What can't functional neuroimaging tell the cognitive psychologist?

Mike P. A. Page

School of Psychology, University of Hertfordshire.

m.2.page@herts.ac.uk

t: +44 (0) 1707 286465

f: +44 (0) 1707 285073

Abstract

In this paper, I critically review the usefulness of functional neuroimaging to the cognitive psychologist. All serious cognitive theories acknowledge that cognition is implemented somewhere in the brain. Finding that the brain "activates" differentially while performing different tasks is therefore gratifying but not surprising. The key problem is that the additional dependent variable that imaging data represents, is often one about which cognitive theories make no necessary predictions. It is, therefore, inappropriate to use such data to choose between such theories. Even supposing that fMRI were able to tell us where a particular cognitive process was performed, that would likely tell us little of relevance about how it was performed. The how-question is the crucial question for theorists investigating the functional architecture of the human mind. The argument is illustrated with particular reference to Henson (2005) and Shallice (2003), who make the opposing case.

keywords: functional neuroimaging, cognitive psychology

Introduction

In the past 15 years or so, the development of functional neuroimaging techniques, from Single Photon Emission Tomography (SPET) to Positron Emission Tomography (PET) and, most recently, to functional Magnetic Resonance Imaging (fMRI), has led to an explosion in their application in the field of experimental psychology in general and cognitive psychology in particular. In this article, I make the case that the huge investment of time and money that has accompanied this trend has not resulted in a corresponding theoretical advancement, at least with respect to cognitive psychological theory. As a consequence, I ask whether the time has come for reflection and reappraisal on behalf of both practitioners and funders.

The relevant functional neuroimaging literature is already of such a bewildering size that it would be difficult for a single person adequately to survey the literature in a single article. I am fortunate, therefore, that Henson (2005) and Shallice (2003) have recently set out their claims for the usefulness of functional neuroimaging in more general terms. Their contributions are extremely welcome in that they allow a theoretical debate to be conducted somewhat independently of debates concerning individual findings. Of course, both have illustrated their arguments with reference to particular studies and I will attempt to do the same, not least in order to question the conclusions that they draw. Nonetheless, I shall try, as they have, to abstract general arguments away from individual claims. In this sense, this article should not be thought of as a direct response to either of these two articles, even though the more general case it makes is structured around the issues that they raise.

Several points should be made very clearly from the start.

First, this article is not about "good" versus "bad" science. Much of the neuroimaging work to which I will refer is above reproach in terms of its scientific credibility. I shall nonetheless maintain that it does not constitute good cognitive psychology.

Second, this article is not about future and/or different technologies. I will concentrate almost exclusively on fMRI technology, since this seems to be the current state of the art in functional neuroimaging. I will not question the usefulness of fMRI as a technique in general, restricting myself to a consideration of its usefulness to cognitive psychology. My arguments may well apply to other techniques such as electroencephalography (EEG) or magnetoencephalography (MEG) but for the sake of clarity I will not consider these here. Nor are these arguments directed towards the next generation of technologies that are (quite reasonably) claimed to be just around the corner. It is doubtless better to wait for such developments to occur before trying to assess them.

Third, this article is not intended to be vexatious, that is, it is not written to annoy functional neuroimagers. It is intended to be a serious contribution to a worthwhile debate. Having said that, it is my intention to express the argument reasonably robustly. In this spirit, I will deliberately make reference, towards the end of the article, to relevant strategic/political matters. This runs the risk of upsetting certain readers but I do not see how the discussion of such an expensive and resource-intensive technology can be separated from its consequences, particularly as those relate to research funding.

Fourth, my intention here is to raise questions about functional neuroimaging as applied to cognitive psychology. Those questions might well have perfectly good answers, though not ones that have yet been given a good airing in print. If this paper draws out such answers in any reaction it provokes, then so much the better. I shall try to give the argument its best shot in the hope of stimulating just such a response.

Finally, it is important to acknowledge from the outset that I am far from being the first to question the contribution of functional neuroimaging to experimental cognitive psychology. Coltheart (2002; this issue) has been particularly erudite in his raising of concerns, and he has drawn on work by Fodor (1999), Harley (2004a, 2004b), Paap (1997), Uttal (2001), Van Orden and Paap (1997), among others. Nonetheless, in Coltheart (this issue) he has concentrated on whether functional imaging has *yet* told us something about the functioning of the human mind; he concluded in the negative, though the majority of his respondents disagreed. In this article I shall take Coltheart's side, but will try to go deeper in investigating *why* the contribution of functional imaging to the study of mind has apparently been so slim and, indeed, whether it might always tend to be so.

Having dealt with preliminaries, it is now necessary to state the basic premise of my argument. It is not particularly radical. It is that a cognitive psychologist's job is to seek to explain, on a functional level, the workings of the (human) mind, that is, to explicate the mind's *functional architecture* (Pylyshyn, 1980; Coltheart, 2002). In a slogan, the cognitive psychologist wishes to know *how* the *mind* works, not *where* the *brain* works. Of course, nobody would deny that there is a link between the two, but my point is that it is not a *necessary* one, at least not at the scale addressed by fMRI. I further contend (contrary to some who might otherwise be in my corner) that learning how the brain implements the mind is both interesting and relevant to cognitive psychology. Nevertheless, I maintain that this latter question will not be satisfactorily addressed by fMRI, except perhaps as a weak enabling technology (see the later discussion of localization). By the time that cognitive models are

sufficiently well specified to be able to make genuinely necessary and differential predictions regarding, say, the Blood-Oxygen-Level Dependent (BOLD) signal, then they will very likely already have enough behavioral clout to be distinguished without reference to neuroimages.

In what follows, I will expand upon this premise with particular reference to Henson's (2005) and Shallice's (2003) arguments and examples, as well as with more general reference to disputes such as those involving single- and dual-route theories, in which functional neuroimaging might, prima facie, be thought to be decisive. I shall start by trying to get to grips with some fundamental issues.

Fundamental issues

fMRI as a dependent variable and the function-to-structure mapping

Proponents of functional neuroimaging in cognitive psychology often start by maintaining, as does Henson (2005), that the BOLD signal in an fMRI experiment is simply another dependent variable, like reaction time (RT) or accuracy, and that the availability of such a extra dependent variable cannot in itself be a bad thing. Of course in one sense they are correct, although the rather obvious counterargument is that there is a very large number of variables that one might *potentially* measure in an experiment (e.g., a participant's body temperature) and that normally, therefore, we would restrict ourselves to measuring dependent variables about which the theories under test have something necessary to say. (This restriction would apply with increased force to the extent to which a particular dependent variable was either difficult or expensive to collect.) Henson is well aware of this rejoinder, and therefore goes to some lengths to justify the assumption that imaging data are indeed such a relevant dependent variable. To this end he develops the notion of a "function-to-structure mapping", maintaining that "functional neuroimaging data are only relevant if there is some systematic mapping between "which" psychological process is currently engaged and "where" activity is changing in the brain" (Henson, 2005, p.196). Even this seemingly innocuous statement raises some pertinent questions, but most of these are more profitably posed in relation to two more specific components of Henson's proposed mapping, namely his "function-to-structure deduction" and his "structure-to-function induction". I shall therefore consider each of these in turn.

Function-to-structure deduction.

Henson's (2005) function-to-structure deduction was expressed thus:

"if conditions C_1 and C_2 produce qualitatively different patterns of activity over the brain, then conditions C_1 and C_2 differ in at least one function, F. The definition of "qualitatively" is

considered in greater detail later but entails a reliable statistical interaction between conditions C_1 and C_2 and at least two brain regions R_1 and R_2" (Henson, 2005, p.197).

Of course, C_1 and C_2 may well differ in at least one function in the absence of any (detectable) differences in activity, and Henson duly noted the fact. That means, of course, that no imaging data are capable, in principle, of contradicting a theory that predicts the engagement of two different functions in conditions C_1 and C_2.

In a previous discussion of this point (Page, 2004), I used the example of a tonotopic map. In such a map (for which there happens to be respectable evidence), different parts are auditory cortex are activated in response to tones of different frequencies, such that the relationship is relatively systematic. If we take the presentation of a low tone to constitute condition C_1 and the presentation of high tone to constitute condition C_2, then following Henson's logic we are encouraged to use the different spatial distributions of activation in support of any (arbitrary) theory that proposes that the detection of different frequencies is accomplished by processes that "differ in at least one function". (Note that this argument would hold even for tones differing by a just noticeable difference, provided that our functional imaging was of high enough resolution.) Given that frequency detection is being performed in each task, it is not obvious, however, what this different function might be. Henson addressed this issue, suggesting that although the deduction would indeed be misleading if one were to treat frequency detection as the function in question (because it suggests two functions where only one is present), the deduction would be sound if one were to treat high-frequency detection as a different function from low-frequency detection. This does indeed save the deduction, but at an unreasonable cost: after all, if conditions C_1 (low tone) and C_2 (high tone) are functionally distinct *by definition* then the patterns of brain activation are irrelevant, whether they suggest functional distinction or not.

The key point here is that one can imagine trying to choose between two theories, one that predicts the engagement of different functions (i.e., qualitatively different mechanisms) for tones of different frequencies and one that does not. What seems awkward from Henson's (2005) point of view is that the functional imaging can, if anything, lead one to make the unwarranted inference of two distinct functions. He suggests that one can avoid this error, at least for the specific case of the tonotopic map, by specifying in detail the precise neural mechanisms under consideration. Scanning might be decisive, he claims, in choosing between a pitch-detection mechanism that involves neurons tuned to respond to particular frequencies

such that those neurons are arranged systematically into a spatial map, and an alternative that has pitch montonically related to the activation of a single set of neurons (i.e., not systematically arranged). He may be right (it depends on the bridging assumptions) but his example implies that for scanning to be effective in choosing between two theories in general, the full neural implementation of each will need to be spelled out in advance, in terms of activation and spatial distribution. For most cognitive-level theories, this will not be remotely practicable.

Generalizing away from the tonotopic map, and assuming a broadly materialist position, it is a logical necessity that any two stimuli that give rise to different percepts or behaviors, that is, any two stimuli that are in any way discriminable, must give rise to different patterns of brain activity. After all, the brain is not in the business of performing impossible discriminations. It is overwhelmingly likely, therefore, that very many stimulus pairs will give rise to activation patterns that are qualitatively different in the strict (statistical) sense required by Henson's (2005) definitions. To be sure, there is a question regarding whether one's current scanner will be sensitive enough to detect the difference. Assuming that it is so sensitive, the function-to-structure deduction seems to propose a proliferation of distinct functions, one for each discriminable stimulus-pair. By extension, as technology improves, and as scanners become more sensitive to activation differences at smaller scales then the number of hypothesized functions will proliferate further. This is surely absurd: that is just not the way a cognitive psychologist attributes functions. Functions are hypothesized on the basis of a (potentially incorrect) theoretical analysis of the task at hand. To reiterate, it would seem that for such a functional decomposition of a task to license predictions about activity patterns at the scale appropriate to, say, fMRI, each corresponding cognitive-psychological model would have to be specified not only in terms of the neural hardware with which it is hypothesized to be implemented, but also in terms of the *necessary* spatial distribution of that hardware in the brain. Note that this is not the same as having a computational or connectionist (neural network) model of the relevant function: such models are rarely either specified in terms of actual neural hardware or specified spatially (as opposed to topologically). If our cognitive-psychological models don't make necessary predictions about the spatial distribution of their functional parts, then how can we use spatial distribution of neural activation to choose between them?

Perhaps I'm being too literal in my interpretation of Henson's (2005) function-to-structure deduction. One might say that my tonotopic map example is likely to involve far too "local" a

region-by-task interaction to license functional conclusions. That is to say, a region-by-task activation will be the more impressive if it involves regions that are more widely separated in space. In fact, Henson makes no such stipulation, but his examples do tend to show something of the sort. So should a nonlocal region-by-task interaction in BOLD signal be considered good evidence for the engagement of different functions? There are several reasons to be cautious. First, there is the problem of epiphenomenal activity, that is, brain activity that is a consequence of task-related processing but that can be considered a nonfunctional byproduct *with respect to the processes under consideration*. A couple of examples should suffice.

Seron and Fias (this issue) discussed the role of the intraparietal sulcus (IPS) and noted its apparent involvement in the representation of abstract number semantics. Given this involvement they proposed that activation of IPS might be used to settle a dispute about the transcoding of numbers. Briefly, this dispute concerns whether the activation of abstract number semantics is a *necessary* component of, say, the transcoding of Arabic numerals to spoken output. Among the problems that Seron and Fias themselves identify with using IPS activation as a gauge of abstract-semantic involvement, is that such representations might well activate during the transcoding task even though such activation is not *necessary* to perform the task. Since the debate is precisely about necessity, any semantic activation that is observed in a neuroimaging study can be considered epiphenomenal according to one theory. That theory will survive any number of demonstrations that the relevant area is "active" in the transcoding task.

As a second example among many, Pulvermüller (1999) has suggested that, for example, the cortical representations of action words and perception words will differ in spatial distribution. Thus action words like "run" will be associated (via Hebbian learning) with neurons in motor, premotor and prefrontal cortices, whereas perception words like "rough" will be associated with cortical areas representing somatosensory qualities. Pulvermüller maintains, among other things, that the auditory presentation of each word will activate these associated areas, leading to distinct patterns of distributed activation for each. And yet nobody interested in, say, the lexical access function would use this region-by-word interaction as evidence that the processes of lexical access were themselves different in the two cases. This is particularly the case given the fact that the two words "run" and "rough" will likely compete during lexical access owing to their phonological overlap.

Apart from the problem of epiphenomenal activity, there is an additional problem for the epistemological status of a region-by-task interaction. That concerns whether the engagement

of two different regions, even regions well separated in the brain, *necessarily* implies two different functions or, more particularly, two functions that differ in type. It is worth considering two cases.

In the first case, the regional interaction in activity would be accompanied by some qualitative difference in behavior. In this case, any inference that can be made on the basis of the brain activity seems rather superfluous: if there is a demonstrable qualitative change in behavior then there could hardly be *no* difference in brain activity (whether or not such is detectable). Whether such a difference represents a change in the *type* of processing will not itself be indicated by the corresponding neuroimages.

In the second case, any regional interaction in activity across conditions would not be accompanied by a qualitative change in behavior. Considering this case, though, one could hardly hope for better evidence that the same function could be performed in two different regions (assuming one had ruled out epiphenomenal activity as a possible explanation). One might, of course, resist this explanation and conclude that one's behavioral measures were not sensitive enough to detect the presence of two qualitatively different types of process. Yet such a finding would be purely adventitious: It's difficult to see why one would design an imaging study to investigate whether two types of process were operative in a task for which the behavioral evidence suggested no such decomposition. Perhaps such studies are common, in which a function that has previously been considered unitary is shown to engage qualitatively different brain areas under circumstances in which both epiphenomenal activity and a one-to-many function-to-structure mapping can be ruled out. Perhaps such studies routinely lead to the adoption of new behavioral measures sensitive enough to capture the processing difference. If so, then I have no doubt that respondents to this article will give details. Having said that, if the usefulness of cognitive neuroimaging is limited to such cases, one would want to be persuaded that the results were worth the investment.

Both these considerations, of epiphenomenal activity and of one-to-many function-to-structure mapping, might seem rather like special pleading, were it not the case that they both crop up in relation to one or other of the examples that I discuss below. Before I consider those examples, I turn to Henson's (2005) structure-to-function induction.

<u>The structure-to-function induction.</u>

Henson's (2005) structure-to-function induction was stated thus

"if condition C_2 elicits responses in brain region R_1 relative to some baseline condition C_0

and region R_1 has been associated with function F_1 in a different context (e.g. in a comparison of condition C_1 versus C_0 in a previous experiment), then function F_1 is also implicated in condition C_2", (Henson, 2005, p. 198).

Again, this raises several questions. First, how was function F_1 associated with the region R_1 in the first place. Any straight comparison between activation patterns in conditions C_1 and C_0 will result in some regions for which the activation pattern differs reliably (provided resolution is good enough) and other areas where the activation patterns are not reliably different, though they will very likely differ "numerically". As Henson acknowledges, any of these regions might be the locus of the hypothesized functional difference between the two conditions. Second, the structure-to-function induction seems straightforwardly to discount the possibility that the same brain region (loosely defined) might perform different functions depending on the task at hand. Note that this does not require any commitment to its being the same neurons involved in both cases: a single region might contain two spatially interdigitated functional systems with no neural overlap. It may also be that the exact same neurons activating in different configurations, or bound in different ways, might implement two or more different functions. Henson himself refers to the studies of Duncan (2001) in which networks of frontal-lobe neurons, respond quickly to changing task demands, presumably by some sort of functional reconfiguration. Whether such functional reconfiguration involves the same neurons or, instead, a switch between interdigitated subnetworks, there doesn't seem to be much a priori support for a one-function-per-region assumption, particularly at the resolution of fMRI.

The third question regarding the structure-to-function induction is one that raises a fundamental question about the BOLD signal. Henson (2005) induces the "implication" of function F_1 from a "response" in region R_1. But what if the experimental condition C_2 requires the suppression of function F_1? Would this also show up as a response in the region R_1 with which that function is associated? Or suppose that in condition C_2, neural hardware in region R_1 attempts to implement processes that are associated with function F_1 but does not, for whatever reason, generate a response that subsequently affects some to-be-explained behavior (a version of the epiphenomenal activity problem). Is it possible that this will also result in activation in region R_1, from which the successful operation of function F_1 might be incorrectly inferred?

In order to address this question, it is necessary to look carefully at what exactly the BOLD

signal indicates. The classic (although relatively recent) enabling texts in this regard are the article and companion book chapter by Logothetis and colleagues (Logothetis et al., 2001; Logothetis, 2003). In their pioneering work, they measured the BOLD response while simultaneously performing intracortical recordings of neural signals, in particular taking measurements of local field potentials (LFPs) and single- and multi-unit activity. Given its frequent citation, you might think that this work would be fully supportive of the scanning enterprise. In fact, there are important caveats. The first of these is expressed very clearly by Logothetis et al. (2001):

"These findings suggest that the BOLD contrast mechanism reflects the input and intracortical processing of a given area rather than its spiking output." (Logothetis et al. 2001; p.150)

and

"The present findings also imply that the greater portion of the haemodynamic signal changes reflect the energetically expensive synaptic activity such as that related to the LFP signals. Both our physiological measurements and the spectroscopy results are incompatible with models suggesting a quantitative relationship between the spike rate of neurons and the haemodynamic response" (Logothetis et al., 2001, p.154).

The point is reinforced by Logothetis (2003) who summarizes

"the BOLD signal primarily measures the input and processing of neural information within a region and not the output signal transmitted to other brain regions." (Logothetis, 2003; p.62)

These quotations quite clearly illustrate that when one sees differential activation in a given region of the brain, one cannot conclude that that region, and any of the (possibly multiple) functions it implements, are "implicated" in any behavior-generating process. Neuroimagers assume that when a brain part "activates" it is likely to be functionally engaged in a task. This work of Logothetis and colleagues shows that a brain region can activate according to the fMRI measure without producing any outputs. In the absence of outputs it is highly unlikely to be driving behavior, that is, it is unlikely to be functionally engaged. Moreover, input to an area, and processing within it, are almost certainly necessary for the (active) disengagement of that region's function. The clear implication is, therefore, that such disengagement will also show up as an fMRI signal. In summary, it appears that activation of a given region in an fMRI scan might imply either functional engagement, functional disengagement, or some modulation in between. To infer only one of these is unacceptable.

There's another aspect of Logothetis and colleagues' work that might give some cause for

concern regarding the scanning project. This is illustrated by the following quotation:

"In all of the measurements, the signal-to-noise ratio of the neural signal was an average of at least one order of magnitude higher than that of fMRI signals. This observation indicates that the statistical analyses and thresholding methods applied to the haemodynamic responses probably underestimate a great deal of actual neural activity related to the stimulus or task, and suggest that a certain degree of caution is called for when interpreting mapping studies, particularly when precise localization of activity is required." (Logothetis et al., 2001, p.154)

We are all used, by now, to seeing images of brains on which a particular region has been colored to indicate that it activated reliably differently in two conditions. This quote quite clearly implies that what is indicated in such images is a subset, perhaps even a very small subset, of the brain regions at which significantly different neural signals were present in those conditions. The uncolored regions represent in graphical terms the implicit assertion of a null result, where the relevant statistics have been performed on a signal that is between ten and one hundred times noiser than the (possibly functional) neural signal itself. What are we to make of such images? Henson (2005) is rigorous enough to admit that we cannot make anything in the absence of at least statistically reliable condition-by-region interaction in activation, but it would be interesting to know whether all imaging results are reported so fastidiously. My strong suspicion is that they are not.

<u>Summary of the function-to-structure mapping.</u>

What implications does all this have on the status of the structure-to-function mapping and its application to the development of theory in cognitive psychology? It appears to imply that virtually nothing concrete can be inferred from BOLD activation patterns with regards to the operation, suppression or modulation of various putative functions. Unless at least two candidate psychological theories are each accompanied by a precise neural mechanism (not just one from a variety of equipotent options), that specifies the necessary spatial arrangement of its components (rather than their topology) and explicitly describes which systems are functional, which are modulated and which are activated but nonfunctional (perhaps even suppressed) in a given task, then it's difficult to see how, say, fMRI images can be helpful in choosing between them. Of course, if theories are specified in such detail, it seems likely that they would make some differential behavioral predictions too. If they did, of course, there would be little or no point in doing the neuroimaging.

This last point highlights why imaging data are not just another dependent variable like RT or accuracy. Even though behavioral data don't *always* permit the falsification of all but one

theory, they are at least in principle capable of falsifying theories that make behavioral predictions. In many cases, therefore, observed behavior is the thing about which cognitive-psychological models do (and should) make predictions. That is not to say that observed behavior is the only thing about which theories can make predictions. There is a long history of the use of physiological measures (e.g., galvanic skin response, pupil diameter) in cognitive psychology. Nonetheless, the quality of the theoretical inferences that can be made on the basis of such measures is heavily dependent on the strength of the bridging hypotheses that relate them to the engagement of particular psychological functions. The BOLD signal as measured by fMRI is not an aspect of behavior, so any functional inferences drawn from it must be grounded on strong bridging assumptions. Henson's (2005) function-to-structure deduction and his structure-to-function induction have been formulated to play precisely this role. In this section, I hope to have demonstrated that they are in no way strong enough to bear the inferential weight that Henson and others wish to place on them.

Localization

Setting aside for a moment the incommensurate nature of psychological models on the one hand and the BOLD signal on the other, it has been asserted that fMRI (among other methods) might be able to locate (rather generally) particular functions to particular parts of the brain. Of course, this would presume some use of a structure-to-function mapping, but it might be a rather coarse-grained one, not itself directed towards choosing between rather similar but competing psychological theories but instead focused on a more general mapping of classes of theories to broadly defined brain areas. I absolutely concede the point and take as an example the paper by Indefrey and Levelt (2004), in which the authors presented a comprehensive meta-analysis of the imaging literature related to word production with the aim of specifying the brain areas that subserve the various components of this multistage task. Nonetheless, no matter how meticulous the contribution to the broad localization of function, I cannot concede the more general contribution to cognitive psychological theory. This is because, taking the Indefrey and Levelt example, all of the imaging data were interpreted by reference to a single model of word production, namely that of Levelt et al. (1999). Nothing in the meta-analysis could, therefore, have been used to disconfirm the predictions of the model. (Coltheart, this issue, makes essentially the same point regarding papers by Winston et al., 2004, and Smith and Jonides, 1997.) The Levelt et al. model is one of a class of models that assume a more or less common set of processes underlying speech production. Again, to illustrate my point, I cannot do better than to quote the authors themselves

"The theory explicates the successive computational stages of spoken word production, the representations involved in these computations, and their time course. The results of the meta-analysis, however, do not hinge on this particular choice of theory, since differences between the sequential LRM model and other models of word production...do not concern the assumed processing levels but the exact nature of the information flow between them. The method and design of the neuroimaging experiments analyzed here were not suited to identify these rather subtle differences between current models." (Indefrey & Levelt, 2004, p.102)

To which I would do no more than add that it is not clear that the method and design of any (practical) neuroimaging experiment would have been so suited.

The point should be clear: Cognitive theory is not advanced by the localization of function per se. As Fodor (1999) put it with characteristic vividness,

"If the mind happens in space at all, it happens somewhere north of the neck. What exactly turns on knowing how far north? "

One further illustration: Suppose that you receive an spreadsheet from a trusted colleague, listing in one column all the functional stages of a highly worked-out model of, say, word production, and in the adjacent column a corresponding list of the brain areas at which each of the functional stages had been reliably located. You might indeed be impressed at the accomplishment. Now suppose the next day you receive an apologetic note from the same colleague indicating that the columns in the previous day's spreadsheet had been incorrectly aligned, and that the corrected mapping was now attached. What exactly would change regarding your appreciation of the functional architecture of word production from one day to the next? Nothing, I would suggest.

There is, I think, one proviso. Earlier, I referred to the possibility that functional neuroimaging might play a role as an enabling technology in discovering the way in which the brain implements the mind. To be more specific, it may be the case that once particular functions have been approximately localized in the brain (notwithstanding the analysis in the previous sections), then other technologies (such as single-cell recordings or whatever) might be used to study detailed properties of the particular brain areas so identified, in order to pin down the particular mechanisms at work. I certainly don't rule this out. Nonetheless, there are still some hurdles to be negotiated. First, if the follow-up technologies are invasive, then there will ipso facto be ethical problems associated with their use. This is particularly problematic with regard to cognitive psychology, since many research areas of cognitive psychology (e.g., study of language, reasoning, emotion, etc.) specifically require the use of human, as opposed

to other animal, participants. Second, the strategy may run up against the fact that the structure of neocortex is approximately invariant across different areas. This would tend to imply that knowing about the neocortical localization of a particular function will not in itself furnish strong constraints on possible mechanism. This might be too pessimistic an assessment: the different Brodmann areas (Brodmann, 1909) into which the neocortex is often parcelled were originally delineated on the basis of relatively subtle differences in cytoarchitectonic properties. If those cytoarchitectonic properties are sufficient to constrain possible mechanism, and, crucially, if the resulting constraints on possible mechanism are themselves sufficient to permit the distinction between rival *psychological* theories (presumably expressed in some form of connectionist model), then the locating of particular functions to particular area of cortex might permit some rather indirect leverage on the functional architecture of the human mind. I must admit, though, that it seems something of a long shot.

Perhaps everyone can agree that if this is the way functional localization is to come to the aid of cognitive-psychological theorizing then we should start insisting on some more obvious efforts in that direction. Dispiritingly often, imaging papers go to great lengths to establish the locus of a particular function, only to leave it at that. For example, a recent paper by Botvinick, Cohen and Carter (2004) makes a good case that the activation of the anterior cingulate cortex (ACC) has something to do with the psychological process of conflict monitoring. It is far from clear, though, how this advances theorizing about the process itself. Tasks, such as the Stroop (1935) task, are described as involving conflict not because they activate ACC but because that is the way the behavioral effect has best been characterized. No facts about the activation of ACC could reasonably have been supposed to argue against the existence of some sort of conflict in the Stroop task; and no cytoarchitectonic facts about ACC are brought to bear on the plausibility of various psychological models of conflict monitoring. Botvinick et al. present the results of neural network simulations, but these too seem to be independent of any facts about ACC or its activation, being driven, quite appropriately in my view, by the requirements of the behavioral data. If localization is going to enable theoretical development via a close consideration of the constraints imposed by regional differences in cytoarchitectonic structure, then let us have the claims made explicit so that they can be straightforwardly assessed.

Some examples from Henson (2005)

Having set out some of the principal theoretical issues, I will now discuss some of the examples that Henson (2005) uses to illustrate his case. I will try to show that these are less convincing than he maintained. To avoid the necessity to reproduce large numbers of color

plates here, I shall refer to Henson's figures directly.

Recognition memory

The first of Henson's (2005) examples involves recognition memory. He briefly discusses such memory in relation to the Remember/Know (R/K) distinction introduced by Tulving (1985). In this paradigm, participants are presented with a series of stimuli in an exposure phase; in a later test phase they are presented with further stimuli, some of which are "old" in the sense that they were presented in the exposure phase, and some of which are "new". The task is to distinguish between the old and new stimuli and, further, for those items described as old, to distinguish between decisions based on recollection of the prior exposure ("Remember", R) and those based on a sense of familiarity ("Know", K). Behavior in such tasks has been accounted for using "dual-process" models (e.g., Yonelinas, 2002) and so-called "single-process" models (e.g., Heathcote, 2003; Donaldson, 1996). Henson (see his Figure 3a) described a reliable cross-over interaction in activation between left inferior parietal cortex, whose event-related BOLD response is strong for R judgements and weaker for both K judgements and "New" (N) judgements, and right dorsolateral prefrontal cortex, that activates strongly for K responses, less strongly for N responses, and less strongly still for R responses. This pattern of results led Henson et conclude that "the imaging data support dual-process models over single-process models".

There is much to be said here. First, a warning regarding a potential confounding between vehicle and content: there is no good reason to expect a continuum of memory strength to be reflected in a continuum of activation strength; this would be to commit a vehicle-content error. We do not expect actvations corresponding to memories of red things to be red, so we cannot expect strong memories necessarily to involve strong activations. It is not clear whether Henson (2005) himself commits this solecism, although he comes close when he claims that "it would be difficult to explain these imaging data in terms of purely quantitative differences in memory strengths" — as if that is the way they would have to be explained. Second, there's something funny about the data: if activation of right dorsolateral prefrontal cortex is something to do with a measure of, or a judgment regarding, familiarity (which is how Henson characterizes it both here and elsewhere), then why is that region activated less for R responses (which are presumably familiar as well as recollected) than K responses and even N responses (the latter being not familiar at all). The pattern more naturally suggests that the qualitative difference in activation between R and K responses is more to do with the responses themselves rather than any memory-strength representation or judgment. But if the imaging data are telling

us that there is something different about R and K *responses*, that is something we already knew. After all, for R responses participants say (and mean) "Remember" and for K responses say (and mean) "Know"! Did anyone really believe that there was *no* difference in brain activation associated with participants issuing demonstrably different responses?

In fact, Henson (2005) himself (and Coltheart, this issue) dismantled the argument derived from the imaging data for dual-process over single-process models. He pointed out that

"in subsequent experiments, we interpreted the right dorsolateral prefrontal response in terms of post-retrieval decision processes rather than familiarity or retrieval strength...and have since associated familiarity with response reductions in anterior medial temporal cortex..." (Henson, 2005, p.202)

Oddly, anterior medial temporal cortex didn't get a mention in relation to the original experiment, even though the subsequent experiments apparently fingered it as the site of familiarity-based reductions in activation. We are not told whether R responses were also associated with such reductions, but reference to Henson et al. (2003) makes clear that the familiarity signal is "acontextual". This supports the new localization of the familiarity signal to the anterior medial temporal cortex (for both R and K responses), but actually rather weakens Henson's argument that qualitatively different R and K processes were at work. This is because the original significant crossover interaction between task and region has now been replaced by what we presume to be a noncrossover interaction. Although Henson (2005) was careful not to require a crossover interaction in activation before inferring a qualitative difference, there is no doubt that the noncrossover pattern is more compatible than the crossover with different brain regions having rather different, though monotonic, response functions to a single underlying memory continuum.

And it gets worse, because it turns out that so-called single-process models are not single-process at all, at least not in the sense that Henson (2005) implies. Taking Donaldson's (1996) model as an example, it turns out that Donaldson's theory is that R and K *responses* are based on two thresholds applied to a single continuum of memory strength. Memory strengths that exceed the lower threshold but not the upper are designated K responses, with those exceeding the upper threshold designated R responses. Given that there are two thresholds, and thus two threshold comparison processes (not to mention two different types of response, attitude, etc.), on what basis is such a model described as a single-*process* model? And in what way is it obviously inconsistent with one brain region that responds to old stimuli whether they are known or remembered (i.e., exceeding the lower threshold), and another brain region that only

responds when they are remembered (i.e., exceeding the higher threshold)?

In fact, even if you consider Donaldson's dual-thresholds to be implemented by a single process, whose operation and consequences are constrained to within a single brain area, Henson's (2005) logic *still* doesn't hold. Why? Because Donaldson doesn't even deny the existence of separate processes associated with recollective memory. As he puts it,

"I am not claiming that there is no distinction to be made between recollective and nonrecollective memory, only that this introspective technique [i.e., the remember/know technique] fails to capture the distinction...to deny that people can sometimes recollect encoding details, or that there are occasions when they cannot, would be silly." (Donaldson, 1996, p.532, my parentheses)

Heathcote (2003) makes a related point, the basic observation being that neither is denying the possibility of recollection. Rather they are discussing whether and how such recollection affects the *behavior* exhibited in the remember-know task. But if, say, Donaldson, a so-called single-process theorist, is quite content that both recollective and familiarity-based processes are taking place, but with only one of them having an effect on R/K behavior, then in what way could the imaging data help to decide the issue relative to such behavior? This is another example of the epiphenomenal activity problem.

<u>Memory encoding and the subsequent memory paradigm</u>

The second example Henson (2005) gave concerned the subsequent-memory paradigm, in which participants are presented with several stimuli on which they perform a simple task while being scanned. They are specifically not invited to try to remember the stimuli. Following a later memory-test phase, during which participants indicate which stimuli they remember, the original scans are backsorted into those deriving from the presentation of stimuli that were subsequently forgotten, and those from subsequently remembered stimuli. The context in which Henson discusses this paradigm is in relation to two theories of memory encoding. I will use his words, so we can be quite clear about the claims he makes:

"According to what I shall call "structural" theories, there exists a cognitive system specialized for episodic memory...According to "proceduralist" theories on the other hand...memory is better viewed as a by-product of the processes performed when a stimulus is encountered....An important difference between these two types of theories is whether successful remembering always involves a specific psychological process, supported by a specialized memory system, or whether remembering can be associated with different processes on different occasions. This can be tested by comparing subsequent memory effects under

different study tasks: according to the structural theories (T_0), the brain regions correlating with subsequent memory should not differ across tasks, whereas according to procedural theories (T_1), they should." (Henson, 2005, p.202)

Before tackling the questionable logic of this statement, I shall briefly describe the imaging results. Otten et al. (2001) had backsorted scans according to subsequent memory, for words that had either been semantically processed or orthographically processed at study. They had found that for both encoding contexts there was a correlation between left medial-temporal-lobe activity and subsequent memory. This was taken to support a structural view. In a different experiment, Otten and Rugg (2001) compared semantic and phonological encoding-contexts and found "a region within anterior medial frontal cortex that showed greater subsequent memory activations under the semantic task, and regions within left intraparietal sulcus and superior occipital cortex that showed greater subsequent memory activations under the phonological task". This was taken to support the procedural view.

The ambiguity in these results only goes to highlight the faulty underlying logic. First, there is no necessary incompatibility between the structural and proceduralist theories: subsequent memory might depend on both stimulus-specific processes *and* general memory processes; indeed, that seems rather likely a priori (by which I mean, given previous behavioral data). Neither imaging result ruled out either theory; neither did the imaging results particularly support one over the other. To be explicit, although Otten et al. (2001) had found a common brain region that correlated with subsequent memory for differently encoded stimuli, they couldn't plausibly have asserted that it was the same neurons that were involved each time. Of course, if memory networks are simply defined by where they are generally in the brain (e.g., medial temporal lobe, MTL) then one is tempted to make this identity, but it might just as well have been one set of MTL neurons whose activity correlated with subsequent memory for semantically encoded material, and another set whose activity correlated with subsequent memory for orthographically encoded material. In which case, the memory systems involved would have been "procedural all the way up", not withstanding the incorrect inference from the imaging data.

Likewise, finding that the activation of different regions correlated with subsequent memory for different encoding contexts is hardly conclusive. Did anyone seriously maintain beforehand that exactly the same neurons were involved in exactly the same way for doing semantic judgments on one hand and phonological judgments on the other? Given that, did anyone

seriously entertain that *nothing* about the relevant processes would differ for subsequently remembered items as opposed to subsequently forgotten ones? The only interesting question seemed to be whether or not the fMRI scanner was sensitive enough to spot whatever it was that differed. Even given that the scanner was able to spot activation differences, there was no evidence regarding which, if any, of these activation differences were causally related to the *memorial* process at encoding. To expand upon this point somewhat, suppose a given trial induces epiphenomenal activation whose strength is correlated (for whatever reason) with the activation of memorial processes that are genuinely causally related to later performance. How do we know which of the regions whose activations are correlated with subsequent memory, were involved with *memorial* encoding at the time, as opposed to the other epiphenomenal (but causally unrelated) processes? The point is rather reinforced by the fact that Henson (2005), in his summary of Otten and Rugg (2001), only mentions two of the several regions whose activation correlated with later memory in the phonological task. What of the other regions? Was their activation considered epiphenomenal? And how might one tell?

To summarize the second of Henson's (2005) examples, proceduralist theories "predicted" one thing, structuralist theories "predicted" another; as it turned out, either both things happened, or they didn't, or it was some mixture of the two.

Short-term memory

Henson's (2005) next example used experiments on short-term memory to illustrate the operation of the structure-to-function induction. He discussed the difference between two visually presented tasks: an item-probe task, in which a list of items is presented followed by a test-item that the participant has to identify as either a member of the preceding list or not; and a list-probe task, in which the probe is another list containing the same items as the first but either in the same or different serial order. These two tasks were combined with the manipulation of the temporal grouping of the six-item stimulus list, which was either presented with a constant stimulus-onset-asynchrony (SOA) or with a long SOA separating two groups of three items. Henson also noted that grouping has been held to reflect the operation of a "timing" signal by some (e.g. Burgess & Hitch, 1999; Brown et al., 2000). The scanning findings, at least the subset of them presented by Henson (2005), are rather uninformative. He drew attention to an area of left dorsal premotor region that was "more active in the list probe than the item probe task, but less active in the grouped than the ungrouped list probe task". Henson claims that this is "consistent with utilization and modulation, respectively, of a timing signal like that proposed by Burgess and Hitch (1999)". Well yes. It is also consistent with

models of serial recall that include a positional signal that is not derived from a timing signal, and would potentially even be consistent with models of serial recall that acknowledge neither a timing nor a positional signal, though no such models exist owing to the overriding constraints placed on models by the behavioral data. In fact, the activation of the highlighted area might conceivably even indicate task difficulty, given that we know very well that grouped lists are easier to recall than ungrouped lists.

It is at this point that Henson (2005) invoked the structure-to-function induction, citing another imaging study in which the "same" region was "more active for sequential rather than repetitive finger movements", and a neuropsychological study in which patients with damage to "this region" had difficulties reproducing rhythmic motor sequences. Neither of these citations, however, does anything to cement the link between that brain region and a timing signal, as opposed to a positional signal, or even sequence representation more generally. One is reminded of the Abraham Maslow's quip, "When all you own is a hammer, every problem looks like a nail". It seems that when one favours a timing theory, one sees its confirmation in a variety of brain measures, even though these are just as consistent with any one of the rival theories.

In summary, Henson's (2005) application of the structure-to-function induction to short-term memory is completely uninformative with regard to the merits or otherwise of the competing theories in the area.

Given that Henson's final example involves reasoning on the basis of MEG data, it is not within my current purview beyond noting that his account suggests that the imaging data were actively unhelpful in shaping his opinions about repetition priming. I will move, therefore to discussing the contribution of Shallice (2003), in which a similar story unfolds.

Shallice (2003)

In his defense of the use of functional neuroimaging in the development of cognitive theory, Shallice (2003) goes further than did Henson (2005) in the assumptions that he considers necessary. In particular, he requires three things to be true:

1. processing is carried out by "isolable subsystems".
2. no [brain] region realizes more than one subsystem per task.
3. "the resource required by a subsystem in performance of a given task can be thought of as being monotonically related to the "local average neural activity" (p. S148) and further that "the relation between specific resource requirement of a task and activation is linear" (Shallice, 2003, p. S149).

The first of these is expressed as a hope: unless it's broadly true (and it might not be) then a good deal of theorizing in cognitive psychology is vulnerable. The second seems to be to be little more than speculative and is additionally heavily dependent on how one defines a brain "region". It is far from clear that the brain is made up of spatially segregated regions (especially at the resolution of fMRI), in each of which a given task "activates" only a single subsystem. The third assumption is one at which Henson himself blanches. Specifically, Henson is happy to conclude that an area is "engaged" in a task, whenever there is differential activation in that area under the relevant conditions. He expressly does not care whether there is a reduction in BOLD signal or, alternatively, an increase; as long as there is a change in BOLD signal then Henson is content to consider the region as being involved, provided his statistical conditions are met. As I've tried to point out above, this leads Henson into inducing behavioral engagement of a function even if its engagement is epiphenomenal or even explicitly suppressed. Nonetheless, Shallice appears to commit the greater vehicle-content error when he assumes a monotonic linear relationship between regional activation and cognitive "resource". The assumption is still the more puzzling because we are given no idea where it might have arisen.

To take an example, suppose one were a proponent of a dual-route model of word reading (of which more later), in which one (functional) route acts on the basis of "regular" spelling-to-sound correspondences, and is able, therefore, to produce regularized pronunciations of words and plausible pronunciations of nonwords. The other route is a lexical route within which the (possibly irregular) pronunciations of words are stored and accessed. In order to explain the relevant behavioral data, all such models have to assume that both routes are activated on presentation of both types of word (regular and irregular) as well as on presentation of a nonword. Given this fact, it would seem relevant to Shallice's argument to ask whether the lexical "resource" would be activated more for a low-neighborhood high-frequency word or a high-neighborhood nonword. I am not in a position to do the relevant simulations, though given Shallice's qualms about relating connectionist models to imaging data (Shallice, 2003, p. S147), it's not at all clear what simulations one might do. Nevertheless, it seems that using as a guide the "local average neural activity" in a lexical route, it would not be in the least bit surprising to conclude that the nonword uses more of a lexical resource than the highly frequent word. These are the sorts of conclusions one is led to make when one mixes talk of "cognitive resources", specified relative to some theory, and the local average activation of brain regions. The situation is not improved by assuming a linear relationship between the two.

Shallice (2003) bolstered his case by reference to the Hemispheric Encoding and Retrieval Asymmetry (HERA) model of Tulving, Kapur, Craik, Moscovitch and Houle (1994). This constituted a claim about the "*relative* lateralization of the encoding and retrieval stages of [verbal] episodic memory task performance" (Shallice, p. S149). As the reader might have guessed, this claim was itself primarily based on functional neuroimaging data of various sorts. There are are several points to make about Shallice's choice of HERA as a worked example.

First, it is rather an odd model for Shallice (2003) to choose if he wished to illustrate the usefulness of imaging to cognitive theorizing. That is because there seems very little of cognitive substance to the claim that HERA embodied. Unlike the vast majority of cognitive theories, HERA expressly drew attention to the neural localization of various processes: encoding (relatively) left hemisphere; retrieval (relatively) right. Nothing about the functional architecture of the system would have been (necessarily) different if the mapping were reversed or, indeed, nonexistent. It was presumably clear from the start that encoding and retrieval were not literally the same process; had it not been, the scanning tasks would have been somewhat tricky to design. It was also presumably clear on logical grounds that encoding and retrieval must engage at least some common brain parts (for a given participant and episode), lest the brain be supposed to function like a stage magician, putting things in one box and retrieving them from another. So what of functional significance was contributed by the hemispheric distinction per se?

Second, as Shallice's (2003) HERA example unfolds, it becomes increasingly clear that the functional imaging results were actively unhelpful in locating encoding and retrieval processes, even for those interested in establishing their locations. I shall not reprise the whole story here but it turned out that, contrary to the most straightforward prediction, patients with right-hemisphere damage were not impaired in their levels of verbal episodic retrieval (Stuss et al., 1994; Swick & Knight, 1996, etc.). Shallice suggested three reasons for the disparity, each of which is illustrative (for my purposes) of a wider issue. His first reason involved noting that "prefrontal cortex" (towards which the HERA claim was directed) is a big place; more specifically he noted that other studies had found that different parts of prefrontal cortex were differently engaged by different retrieval tasks. This straightforwardly questions the wisdom of talking about brain regions as if they were functionally circumscribed. His second point was that although "in many studies the right prefrontal activation is stronger than the left...or it is only the right prefrontal that is significant...in other studies in which activation has been found in right prefrontal cortex in experiments on episodic memory retrieval , a comparable if

frequently lower degree of activation has been found in left prefrontal cortex". Whatever this means with regard to the functional conclusions that can be drawn, it certainly implies that Henson's (2005) strict statistical requirements for activation-by-region interactions were somewhat relaxed in interpretation of the earlier data. Finally, Shallice's third and "most critical" point was that the earlier function-to-structure deduction regarding right-hemisphere retrieval was probably incorrect in the first place! He noted that retrieval had previously been considered as engaging a multitude of functions, including a post-retrieval checking mechanism (Burgess & Shallice, 1996). Shallice is quite clear that this consideration had itself stemmed from a neuropsychological/behavioral observation, namely that patients with right-hemisphere damage tended to repeat themselves in free recall (Stuss et al. 1994). To cut a long story short, right hemisphere "activation" was eventually attributed to this checking mechanism, with later imaging studies settling on the left prefrontal cortex after all as being the site for recollection (though not familiarity judgment, see earlier). At no point that I can see did the imaging results lead the functional decomposition of the task; they were used solely to localize the constituent functions, sometimes inappropriately.

Single- and dual-mechanism models

Having previously delivered various versions of this article in lecture form, I have noted that the distinction between single- and dual-mechanism models is often raised in subsequent discussion. It is therefore worth discussing, albeit briefly. The basic intuition seems to be that functional neuroimaging will be helpful in choosing between single-mechanism models on the one hand and dual-mechanism models on the other. Because they are the most frequently discussed topics of this type, I shall consider the question with reference to regular/irregular past-tense production and, first, to models of single-word reading (spelling-to-sound).

Single word reading

Single-word reading has (on the face of it - see below), been the subject of debates between dual-route theorists, who make a distinction between lexical and nonlexical routes in reading, and single-route theorists, who do not. With regard to the implications of functional neuroimaging, I am fortunate that Coltheart (2002, this issue) has already made many of the principal arguments. Nonetheless, I think they bear developing here for completeness. Coltheart is, of course, in an interesting position here, because as one of the principal advocates of dual-route models in this area he potentially has much to gain from imaging data, if such appear to support his position. Indeed, it seems that he (and those on his side of the debate) have nothing to lose from allowing imaging data into the debate. And that is exactly the point!

Among other things, Coltheart is unwilling to accept any imaging data as being supportive of his position precisely because *no imaging data could contradict his position*. Nothing about the dual-route position mandates that the two routes should be located in different regions of the brain. Indeed given that (in respect of reading) both its lexical and nonlexical routes draw on a common orthographic input stage and drive a common phonological output stage, it might be supposed, a priori, that the two routes would be physically close, even interdigitated. No imaging result that showed the "same regions engaged" in the reading of exception words and nonwords, could possibly be interpreted as contradicting either a dual- or a single-route model.

But what if neuroimaging suggested two regions, one that "activated" during irregular-word reading, and one that "activated" for nonword reading (or some such comparison). Wouldn't that be good evidence against single-route models (as Henson, 2005, footnote 10 insists) and in favour of dual-route models? Well no, because the dual-route models make no such necessary prediction. As I hope to have made clear above, it is a fundamental feature of dual-route models that both routes are engaged even for irregular words and nonwords; this feature is necessary to explain, for example, the reaction time cost of irregularity and its interaction with frequency. Even a putative dual-route model that offered some other explanation for these behavioral effects and that posited, say, that an irregular word suppressed the sublexical route, would have quite arbitrarily to assume that the shutting down of a given route would not lead to metabolic activity in the corresponding brain region, contra the work of Logothetis et al., as described earlier.

As a final gambit it is possible, I suppose, that even if a two-region imaging pattern were not consistent in detail with a dual-route model, some dedicated dual-router might maintain that it was even more straightforwardly inconsistent with a single-route model. Obviously this would be somewhat clutching at straws. In response, I would note that even those spelling-to-sound models that have traditionally been conceived as single-route are actually, as far as imaging is concerned (and perhaps as far as theory is concerned too - but that's another question), more like dual-route models than is usually conceded. Taking the familiar triangle model (Plaut, McClelland, Seidenberg & Patterson, 1996) as representative, it has two functional routes from spelling to sound: one is relatively direct, bridging the orthographic and phonological representations with a single layer of hidden nodes; the other proceeds somewhat indirectly via a representation of semantics. As a consequence, the real distinctions between the "single-route" triangle model and its more explicitly dual-route competitors are whether its "semantic" route includes lexical representations (possibly with semantics "branching off"), and whether

the more direct route, with its absence of explicit lexical representations, maintains an independent ability to read exception words. Neither of these issues is ever likely to be decided by fMRI and, ipso facto, neither is the debate between "single-" and dual-route models of single-word reading.

Past-tense representation

The situation in relation to the representation of the English past tense turns out to be similar (and possibly related). The classical debate has once again been between proponents of dual-mechanism and single-mechanism models. Dual-mechanism theorists believe that mappings between stems and regular past-tense forms (jump → jumped) are implemented by a rule of the form $X \rightarrow X+'ed'$ (with some phonological adjustment), whereas mappings between stems and irregular past-tense forms (bring → brought) are stored by a purely associative process linking the undecomposed lexical forms. By contrast, single-mechanism theorists believe that a single process suffices to explain both regular and irregular mappings. As with the "single-route" spelling-to-sound models, we have to be careful with what is meant here. Probably the most prominent single-mechanism model is that of Joanisse and Seidenberg (1999), with which the authors attempted to account for neuropsychological-behavioral data that otherwise seemed to support a dual-route account. The details are not germane, but Joanisse and Seidenberg showed that similar data could be simulated by their model by virtue of differential damage to two loci: a semantic locus (that was actually lexical); and a phonological locus. Their simulations suggested that irregular past tenses were more adversely affected by damage to an area representing what they called semantic information (that was in fact lexical information). Regular past tenses were more affected by damage to areas representing phonology. Whatever the implications of this model for the interpretation of the neuropsychological-behavioral data (and, rather crucially, for the taxonomy of models into single- and dual-mechanism), it should already be clear that functional neuroimaging is not going to be decisive. Given that both classes of model suggest differential reliance on two flavors of processing (semantic/lexical and phonological in one case, lexical and morphophonological in the other), both seem able to deal with any neuroimaging pattern that might plausibly emerge.

This is potentially an illustration of the problem referred to earlier, namely that which stems from assuming that different regions must implement different functions. Joanisse and Seidenberg's (1999) model involves a single set of "hidden units" that receive input from two separate sets of units labelled "speech input" and "semantics". These hidden units in turn send signals back to the semantic units and, separately, to a set of "speech output" units. It is clear

that it is a topological model rather than a spatial one. Suppose that we take some of the "hidden units" and move them to a place in the brain physically nearer to the representation of semantics. Some other units might be moved to a place in the brain nearer to the representation of speech output. Suppose further that, although the general topology of connections was maintained, the error-term from nearer output units (speech or semantics) affected more strongly the development of connections to and from the corresponding hidden units. In this way, although the model would be topologically single-route, and perhaps even definitionally so (their being no clear distinction between rule-based learning and associative learning), it might show characteristics of a dual-route model if examined solely with respect to patterns of spatial activation.

Perhaps this is why, in some recent spirited comment on this topic, proponents of neither the single-route nor the dual-route view referred to any functional neuroimaging data (McClelland & Patterson, 2002a, 2002b; Pinker & Ullman, 2002a, 2002b, 2003; Ramscar, 2003; Seidenberg & Joanisse, 2003). Instead, the currency in which the debate has been and continues to be transacted is firmly that of behavioral and neuropsychological-behavioral data (e.g., Longworth et al, 2005; Miozzo, 2003; Tyler et al. 2004). In a brief follow-up Marslen-Wilson and Tyler (2003) did refer to preliminary imaging results, though in their reply McClelland and Patterson (2003) insist that "all of the findings are consistent with [their] account" without referring explicitly to the imaging data. Interestingly, Marslen-Wilson and Tyler (2003) conclude their account, that includes reference to a variety of neurophysiological findings, by claiming that they "remain agnostic as to the types of mental computation implicated by these results". This is a slightly odd claim because if the past-tense debate is about anything, then it is about "types of mental computation". Maybe they were drawing attention to the fact that no amount of neuroimaging will indicate the types of mental computation that are being imaged.

Why does it matter?

If I have made the case with any success that functional neuroimaging is unlikely decisively to advance cognitive theory, then the question might be raised as to whether this matters terribly much. After all, I've tried to be clear that there is much good science to be found in the neuroimaging literature. In the latter part of this article I will outline a case that it does indeed matter and, moreover, that the overplaying of functional neuroimaging might have undesirable consequences for the future development of cognitive psychology as a discipline.

The first point to note is that the current vogue for cognitive neuroimaging would be of little negative consequence if the method were cheap to practice and resource-light. Unfortunately it

is anything but. Although the figure is sometimes difficult to obtain, I understand that somebody wishing to run an experimental participant for one hour in a typical fMRI study would not expect to receive much change from £1000 ($ US 1750). This is two orders of magnitude higher than the typical payment made to participants in more conventional experimental settings. Whether these costs fully cover the very high purchase costs and running costs that attend an MRI scanner is also difficult to ascertain: at least some portion of the salaries of additional personnel (e.g., physicists, technicians) that are employed directly by interested psychology departments might be considered to come on top of the per-subject cost. Whatever the exact figure, this represents a very large investment on behalf of psychology departments and, hence, psychology funders (among others).

Second, we should note that large capital investments tend to concentrate subsequent funding around them. Funding bodies that have committed large amounts of money to a capital project are, for understandable reasons, loathe to let that project fail for lack of ongoing funding. Indeed, they may even be tempted to provide lavish funds in an attempt to justify their original decision. Somewhat perversely, this might be especially the case if the original decision starts to look shaky. (I leave any more involved social psychological speculation to those better qualified to indulge in it.)

Both these observations point to a fairly obvious fact, namely that functional neuroimaging does, and will continue to, absorb resources at a relatively high level. Again, this would not be a problem for cognitive psychology per se, if those resources were not coming from a limited pool of research funds from which experimental cognitive psychology must itself draw. But it seems clear that some, if not most, of the money that currently funds functional neuroimaging comes from a pool from which experimental cognitive psychology has traditionally drawn. And here is the nub: if cognitive neuroimaging either hasn't yet paid (Coltheart, this issue), or by its very nature won't pay (above), its way in terms of theoretical advance, then cognitive psychology must indeed suffer the consequences. Lest this be seen as mere scaremongering, I have already heard of at least one example (from the US) in which a grant application in the area of experimental cognitive psychology (i.e., behavioral work) was refused. The applicant subsequently resubmitted the (considerably more expensive) request, with the same behavioral work outlined but with additional neuroimaging proposals in which s/he had no real interest. The second request was met with a positive response. One can only assume that the additional funding allocated to unwanted neuroimaging was withdrawn from some other unlucky applicant.

There is another resource apart from money that may well (have) become concentrated around cognitive neuroimaging, that is, people. It seems that in recent years some (though thankfully not all) of the very best psychology postgraduates and postdocs have "gone into scanning". No doubt this is sound career move, particularly if my analysis of the monetary resource issue is anything like correct. However, it does risk the impoverishment of more traditional experimental cognitive psychology as a field. Even if those attracted into neuroimaging still think of themselves as cognitive psychologists, the large amount of extra time and knowledge that must be invested in their new trade can only detract from the theoretical contributions they might otherwise have made.

To be successful as a field, I maintain that cognitive psychology needs a readier supply of "mindscanners" than of brainscanners. I take a mindscanner to be an individual trained in (cognitive) psychological theory, able to identify areas of significant theoretical dispute, capable of designing, running and analyzing appropriate experimental tests, and willing to relate quantitatively the results of those tests to the predictions of the relevant theories (perhaps embodied as computational and/or connectionist models). Even if readers are more optimistic than am I regards a significant role for functional neuroimaging in the development of cognitive theory, it should still be clear that a secure supply of mindscanners is absolutely necessary if there are to be theories to test, tasks to scan, and inductions to be made.

This appeal to the redirection of resources from machinery to people might seem rather sentimental. I anticipate that the word "Luddite" will not be far away from the lips of my critics. Perhaps so. But the analogy is only apposite if one refuses to accept the arguments developed above. After all, the Luddites were protesting against the introduction of machines that were (apparently) less expensive and more efficient at weaving cloth than were they. If cognitive theory can be taken to be the cloth that cognitive psychologists are attempting to weave, then it is precisely my case that functional neuroimaging is, in this regard, *more* expensive and *less* efficient. The exact reverse of Luddism.

I have two further points to make, and both reflect on the status and image of cognitive psychology among the general public. Seldom a week goes by without their appearing on, say, the BBC news website, some reference to a functional neuroimaging experiment. I'm sure the same can be said for media outlets in other parts of the world. In most cases, the article is accompanied by a picture purporting to show a part of the brain "activating" in response to some task or other. Very often, the detail of the study and, indeed, the means by which the picture was obtained, are eschewed in favor of some summary such as "Scientists have

discovered that when people do X, a part of their brain activates". I am not naive enough to believe that journalists never spice up a story, or always report accurately what they are told by the scientist. Nonetheless, the frequency with which this sort of vacuous statement is made is starting to cause some concern. To whit, if the general public gets the idea that producing pictures of the brain is what cognitive psychologists do and, indeed, that cognitive psychologists who don't produce such pictures are not doing their job, then there is a risk that the tyranny of the graphical over the contentful will be exercised in our own realm as it has been in so many others (e.g., mainstream film production).

Finally, I would like to draw attention to the disadvantages inherent in the creeping medicalization of cognitive psychology. I can see clear dangers looming if, perhaps via on over-reliance on neuroimaging techniques, we allow cognitive psychology to become seen as a branch of the medical sciences. The problem is not so much in the short-term, since money for "medical" purposes might initially be easier to come by. The drift might prove counterproductive, however, in the longer term. If money for basic cognitive psychology is increasingly dressed up as being for medical research, while reliance is decreased on funding sources that see cognitive psychological knowledge as an end in itself, then sooner or later somebody will ask what *medical* benefits have thereby been gained. Since, for many branches of cognitive psychology, these are likely to be slim or even nonexistent, a sudden reduction in available funds might be precipitated. This is very possibly an over-dramatization, but for those who are doubtful that anything like it could happen, I draw attention to the ominous portents of a recent reorganization at the U.S. National Institute for Mental Health, as described in Holden (2004). In brief, the reorganization resulted in basic cognitive and behavioral research being lowered in priority unless it could be shown to have a "strong disease component", an outcome that caused some consternation among leading cognitive researchers. This indicates very clearly what happens when behavioral researchers come to rely on funding from medical sources and when those medical sources subsequently, and possibly quite reasonably, decide to concentrate their resources on the treatment of illness. Somewhat by way of a (telling) aside, such concerns regarding the medicalization of psychology were buttressed recently when the UK national broadcaster (i.e., the BBC) chose as the presenter of a major new series on "The Human Mind" a (very eminent) embryologist, someone whose proper business might, to paraphrase Fodor, be deemed to be some way south of the neck!

Conclusion

In this article, I have sought to do no more than raise questions regarding both the

retrospective and prospective contribution of functional neuroimaging regards the development of cognitive theory. When I first started writing the talk on which this article was based, I thought it inconceivable that some fifteen years of (relatively expensive) work in the area would not have contributed significantly to cognitive theory. Coltheart (2002, this issue) raised some doubts, and writing this paper only served to crystallize them. If responses to this article and Coltheart's are able to demonstrate where and how such a contribution has been made, then perhaps such demonstrations will help increase the future hit-rate. If they are not, then some serious reflection is in order.

References

BOTTVINICK MM, COHEN JD, and CARTER CS. Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences,* 8: 539-546, 2004.

BRODMANN K. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien Dargestellt auf Grund des Zellenbaues. 1909.

BROWN GDA, PREECE T, and HULME C. Oscillator-based memory for serial order. *Psychological Review,* 107: 127-181, 2000.

BURGESS N and HITCH GJ. Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review,* 106: 551-581, 1999.

BURGESS PW and SHALLICE T. Confabulation and the control of recollection. *Memory,* 4: 359-411, 1996.

COLTHEART M. What has functional neuroimaging told us about the mind (so far)? *Submitted to Cortex*.

COLTHEART M. Cognitive neuropsychology. In Wixted J. (Ed.) *Stevens' handbook of experimental psychology, vol 4. Methodology*. New York, 2002, pp. 139-174.

DONALDSON W. The role of decision processes in remembering and knowing. *Memory and Cognition,* 24: 523-533, 1996.

DUNCAN J. An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience,* 2: 820-829, 2001.

FODOR JA. Let your brain alone. *London Review of Books,* 21, 1999.

HARLEY TA. Does cognitive neuropsychology have a future? *Cognitive Neuropsychology,* 21: 3-16, 2004a.

HARLEY TA. Promises, promises. *Cognitive Neuropsychology,* 21: 51-56, 2004b.

HEATHCOTE A. Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29: 1210-1230, 2003.

HENSON RNA. What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology,* 58A: 193-233, 2005.

HENSON RNA, CANSINO S, HERRON JE, ROBB WGK, and RUGG MD. A familiarity signal in human anterior medial temporal cortex. *Hippocampus,* 13: 259-262, 2003.

HOLDEN C. Nimh takes a new tack, upsetting behavioral researchers. *Science,* 306: 602, 2004.

INDEFREY P and LEVELT WJM. The spatial and temporal signatures of word production components. *Cognition*, 92: 101-144, 2004.

JOANISSE MF and SEIDENBERG MS. Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences USA*, 96: 7592-7597, 1999.

LEVELT WJM, ROELOFS A, and MEYER AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22: 1-75, 1999.

LOGOTHETIS NK, PAULS J, AUGATH M, TRINATH T, and OELTERMANN A. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412: 150-157, 2001.

LOGOTHETIS NK. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. In Parker A. (Ed.) *The physiology of cognitive processes*. London, 2003, pp. 62-116.

LONGWORTH CE, MARSLEN-WILSON WD, RANDALL B, and TYLER LK. Getting to the meaning of the regular past tense: Evidence from neuropsychology. *Journal of Cognitive Neuroscience*, 17: 1087-1097, 2005.

MARSLEN-WILSON WD and TYLER LK. Capturing underlying differentiation in the human language system: Comment. *Trends in Cognitive Sciences*, 7: 62-63, 2003.

MCCLELLAND JL and PATTERSON K. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6: 465-472, 2002a.

MCCLELLAND JL and PATTERSON K. 'Words or rules' cannot exploit the regularity in exceptions: Reply to Pinker and Ullman. *Trends in Cognitive Sciences*, 6: 464-465, 2002b.

MIOZZO M. On the processing of regular and irregular forms of verbs and nouns: Evidence from neuropsychology. *Cognition*, 87: 101-127, 2003.

OTTEN LJ and RUGG MD. Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex*, 11: 1150-1160, 2001.

OTTEN LJ, HENSON RNA, and RUGG MD. Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain*, 124: 399-412, 2001.

PAAP KR. Functional neuroimages do not constrain cognitive models of language processing. Paper presented at the 22nd Annual Interdisciplinary Conference, Jackson Hole, Wyoming, February, 1997.

PAGE MPA. What can't functional neuroimaging tell the experimental psychologist? Paper presented at the Conference of the Experimental Psychology Society, London, January, 2004.

PINKER S and ULLMAN MT. The past and future of the past tense. *Trends in Cognitive Sciences,* 6: 456-463, 2002a.

PINKER S and ULLMAN M. Combination and structure, not gradedness, is the issue: Reply to mcclelland and patterson. *Trends in Cognitive Sciences,* 6: 472-474, 2002b.

PLAUT DC, MCCLELLAND JL, SEIDENBERG MS, and PATTERSON K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review,* 103: 56-115, 1996.

PULVERMÜLLER F. Words in the brain's language. *Behavioral and Brain Sciences,* 22: 253-336, 1999.

PYLYSHYN ZW. Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences,* 3: 111-169, 1980.

RAMSCAR M. The past-tense debate: Exocentric form versus the evidence. *Trends in Cognitive Sciences,* 7: 107-108, 2003.

SEIDENBERG MS and JOANISSE MF. Show us the model: Comment. *Trends in Cognitive Sciences,* 7: 106-107, 2003.

SERON X and FIAS W. How images of the brain can constrain cognitive theory: The case of numerical cognition. *Submitted to Cortex*.

SHALLICE T. Functional imaging and neuropsychology findings: How can they be linked? *Neuroimage,* 20: S146-154, 2003.

SMITH EE and JONIDES J. Working memory: A view from neuroimaging. *Cognitive Psychology,* 33: 5-42, 1997.

STROOP JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 18: 643-662, 1935.

STUSS DT, ALEXANDER MP, PALUMBO CL, BUCKLE L, and AL. E. Organizational strategies with unilateral or bilateral frontal lobe injury in word learning tasks. *Neuropsychology,* 8: 355-373, 1994.

SWICK D and KNIGHT RT. Is prefrontal cortex involved in cued recall? A neuropsychological test of PET findings. *Neuropsychologia,* 34: 1019-1028, 1996.

TULVING E. Memory and consciousness. *Canadian Psychology,* 26: 1-12, 1985.

TULVING E, KAPUR S, CRAIK FIM, MOSCOVITCH M, and HOULE S. Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences USA,* 91: 2016-2020, 1994.

TYLER LK, STAMATAKIS EA, JONES RW, BRIGHT P, ACRES K, and WILSON WDM.

Deficits for semantics and the irregular past tense: A causal relationship? *Journal of Cognitive Neuroscience,* 16: 1159-1172, 2004.

UTTAL WR. *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press, 2001.

WINSTON JS, HENSON RNA, GOULDEN MRF, and DOLAN RJ. Fmri-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology,* 92: 1830-1839, 2004.

YONELINAS AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46: 441-517, 2002.

Author Notes

I would like to thank Max Coltheart, Rik Henson, Tim Shallice, Dennis Norris, John Duncan, James McQueen, Holger Mitterer, and Jan-Peter de Ruiter for interesting exchanges.

All correspondence and requests for reprints should be sent to Mike Page at the School of Psychology, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, U.K.