images that were obtained, while intriguing, were very difficult to interpret. The problem was simply that much of the migraine aura story had already passed, and the data had more to do with the complicated sequence of recovery than with the onset of the phenomenon.

It was not clear whether any reportable study could have been completed, until one fortuitous day a volunteer was found who routinely induced migraine headaches via his behavior. The subject knew that, if he played vigorous basketball for a sufficient period of time, he was very likely to have a migraine attack with visual aura in the period shortly afterward. The subject was willing to do this in a planned way. Therefore, the investigators could schedule an imaging session, send the subject across the street (where there was a local gymnasium with basketball court), and the subject would return when he was confident than a migraine attack was likely. In this manner, the investigators managed to collect fMRI data for a period of 30–40 min that spanned the entire subjective experience of the aura: its initial appearance, its development and movement across the visual field, and its final disappearance. The resulting data were the basis for the scientific report [11]. Although other subjects were scanned and their results were compatible with the story that unfolded due to this exceptional subject, the understanding of the data was only likely to have been obtained because of the fortuitous finding of this unusual and exceptionally cooperative individual.

Single Subjects: Required in Practice

There are, of course, several situations where it is only data on individual subjects that is of interest. The most obvious is in clinical evaluations. One does not go to a physician to find out the average location of an epileptic focus across a population. And there are other contexts where testing is used at the level of the individual: from employment applications to the criminal justice system to dating services. The latter domains have not yet become the routine purview of fMRI (nor, indeed, has fMRI become standard in the domain of neurosurgery), but the potential is real.

At present, the vast majority of single-subject fMRI-based work is in presurgical planning. When contemplating a surgical procedure to excise a portion of a patient's brain (to remove a tumor or to remove the focal source of intractable epileptic seizures, for instance), all neurosurgeons perform tests during the surgery to know where, in this particular patient's brain, the control of fine motor movements of the hand and tongue and the processing and generation of speech are located. The reason for this interest is that loss of these abilities are some of the most dreaded by neurosurgical patients. Some neurosurgeons want to have a prospective map of these locations prior to the surgery, and for this purpose, fMRI has become the method of choice. (Memory loss is another important concern during resection of the hippocampus for the treatment of intractable epilepsy, but testing the lateralization of memory function is more difficult in a brief fMRI scan.)

Of course, the surgeons know generally where these areas are, but in many individual cases—sometimes due to trauma, disease processes, or developmental abnormalities—the exact locations of these functions might be moved relative to the expected location. Since the earliest days of using fMRI to study neural activity, various hospitals and laboratories have been exploring this question for presurgical planning. One review of the experience in a single laboratory is given in [14]. It should be noted that while some neurosurgeons want (and even demand) such fMRI-based studies for their patients, it is not yet a medically reimbursable procedure, and only a small fraction of neurosurgeons use it routinely. More information is gleaned during surgery, using techniques such as the direct electrical stimulation of the exposed brain, but the fMRI-based presurgical mapping can facilitate the efficiency of that process. It can also help the surgeon determine the safest physical path to take through the brain to the lesions. In some cases, the presurgical mapping has even helped the surgeon decide whether the operation should be undertaken at all for a given patient.

While the above summarizes the state of routine use of fMRI for presurgical planning, there are exciting extensions possible when electroencephalogram (EEG) recording is added to the picture. It is now straightforward to record EEG signals during fMRI studies [15], and there are a number of ways in which the data can be combined. One case study is particularly instructive.

Most commonly, when a patient suffers from epileptic seizures, it is possible to find some abnormality in the appearance of structural MR images of the brain. However, in some cases, there is no apparent lesion in the structural MR scans. By combining EEG recording during epileptic discharges with ongoing fMRI of the brain, it was possible in one case [16] to locate a presumptive focal lesion. During the time that the patient was scanned, three separate EEG events were recorded. The MRI data could be analyzed by grouping the images according to the presence or absence of those EEG events, and a focal region of activity was found. The report indicated that this focus was successfully removed, and the patient's seizures abated.

Presurgical planning is, by far, the most actively used application of single-subject fMRI in a clinical context. There is hope and speculation that fMRI will eventually become important in other aspects of diagnostic radiology and become part of the trend toward increasingly personalized healthcare. But there are other socially important (and/or dangerous) uses to which fMRI could be put, in the context of evaluating individuals. Indeed, a new academic discipline, neuroethics, is rapidly growing to examine the possible consequences of using functional brain imaging for evaluating criminal cases, studying choice behavior in commercial marketing, informing dating services, and screening prospective employees. Given our current level of theoretical and practical understanding of brain organization via brain imaging, I suspect that serious worry about these issues is premature. But I also have no doubt that the issue will become more important in the future.

Single Subjects: Not a Panacea

Designing studies with a small number of subjects is certainly not a panacea. While issues associated with aligning brain data across multiple imaging sessions are minimal, there are other problems that arise when using the same subject for all studies.

For example, following my inclination to emphasize many experiments with a small number of subjects, David Cox and I scanned each other multiple times over a period of months as part of a process of attempting to define and then refine a particular experimental paradigm and its associated multivariate data analysis. Eventually, we settled on a particular procedure and collected additional data on ourselves as subjects but also on two new, naive subjects. In some general sense, the quality of the data obtained from the new subjects was superior to the quality of data obtained from subjects who had been run on the same or similar tasks many times [17].

Perhaps the new subjects were simply younger and more motivated by the novelty of the experiment; perhaps the senior author (and senior subject, i.e., me) is getting too sleepy in the magnet; but it may well be that repetition of the same stimulus types, in and of itself, is a predictor of adaptation and decreased strength of MR signal. This is one of the dangers of using the same subject repeatedly. Even in tasks where there may not be a decline due to familiarity with a particular cognitive challenge, there could still be a change due to all the variations of familiarity and boredom with the experimental procedure as a whole.

In one remarkable example of using a single subject, one investigator was run in the same three procedures (using a simple visual task, motor task, and cognitive task, respectively) every day for a month [18]. While there were some issues associated with the way the resulting data were presented and interpreted (leading to a reanalysis [19] and discussion of related issues [9]) there was also extensive discussion in the paper of possible Groundhog Day effects (the idea embodied in the movie *Groundhog Day*, in which a person keeps waking up and reliving the same day over and over). This study presented an extreme example of the possible difficulties associated with doing the same task over and over and over (and over) again. And yet, when properly analyzed and displayed, these data were reasonably similar across the many imaging sessions.

By making the (possibly ill-founded, but bold) assumption that the challenges associated with repeating the same or related cognitive tasks are not any worse than the challenges associated with averaging data across different brains, perhaps one can design a new kind of functional neuroimaging study. The following section is such a speculation.

Speculative Proposal for Small-N Research in fMRI

This section is a speculative (almost whimsical) exercise in taking the idea of using a very small number of subjects to its logical conclusion. To wit, instead of doing the same task on a hundred subjects, how could we make sense of doing a hundred different tasks on one subject? The speculation has four components.

What Are the Tasks?

In many ways, this is the most interesting part of the speculation. It is interesting because it is important, and it is physically feasible (albeit politically problematic). We can begin by observing that there is no standard set of tasks for presurgical planning in fMRI. Each investigator has developed or chosen their own set of tasks and used them repeatedly. The fact that there is, as yet, no standard set of tasks across hospitals is probably one of the reasons that presurgical planning using fMRI is still not a reimbursable procedure. On the other hand, it certainly makes sense to use a set of tasks repeatedly in a clinical setting. Physicians and surgeons who receive the output from a diagnostic procedure want that output (and the tests on which the output is based) to be in a highly standardized form. One such list is presented in [14], and it reflects the general consensus of the presurgical planning world that fine motor performance, visual perception, and language skills are particularly important to people—especially people who are about to lose parts of their brain.

In the context of an alternative (and nonsurgical) functional neuroimaging experiment, the constraints are different. To be sure, we could start by asking a few presurgical planners for their favorite tasks. But once we have flickering visual checkerboards, simple finger-to-thumb movements, and picture naming, how do we proceed to get other interesting but reliable and robust paradigms?

The proposed procedure for obtaining 100 critical tasks for use in this study is as follows. List 100 researchers in different areas that have the most publications using fMRI. Ask them to imagine that they are betting their home on the outcome of an fMRI-based study. Let them pick one, two, or three tasks that they guarantee will activate a particular part of the brain or a particular constellation of brain areas. However, do this for one investigator at a time and let them know all the previously selected tasks and the constraint that they have to choose different ones to add to the collection. In other words, they cannot all pick "finger tapping" to save their houses. (On the other hand, they can decline to participate.)

The goal is to find, say, 100 unique tasks that will reliably activate most, if not all, of the brain in various combinations.

In return for becoming part of this wonderful new (and soon-to-be-famous) collection of tasks, each contributor would also have to supply their software and data for running the tasks. (This, of course, will be the trickier political problem. It is one thing to give away the raw data from an individual experiment. It is another to give away the unique means for conducting future experiments.)

Let us assume, now, that we have a wonderful collection of tasks that spans the domains of psychology and behavioral neuroscience. To whom shall these tasks be administered?

Who Are the Subjects?

I would be happy to hear other ideas, but this seems simplest and most appealing to me: Invite four new graduate students to make their brain the topic of their Ph.D. dissertation. The idea would be that they each are imaged once or twice a week for the duration of their graduate training. fMRI might be the primary modality, but EEG, magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), PET, and anything else that might make sense would be considered. Probably, it would be best to have two male and two female subjects. For reasons elaborated below, it would probably be nice to have them at different institutions. It would be best to start with a fifth subject, on whom much of this is done first, to get the bugs worked out of the experimental and imaging protocols before using the four core subjects. Certainly, the subjects should be screened on general reliability and maturity grounds, in addition to the obvious issues associated with MR safety. And no one would claim that these subjects would reflect the average population. Still, I expect that it would be easy to find interested students.

It has been pointed out to me that there will be significant ethical issues that may preclude using students for this hypothetical enterprise. Institutional review boards (IRBs) are unanimous in guarding subjects' right to withdraw from an experiment at any time. Coercion of any form is not permitted, and coercion is very broadly interpreted. (Indeed, at least one IRB asserts that is not permissible to use bold font in advertising the amount a subject will be paid for a study, as it is possible to interpret such bold writing as a form of coercion!) And while the parenthetical example may seem silly, the ethical issues associated with the present proposal are substantial and particularly so for a student. Even though graduate students might be the most interested and motivated to participate in such a study, the fact that their dissertations and graduation would likely be directly connected with their continuing participation would definitely raise ethical concerns. Thus, the subjects may need to be postdoctoral fellows, who are arguably less subject to that form of coercion. But the issue is not trivial for anyone whose career is tied to the study, and that may apply to postdocs too. It may be simpler (albeit less fun) to recruit more ordinary volunteers. (Part of the appeal of using students or other neuroscientists as subjects is that they might give particularly useful feedback and participate actively in proposing and developing new tasks as the project progressed.)

Whoever the subjects are, there are opportunities to add some sociological icing on this cake. For instance, each experiment could be run as a Webcast, with real-time data presentation and preliminary analysis, as is now straightforwardly possible via a number of software packages. Of course, the data would be universally available. The goal, as will be repeated below, is to engage a significant fraction of the neuroimaging community in this project. (Again, however, ethical concerns will have to be addressed regarding the subjects' privacy if the data is to be widely shared.)

How Is the Data to Be Interpreted and Integrated Across the Collection of Tasks?

How the data are interpreted and integrated is, in my view, the crux of the scientific matter. Indeed, when I spoke about these ideas at a recent seminar, one participant commented that his former laboratory had actually tried something like the preceding, using about ten different tasks. They could not find a useful way to integrate the data from the various experiments. Will 100 tasks make the integration process easier?

fMRI data is being generated at an astonishing rate. As a reader of the literature I sometimes feel that the field is drowning in its own success. More than 800 peer-reviewed fMRI-based studies were reported in more than 200 different journals in 2001 [20]. Reviews of a small fraction of this material (e.g., [21] and [22]) present lists of activated brain areas that are difficult to integrate and interpret. The entirety of a recent issue of one prominent journal was devoted to a virtual workshop on meta-analysis of functional imaging data [23]. All of these reports rely on activated sites in a wide range of studies using a standard three-dimensional coordinate system for reporting and emphasize finding the commonality across subjects. Judging the value of these ambitious and effortful attempts to integrate the torrent of data is a highly personal matter.

In addition to the aforementioned journal articles, there are a number of governmentally funded efforts to promote sharing of data and quality control in data collection. Raw fMRI data from all the studies published in another prominent journal are available through the fMRI data center [24]. The National Institutes of Health initiated the Biomedical Informatics Research Network (BIRN) to bring the latest computation tools to bear on the problem of data sharing and standardization, along with funded studies to examine (among many other things) the variability due to different MRI device manufacturers, models, and software implementations. The various BIRN projects deal with far more than just human fMRI data; there are structural and functional data from MR and other modalities, and there are animal data. Ultimately, all of these attempts are explicitly or implicitly aimed at some form of integration and interpretation of the burgeoning mountain of data. But at present, it seems that much of the effort is organized around the important task of making the data available to a wide range of users in both the medical and research communities and working to establish standards for measurement and reporting.

My personal background in neural network modeling and mathematical psychology is somewhat at odds with these approaches to summarizing and attempting to integrate the vast array of data. Perhaps functional brain imaging and systems neuroscience might be likened to the brain of a child—exuberantly growing but in need of structured and developmentally sound pruning to make further development possible. Who knows how we might do that? As indicated previously, various groups are making attempts, but I do not think there is anything like widespread agreement that any one approach is, as yet, dominantly successful. In the absence of a clear answer and in the spirit of this speculative section, I'll offer a somewhat different suggestion for getting started.

One of my favorite texts is *The Fundamentals of Neuropsychology* by Kolb and Wishaw. There are many good books in neuropsychology, but the reason I mention this one is that the authors follow each description of clinical syndromes and symptomatology with the current best theories that clinicians have for explaining these phenomena. Given the right imaging tools, each page could be used to outline a testable theory of human neuropsychology.

Perhaps a collection of such theories could be mapped onto the functional neuroimaging data collected from the proposed small number of subjects. The development of software tools to represent those theories and the mechanisms for incorporating the results of any given study would be substantial challenges. It would require the close coordination and cooperation of software experts with neuropsychology experts. In some ways, it might be analogous to the partnership between a few chess experts and the hardware and software developers who built the best chess-playing computers. The present goal is more difficult, but the idea of bringing the difficult-to-articulate expertise of neurologists and neuropsychologists together with computational experts to represent a substantial collection of models in an extensible way is an appealing idea. Of course, neuropsychologists and computer scientists would just be the beginning. Anatomists and other neuroscientists would have to contribute as well. After all, the goal is to build up a representation of our knowledge of the brain starting with functional imaging data but ultimately including other data. Thus, this model would have to be connected with anatomy and biochemistry in due course. Finally, the tools that would be developed could presumably be applied to the larger body of functional neuroimaging data (i.e., beyond the four test subjects) as well.

How Is the Whole Study to Be Run?

One of the ideas behind this speculation is that it has the potential to engage a broad spectrum of investigators associated with human brain mapping. The subjects themselves could help design new experiments as the basic ones are completed. The new experiments would be driven by the theoretical questions generated when trying to tie the previously collected data to the host of functional models suggested by neuropsychology and cognitive neuroscience and embodied in computational representations. Moreover, as indicated in [23], the vast majority of fMRI data analysis continues to be univariate at a single point in a subject's brain. There are many groups using multivariate techniques for data analysis, and, indeed, the other articles in this special issue present some of those.

Summary

The preceding speculations may not sound very novel to some ears. Indeed, when I started to describe the above ideas to any fMRI researcher who has been involved in the field for a substantial number of years, their response is invariably a comment to the effect that, "Oh yes, we are doing something like that right now in our lab." What I think they mean, of course, is that sometime in the uncertain future they might run all their own existing tasks on a few subjects, if they can get around to it and if they can find the software for some of those earlier tasks. The heart (and hard parts) of the present speculation would be creating a repository of (publicly available) interesting tasks and creating software to embody neuropsychological models and the associated imaging data. At the risk of completely losing my audience, it is the image presented in Isaac Asimov's novel *Second Foundation* for integrating theory and information across many different sources that is the true challenge.

Conclusions

A great deal of energy has gone into the development of computational techniques for aligning the images of brains from different people. The goal has always been to find whatever universal truths may exist about human brain mapping-results that apply to the population as a whole. Perhaps it is time to put energy into a small but focused program of collecting brain imaging data with a very small number of subjects but with a large number of carefully selected tasks. New computational methods for integrating such data will need to be developed, both in the limited context of fMRI and in the more complete context of all available brain mapping technologies. Ironically, it may someday turn out that the information from a few brains, thoroughly studied, reveals more about the universal aspects of human brain function and organization than the current torrent of studies from large collections of brains.

Robert L. Savoy received his academic training in applied mathematics at the Massachusetts Institute of Technology MIT (B.S., 1971, M.S., 1975) and experimental psychology at Harvard University (Ph.D., 1980). This period included ten years of work at Polaroid Corporation's Vision Research Laboratory, after which he joined the newly formed Rowland Institute for Science, under the direction of the late Edwin Land, in 1981. In 1991, he first learned of the revolutionary work being conducted at the Massachusetts General Hospital's Nuclear Magnetic Resonance (NMR) Center, using magnetic resonance imaging (MRI) to detect changes in neural activity (via the associated hemodynamic changes in blood flow, blood volume, and blood oxygenation). In 1993, he joined that group and became the director of Functional MRI Education in 1994. He has conducted fMRI training workshops regularly at the MGH NMR Center three or four times per year since 1994, attracting more than 1,000 researchers from around the world. In addition, he has run similar programs at conferences and at other institutions in the United States, Europe, Asia, and Australia. His fMRI-based research interests are wide-ranging, including temporal resolution of fMRI, stereopsis, language, American Sign Language, decision making, and multivariate analysis. He is the president of HyperVision, Incorporated, a small consulting company located in Lexington, Massachusetts. In this capacity, he continues to teach fMRI workshops at both MGH and other institutions around the world as well as serving as a research consultant for various investigators.

Address for Correspondence: Robert L. Savoy, HyperVision, Inc., P.O. Box 158, Lexington, MA 02420 USA. Phone: +1 781 862 5283. Fax: +1 781 862 5559. E-mail: Robert.Savoy@HyperVison.us.

References

[1] K.A. Friston, A.P. Holmes, and K.J. Worsley, "How many subjects constitute a study?," *NeuroImage*, vol. 10, no.1, pp. 1–5, 1999.

[2] J. Talairach and P. Tournoux, *Co-planar Stereotaxic Atlas of the Human Brain.* New York: Thieme Medical Publishers, 1988.

[3] B. Fischl, M.I. Sereno, R.B.H. Tootell, and A.M. Dale, "High-resolution intersubject averaging and a coordinate system for the cortical surface," *Hum. Brain Map.*, vol. 8, no. 4, pp. 272–284, 1999.

[4] M. Beauchamp, NIH, Bethesda, MD, private communication, Oct. 2004.

[5] M.W.L. Chee, K.M. O'Craven, R. Bergida, B.R. Rosen, and R.L. Savoy, "Auditory and visual word processing studied with fMRI," *Hum. Brain Map.*, vol. 7, no. 1, pp. 15–28, 1999.

[6] M.W.L. Chee, R.L. Buckner, and R.L. Savoy, "Right hemisphere language in a neurologically normal dextral: A fMRI study," *NeuroReport*, vol. 9, no. 15, pp. 3499–3502, 1998.

[7] D. Manoach, "Prefrontal cortex dysfunction during working memory performance in schizophrenia: Reconciling discrepant findings," *Schizophrenia Res.*, vol. 60, no. 2–3, pp. 285–298, 2003.

[8] G.K. Aguirre, E. SZarahn, and M. D'Esposito, "The variability of human BOLD hemodynamic responses," *Neuroimage*, vol. 8, no. 4, pp. 360–369, 1998.

[9] R.L. Savoy, "Experimental design in brain activation fMRI: Cautionary tales," *Brain Res. Bulletin*, vol. 67, no. 5, pp. 361–367, 2005.

[10] K. Cheng, R.A. Waggoner, and K. Tanaka, "Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging," *Neuron*, vol. 32, no. 2, pp. 359–374, 2001.

[11] N. Hadjikhani, M. Sanches del Rio, O. Wu, D. Schwartz, D. Bakker, B. Fischl, K.K. Kwong, M.F. Cutrer, B.R. Rosen, R.B.H. Tootell, A.G. Sorensen, and M.A. Moskowitz, "Mechanisms of migraine aura revealed by functional MRI in human visual cortex," *Proc. Nat. Acad. Sci.*, vol. 98, no. 8, pp. 4687–4692, 2001.

[12] J. Fulton, "Observations upon the vascularity of the human occipital lobe during visual activity," *Brain,* vol. 51, no. 3, pp. 310–320, 1928.

[13] A. Mosso, *Concerning the Circulation of the Blood in the Human Brain*. Leipzig: Verlag von Veit & Co., 1881.

[14] J. Hirsch, M.I. Ruge, K.H. Kim, D.D. Correa, J.D. Victor, N.R. Relkin, D.R. Labar, G. Krol, M.H. Bilsky, M.M. Souweidane, L.M. DeAngelis, and P.H. Gutin, "An integrated functional magnetic resonance imaging procedure for preoperative mapping of cortical areas associated with tactile, motor, language, and visual functions," *Neurosurgery,* vol. 47, no. 3, pp. 711–721, 2000.

[15] L. Lemieux, A. Salek-Haddadi, O. Josephs, P. Allen, N. Toms, C. Scott, K. Krakow, R. Turner, and D.R. Fish, "Event-related fMRI with simultaneous and continuous EEG: description of the method and initial case report," *NeuroImage* vol. 14, no. 3, pp. 780–787, 2001.

[16] G.D. Jackson, R.S. Briellmann, A.B. Waites, G. Pell, and D.F. Abbott, "Functional MRI," in *Handbook of Modern Magnetic Resonance*, C.E. Mountford, Ed. Norwell, MA: Kluwer, 2005.

[17] D. Cox and R.L. Savoy, "Functional magnetic resonance imaging (fMRI) 'Brain reading': Detecting and classifying distributed patterns of fMRI activity in human visual cortex," *NeuroImage*, vol. 19, no. 2, pp. 261–270, 2003.

[18] D.J. McGonigle, A.M. Howseman, B.S. Athwal, K.J. Friston, R.S.J. Frackowiak, and A.P. Holmes, "Variability in fMRI: An examination of intersession differences," *NeuroImage,* vol. 11, no. 6, pp. 708–34, 2000.

[19] S.M. Smith, C.F. Beckmann, N. Ramnani, M.W. Woolrich, P.R. Bannister, M. Jenkinson, P.M. Matthews, and D.J. McGonigle, "Variability in fMRI: A re-examination of inter-session Differences," *Hum. Brain Map.*, vol. 24, no. 3, pp. 248–257, 2005.

[20] J. Illes, M.P. Kirschen, and J. Gabrieli, "From neuroimaging to neuroethics," *Nature Neurosci.,* vol. 6, no. 3, p. 205, 2003.

[21] R. Cabeza and N. Nyberg, "Imaging Cognition: An Empirical Review of PET Studies with Normal Subjects," *J. Cognitive Neurosci.*, vol. 9, no. 1, pp. 1–26, 1997.

[22] R. Cabeza and N. Nyberg, "Imaging Cognition II: An Empirical Review of 275 PET and fMRI Studies," *J. Cognitive Neurosci.*, vol. 12, no. 1, pp. 1–47, 2000.

[23] P. Fox, A.R. Laird, and J.L. Lancaster, "Coordinate-based voxel-wise meta-analysis: Dividends of spatial normalization. Report of a virtual workshop," *Hum. Brain Map.*, vol. 25, no. 1, pp. 1–5, May, 2005.

[24] Functional MRI Data Center [Online]. Available: http://www.fmridc.org

[25] Biomedical Informatics Research Network [Online]. Available: http://www.nbirn.net