Functional imaging and neuropsychology findings: how can they be linked?

Tim Shallice\textsuperscript{a,b,\*}

\textsuperscript{a} Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK
\textsuperscript{b} Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy

Abstract

It is argued that in poorly understood domains functional imaging and neuropsychology findings on cognitive processes can be related only through functional models of normal cognition. The psychological concept of “resource” can, however, be simply extrapolated to functional imaging. It is then argued that double dissociations can have analogous inferential power for extrapolation to models of normal cognition in functional imaging as in neuropsychology. The argument is illustrated by the example of the control processes involved in functional episodic memory imaging of experiments.

Introduction

The operations of the brain when carrying out cognitive operations can be understood at a variety of levels. When the brain is considered as a physical entity implementing physiological processes the contrasting potentials of different types of imaging methodology are well understood. Thus fMRI and PET provide more adequate information about the localization of brain processes, EEG and MEG provide more adequate information about temporal aspects. However, if one is concerned with how useful different methodologies are to address issues related to cognitive theories the situation is much less clear. Just because fMRI, say, can provide much better evidence on anatomical localization than neuropsychology, this does not necessarily mean that it provides more definitive information for answering cognitive questions.

I will make the assumption that a valuable conceptual domain for understanding cognitive processes is that of the “functional architecture,” in which the concept is broadly construed. (For a critique see Churchland and Sejnowski, 1989.) Traditionally, neuropsychology has been used as a powerful source of evidence for theorizing on the basic functional architecture of the mental processes in a domain (see, e.g., Shallice, 1988). However, recently certain neuropsychologists have argued that functional imaging of brain processes does not have an analogous potential. Thus Coltheart (2003a) argues that the findings of imaging studies cannot be used to falsify functional theories of cognition. He gives the example of multiple route models of reading and says that one would not reject psychological evidence for their existence just because one could not obtain imaging evidence that they are anatomically distinct. Harley (2003) adds that one would not reject a single-route model just because there appeared from functional imaging to be two anatomical routes. He goes further: “you need to have a complete theory of cognition before you can begin interpreting images. Hence imaging can in principle add nothing. There is a level of psychological theorizing—the cognitive level—which can only be studied at that level, and information from lower levels will tell us nothing about what happens at the cognitive level.”

I will provide a different type of answer. I will consider the following questions:

(1) Consider a domain in which only the basic cognitive architecture is still not well understood. Can functional imaging, like neuropsychology, provide critical information on the basic functional organization of a domain?
If these arguments are valid then the answers to questions 1 and 2 above are also basic for question 3, so I will turn to these. If the way neuropsychological and functional imaging investigations are to be related depends upon how each relates to the cognitive architecture we are faced with the problem that many different types of functional architecture have been proposed as the underpinnings of some or many cognitive domains. Dawson (1998) lists 24 types!

The three main types are:

(i) Isolable subsystems, for which strong evidence can be produced by internal human experimental psychology means and by other neurobiological and in particular neuropsychological means (see, e.g., Posner, 1978; Sternberg, 2001; Shallice, 1988).

(ii) Connectionist models—both distributed (see McClelland and Rumelhart, 1986) and localist (e.g., Page, 2000).

(iii) Production systems (e.g., Newell, 1990).

Neuropsychological evidence has been used extensively to provide evidence with respect to both types of connectionist models (for distributed models, e.g., Mozer and Behrmann, 1990; Farah and McClelland, 1991; Plaut and Shallice, 1993a,b for localist ones, e.g., Houghton et al., 1994; Hartley and Houghton, 1996) but hardly at all for production systems (but see Goel et al., 1997; Rumiati et al., 2000). However, there has been very little attempt to relate fMRI or PET evidence to such models (but see Fiez and Petersen, 1998, and Henson et al., 2000a, for two exceptions). The reason is that for “connectionist”—artificial neural net—models, there appears to be no principled way without detailed knowledge of the underlying anatomy to map the spatial distribution of activity in the cortex onto that of the conceptual “hidden units” of a model, across which there is normally no distance measure (but see the model of optic aphasia of Plaut (2002), which has more potential in this respect).

This leaves one plausible conceptual bridge between fMRI results, say, and cognitive models, namely models based on isolable subsystems. Such models were widely used in the early days of cognitive psychology (e.g., Posner, 1978). However, they are widely guyed pejoratively as mere “box-and-arrow” models. Yet they can be implemented rigorously (e.g., Cooper, 2002). As critically, the separability of one system from another system involved in execution of a task can be rigorously supported, for instance by process decomposition procedures which enable separable modifiability (Sternberg, 2001) to be observed. The best known such behavioral procedure is that of the use of additive factors methodology with reaction time measures (Sternberg, 1969). However, additive factors methodology using chronometric measures has in practice had considerable problems (see, e.g., Sanders, 1998, Chap. 3), for instance due to the existence of phenomena explicable by cascading processes. Yet quite independent of chronometric
procedures in normal subjects, the ubiquity of dissociations in neuropsychological evidence allows such phenomena to be used effectively for process decomposition procedures (e.g., Shallice, 1988). Moreover there are natural extensions of separable modifiability logic to the interpretation of functional imaging evidence (e.g., Sternberg, 2001). These include most directly the use of factorial designs and somewhat more indirectly the use of conjunction designs (e.g., Price and Friston, 1997).

Consider cognitive domains in which the functional architecture is poorly understood. Even today this includes nearly all the underpinnings of thought. The nature of the inference from the observations of dissociations in neuropsychology to the organization of normal cognition has been analyzed by Shallice (1979, 1988). For such poorly understood domains, the use of dissociations was essentially viewed as a discovery procedure, which needs converging evidence other than the dissociation per se in order to support the existence of an isolable system specifically impaired in the relevant patients. This is because the existence of dissociations does not logically entail the existence of separable systems (see also Coltheart and Davies, 2003; Gurd and Marshall, 2003 for relevant discussion). However, Bullinaria and Chater (1995) have shown that if one takes a connectionist architecture and restricts lesions to subsets of hidden units in a feedforward three-layer model—the most straightforward implementation of a conceptually nonseparable system—then double dissociations only very rarely occur. Thus the existence of double dissociations strongly suggests that isolable subsystems exist, although the nature of the underlying subsystems may not be simply describable in terms of the variables manipulated in the empirical study (see Plaut and Shallice, 1993b, for an example).

The reason double dissociations—i.e., complementary dissociations in two patients or groups of patients—were introduced into neuropsychological methodology was that there is a simple way that observed single dissociations may not derive from damage to a separable subsystem. Thus if patient X is impaired in carrying out task A but not in carrying out task B, then this does not necessarily mean that satisfactory performance of task A involves a subsystem (I) which is impaired in patient X and that satisfactory performance of task B does not require subsystem I. Task B may simply be in some way easier than Task A in this respect.

A formalization of dissociation logic developed by Shallice (1979, 1988) was designed to show that double dissociations cannot be explained in an analogous fashion. In this analysis, the concept of resource was used to support the inference from neuropsychological evidence to the organization of normal function. This idea was derived from so-called performance-resource analysis of dual-task attention experiments on normal subjects carried out by Norman and Bobrow (1975). It is assumed that in dual-task attention experiments subjects can vary the amount of attention they pay to one task or the other. This can be controlled by varying the relative rewards for good performance on one task or on the other. As the rewards change so it is assumed that the amount of a resource critical for carrying out one of the tasks varies from 0 to 1, say. As the amount of this resource devoted to one of the tasks increases so performance of the task improves monotonically, producing a so-called “performance–resource” curve, which is the signature of this particular task.

In the application to neuropsychological findings the effect of a lesion on a subsystem is held to correspond to a loss of a certain proportion of the resource of that subsystem (see Fig. 1). Now only tasks that require less than the available resource can be carried out. Glymour (1994) has argued that introducing this concept into neuropsychological theorizing makes theories too little constrained by evidence to be properly testable. However, this argument can be shown to be too specific (see Shallice et al., 2000). On this approach a double dissociation pattern of a crossover type cannot be explained through the different tasks making different levels of demands on the same subsystem, since this can be simply shown to lead to an internal contradiction.

Fig. 1. Performance–resource curves for a given system for carrying out qualitatively equivalent easy and difficult tasks. A patient with an impairment in the domain may be thought of as having a suboptimal level of resource (reprinted, by permission of the publisher, from Shallice, 1988, Fig. 10.4, p. 233).

The concept of resource can be extrapolated to functional imaging. The resource required by a subsystem in performance of a given task can be thought of as monotonically related to the “local average neural activity” using Boynton et al.’s (1996) linear systems analysis model of the BOLD signal in fMRI. The local average neural activity should presumably be limited to the input and local processing rather than the output spikes, given Logothetis et al.’s (2001) findings on the relation between BOLD and local field potentials. Possibly even better, “resource” should be normalized as the proportion of the maximum local average neural activity of the region that the current task requires. Such an approach simply indicates why only a single dis-
sociation in activation values, as occurs in subtraction designs, is subject to the analogous danger to a simple disassociation in neuropsychology. The tasks being compared may simply differ quantitatively in the resources required for satisfactory performance in any of the subsystems held to be being involved in performance of both tasks (see also Price and Friston, 1997).

This, though, leaves open the question of how “region” is defined. Assume that if performance of both of two tasks (I and II) requires use of the same subsystem but one requires it at resource level $R_1$ and the other resource level $R_2$, where $R_1 > R_2$, then for all parts of the closed continuous anatomical region in which the subsystem is realized in the brain, in the terms of Boynton et al.:

$$(\text{local average neural activity})_{\text{task I}} \geq (\text{local average neural activity})_{\text{task II}}.$$ 

Given this assumption applies for all tasks and regions, if a double dissociation in activation values—each involving a significant difference in complementary directions—occurs across two regions when two different tasks are carried out, then a set of inferences analogous to those in neuropsychology can be made. This is that,

(i) given that task processing is carried out entirely by isolable subsystems,

(ii) subsystems are realized in the brain as discussed above, and

(iii) no region realizes more than one subsystem per task,

then at least two separable systems are differentially involved in the performance of the two tasks in a complementary fashion.

Of course the initial observation of the double dissociation is only the beginning of the process of characterizing the total system of which the subsystems are composed and of specifying the properties of the subsystem. However, a “boot-strapping” procedure very analogous to that which occurs in cognitive neuropsychology (see Glymour, 2002)—in which new tasks are developed to stress the hypothetical functions—can then take place using functional imaging too.

The argument developed in this paper is that, contrary to the position of Harley (2003) and Coltheart (2003a, 2003b), functional imaging like neuropsychological methods is able to provide support for the theoretical position that in the normal cognitive system more than one separable system exists in a particular cognitive domain. In neuropsychology (see Caramazza, 1986; Shallice, 1988), one needs to make ancillary assumptions that cannot be checked in our current state of knowledge. For instance to make such inferences one needs to assume that a patient’s behavior can be understood as resulting from the normal system operating with some parts subtracted; one needs to assume that the cognitive systems of all members of a culture are qualitatively equivalent. Analogous inferences from functional imaging, it is argued, also require ancillary assumptions but ones quite different from those required by neuropsychology. These involve assumptions about (i) how subsystems are realized in the brain as given above, (ii) that the relation between specific resource requirement of a task and activation is linear, and (iii) in practice (and unrigorously)—given the way functional imaging results are reported—the degree of anatomical proximity of activation maxima that should count as reactivation of the same region in another task.

**Case history: the HERA theoretical claims**

Consider the domain of the prefrontal processes involved in episodic memory. When functional imaging experiments were first carried out in the domain of episodic memory, it was found that manipulation of the encoding processes led to differential activation in left prefrontal cortex, but manipulation of variables affecting retrieval led to differential activation of right prefrontal cortex (e.g., Kapur et al., 1994; Tulving et al., 1994b; Shallice et al., 1994). In these early experiments encoding processes were not directly contrasted with retrieval ones, so that the double dissociation logic discussed above could not be applied between the two domains. It was, however, shown that variables that affect verbal episodic encoding such as depth of processing (Kapur et al., 1994) or dual task manipulation (Shallice et al., 1994) led to changes in left prefrontal activation, while changes in retrieval demands (e.g., relying on episodic rather than semantic memory) led to differences in right prefrontal activation.

The HERA claim (Tulving et al., 1994a) has been much criticized for not being an adequate empirical generalization. However, this is mainly because it did not account for memory for nonverbal materials (e.g., Kelley et al., 1998), which it was not, however, intended to do! When consideration is restricted to verbal episodic memory metaanalyses support it with respect to the relative lateralization of the encoding and retrieval stages of episodic memory task performance (see, e.g., Lepage et al., 2000; Cabeza and Nyberg, 2000).

Swick and Knight (1996), in an article subtitled “A neuropsychological test of PET findings,” however, argued that “If right prefrontal cortex is crucial for episodic memory retrieval, right frontal patients should be impaired in cued recall” (p. 1020). In a study of patients with frontal lesions they showed that the amount retrieved in cued recall tasks by patients with right frontal lesions was in fact normal. Moreover other studies with patients using other tasks involving episodic memory have shown similar results. Thus Incisa and Milner (1993) found that in free recall their right frontal group did not differ significantly from normals in amount retrieved (see also Kopelman and Stanhope, 1998, for supporting evidence, but see Stuss et al., 1994, for a more complex pattern). Swick and Knight sum-

---

*F. Caramazza, S. Shallice*
marized their findings by writing “The present experiments demonstrate that the regions of prefrontal cortex activated in PET studies of controls are not necessary for patients to perform the cued recall task” (p. 1027). Moreover given that retrieval in free recall in normal subjects activates right prefrontal cortex (Fletcher et al., 1998) the contrast could be extended to that task too. Does this mean that we have a mismatch between the functional imaging and the neuropsychological evidence?

It does not do so, for three reasons. First there are three, if not four, different areas in right prefrontal cortex that have been activated in retrieval from episodic memory. Thus both Lepage et al. (2000) and Fletcher and Henson (2001) review many studies showing different patterns of activation in right anterior, right dorsolateral, and right ventrolateral regions. Thus to argue for a failure to show a correspondence between neuropsychological and imaging evidence one would need to consider subregions of the right prefrontal cortex.

More critically, activation in two of the above regions doubly dissociates. Fletcher et al. (1998) compared retrieval of 16 words from four related categories (e.g., all words related to food) with retrieval of 16 matched words each triggered by a highly specific stimulus (e.g., bread → baguette). The level of recall in the two conditions was closely comparable. However, the free recall task activated the right dorsolateral prefrontal cortex significantly more than the cued recall task. Complementarily, in the right ventrolateral prefrontal cortex, the contrast between activation levels in the two tasks was significantly reversed (see Fig. 2). Moreover a later episodic memory study (Henson et al., 2000) showed a contrast between two key variables in whether there was an effect in the right dorsolateral or right anterior prefrontal structures. Thus the three regions seem to involve two and probably three different functional subsystems.

Second, there is the question of resource and the degree of lateralization of the functional subsystem. In many studies the right prefrontal activation is stronger than the left (e.g., Henson et al., 1999a) or it is only the right prefrontal activation which is significant (e.g., Fletcher et al., 1998a see Table 1). However, 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 the left prefrontal cortex (e.g., Henson et al., 2000a; McCormack et al., 2001). Thus the functional resource at the cognitive level appears to be realized in two complementarily lateralized regions. Thus even a complete lesion of the relevant right prefrontal region will leave the smaller but positive level of resource realized in the left prefrontal area. What level of performance one could expect would depend upon the performance-resource curve for that task for that subsystem. Thus when imaging findings suggest a system is differentially bilaterally localized, the probability of obtaining a complementary neuropsychological deficit is considerably reduced.

The most critical reason is, however, the third. What behavior would be observed following a lesion depends upon the underlying function that the relevant subsystems are carrying out. Now Tulving et al. (1994a), using the HERA model, said “Right prefrontal cortical regions are involved in retrieval of episodic information to an extent that left prefrontal areas are not” and “that does not hold for retrieval of semantic information” (p. 2018). Does retrieval depend on only a single system—a single resource in the

![Fig. 2. The relative activation at recall in two right frontal regions, (34, 36, 32) for the dorsolateral region and (30, 12, 0) for the ventrolateral region, for free recall (task 1) and category-cued paired associate recall (task 2) (reproduced, by permission of the publisher, from Fletcher et al., 1998, Fig. 2, p. 1253).](image)

Table 1

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>Right</td>
</tr>
<tr>
<td></td>
<td>Left</td>
</tr>
</tbody>
</table>

| Cabeza and Nyberg | 11 0 0 | 1 7 4 |
| 9/46              | 1 6 2  |
| 45/47             | 3 0 0  |
| 10/11             | 1 6 5  |
| Lepage et al.     | 0 13 7 |
| Opeculum          | 4 14 3 |
| Polar             |         |
| Fletcher and Henson |      |
| DLPFC             | 10 1 1 | 3 12 5 |
| VLPFC             | 18 0 3 | 3 7 3  |
| AFC               | 1 1 0  | 4 4 9  |

Note. The Cabeza and Nyberg regions are Brodmann areas. DLPFC, dorsolateral prefrontal cortex; VLPFC, ventrolateral prefrontal cortex; AFC, anterior frontal cortex. In each study the relevant initial contrast is coded as left, right, or bilateral. The one discrepancy concerns differential lateralization of anterior/polar activation. This is significant in the Cabeza and Nyberg metaanalysis but completely insignificant in that of Fletcher and Henson (for the Lepage et al. study there is a significant right/left difference for this region at retrieval) (from Cabeza and Nyberg, 2000, Tables 7 and 8; Lepage et al., 2000, Table 3; Fletcher and Henson, 2001, Tables 2 and 5).
the positive extrapolation of a neuropsychological finding to functional imaging. Using a paradigm somewhat analogous to that used by Fletcher et al. (1998), Stuss et al (1994) had studied free recall of word lists by patients. Like other authors discussed above they had noted that it was primarily patients with left frontal lesions who were impaired on the amount recalled. However, if output protocols were looked at in more detail, it was seen that right frontal patients repeated words within them at more than twice the rate of controls. They apparently failed to check their output adequately. Moreover the task used by the Stuss et al. study was very similar to the imaging condition of Fletcher et al. (1998), in which right dorsolateral prefrontal activation was found. Moreover this condition clearly stressed the need of the subject to check the output more than did the paired associate cued recall condition with which it was compared. In this control condition subjects had to respond to a category cue (e.g., bread) with previously presented examples (e.g., baguette). In this condition, if a response word were to come to mind at retrieval, then as the categorical retrieval cues were highly restrictive the putative response would have been unlikely to arise either from another stimulus cue either in the same list or by proactive interference from an earlier list. Subjects who did not check would be unlikely to make an error. The free recall situation did not have that advantage. Any putative response might come from an earlier list. Another reason for subjects to need to check their output is so as not to repeat words. Thus the idea that a subsystem being activated would involve control of checking would explain why the right dorsolateral prefrontal cortex should be significantly more activated in the free recall than in the paired associate condition.

This finding was fortuitous. In later studies, however, retrieval conditions were contrasted according to whether they would stress postretrieval checking processes. One method was to contrast a condition highly dependent on source memory—in fact derived from Jacoby’s (1996) “exclusion” procedure—with one which placed few demands on source memory—related to Jacoby’s “inclusion” procedure. In the former at retrieval a word had to be recognized as having been in the list presented 5 min before and as having occurred in a particular spatial or temporal position. In the corresponding condition all that was required was to recognize whether a word had occurred in the list—its precise spatial or temporal position was irrelevant. In the source memory condition in this study activation of both left and right dorsolateral structures occurred as did that of left and right ventrolateral structures. However, the pattern of findings across different conditions differed between the right and the left dorsolateral structures (see Fig. 3). Thus for the left dorsolateral prefrontal cortex the encoding condition produced the strongest activation, but for the right dorsolateral region it was retrieval in the exclusion condi-
tion. There was a strong double dissociation across these pairs of conditions for those two regions. From the resource perspective, two different systems in left and right dorsolateral prefrontal cortex are being differentially activated between the encoding and the exclusion conditions.

One possibility is that Mandler’s (1980) theory of recognition might explain the pattern of findings in the study of Henson et al. (1999a). Mandler differentiated two processes involved in recognition—recollection and familiarity (see Yonelinas et al., 1996, for supporting results on normal subjects). One possibility a priori is that the inclusion task associated with familiarity might explain the pattern of both sets of results does not fit with the right dorsolateral region holding a system involved in recollection per se. Instead it fits better with the right dorsolateral region being specifically involved in control of checking or monitoring.

Two later studies support the same conclusion. One was an event-related study comparing high- with low-confidence trials. Low-confidence trials were those which activated dorsolateral prefrontal structures (Henson et al., 2000b). This occurred for low-confidence trials in which the response was correct both when the stimulus had been presented before and when it had not. Instead, the Hit versus Correct Rejection contrast activated more anterior right hemisphere structures. In a final study checking was stressed by producing a state of affairs in which there were competing potential responses. At retrieval, subjects were in an interference situation in which three different responses had at different times been connected to the stimulus, but only the last of the three was the correct response. This again led to right dorsolateral prefrontal activation compared with an analogous retrieval condition in which there were no competing responses (Henson et al., 2002).

A neuropsychological analogue

The idea that the right lateral prefrontal cortex is involved in checking can be assessed by neuropsychological means. Reverberere et al. (2002) compared four groups of frontal patients—left and right lateral and superior and inferior medial—on a test of acquisition of spatial rule using a version of the so-called Brixton Spatial Rule Attainment test (Burgess and Shallice, 1996) in which subjects have to guess which of 10 circles will be colored blue on the next card. The blue circle moves across cards according to one of a number of simple rules.

The actual amendment was that just before the rule changed an interfering set of cards containing a red circle that moved according to a different rule was presented. After four red-circle cards the blue circles returned and the subject had to continue to obey the rule that had obtained before the occurrence of the red cards. Left lateral patients were much poorer than right laterals at extracting rules, and indeed the right lateral patients were essentially normal in this regard. However, the right lateral patients made three times the number of capture errors as normal controls. As in the memory experiment the right lateral patients were normal on one measure—extracting of rules—but grossly impaired on a second—being captured by a currently active but irrelevant rule. They failed to check adequately.

What does checking involve? If one returns to the Stuss et al. (1994) neuropsychological study, then patients with right frontal lesions produce a significantly larger number of repeats in free recall. What do normal subjects do to prevent such repetitions? A plausible scenario is as follows: Assume that, if in free recall the repeated response comes to mind, then the subject will experience increased familiarity. If increased familiarity is detected, the subject should interrupt the recall process and instead attempt to recognize whether the item just retrieved had indeed been recalled before. It is supposed that retrieval per se does not present a difficulty for the normal subject or even the right frontal patient; the difficulty arises from creating the intention to retrieve. If a match with a previous response is obtained, then the implicit response needs to be suppressed. These series of operations involve both memory and nonmemory processes. The following nonmemory processes are involved: the specification of the matching criteria for detecting whether the implicit familiarity response is above threshold, the matching process itself, the decision to interrupt ongoing free recall processes, and then the suppression of the incorrect implicit response. Any one of these could involve the right dorsolateral prefrontal cortex. In other memory situations a variety of other checking procedures are possible (see, e.g., Burgess and Shallice, 1996a; Koriat, 2000).

From a resource perspective there are two different reasons one may obtain an apparent mismatch between the results of neuropsychological and those of functional imaging experiments in memory retrieval studies. First, the relatively bilateral realization of the single functional resource at the conceptual level can mean that unilateral lesions produce a relatively small drop in resource. The second more critical reason is that even if a functional system has a role in processes related to task execution its impairment may not be realized as a decrease in an otherwise sensitive measure of task execution. The hypothesized checking function in memory experiments is an example of this. Thus neuropsychological and functional imaging findings can be related only through a cognitive model of task performance.
The more basic point of the paper is that despite their great differences in methodology and preprocessing for inferences about the underlying cognitive system of normal subjects, there can be an isomorphism between the nature of inferences from evidence to underlying models from neuropsychology and from functional imaging. They require very different ancillary assumptions but can have a central common core. Moreover an example was given from a very poorly understood cognitive domain for which there are converging inferences from the two methodologies. This strengthens not only the theoretical conclusions in this particular case but also the assumptions they are each based upon.

Acknowledgments

This research was supported by a grant from the Wellcome Trust and a Co-operative Grant from the Medical Research Council.

References


