Brit. J. Phil. Sci. **45** (1994), 815–835

On the Methods of Cognitive Neuropsychology

Clark Glymour[1]

ABSTRACT

Contemporary cognitive neuropsychology attempts to infer unobserved features of normal human cognition, or 'cognitive architecture', from experiments with normals and with brain-damaged subjects in whom certain normal cognitive capacities are altered, diminished, or absent. Fundamental methodological issues about the enterprise of cognitive neuropsychology concern the characterization of methods by which features of normal cognitive architecture can be identified from such data, the assumptions upon which the reliability of such methods are premised, and the limits of such methods—even granting their assumptions—in resolving uncertainties about that architecture. With some idealization, the question of the capacities of various experimental designs in cognitive neuropsychology to uncover cognitive architecture can be reduced to comparatively simple questions about the prior assumptions investigators are willing to make. This paper presents some of simplest of those reductions.

1 Introduction

Neuropsychology has relied on a variety of methods to obtain information about human 'cognitive architecture' from the profiles of capacities and incapacities presented by normal and abnormal subjects. The nineteenth-century neuropsychological tradition associated with Broca, Wernicke,

[1] Research for this paper was made possible by a fellowship from the John Simon Guggenheim Memorial Foundation and by grant number SBE-9212264 from the National Science Foundation. I thank Martha Farah for teaching me what little I know of cognitive neuropsychology, Jeffrey Bub for stimulating me to think about these issues and for commenting on drafts of this paper, and Peter Slezak for additional comments. Alfonso Caramazza and Michael McCloskey provided very helpful comments on a second draft.

Meynert, and Lichtheim attempted to correlate abnormal behaviour with loci of brain damage, and thus to found syndrome classification ultimately on neuroanatomy. At the same time, they aimed to use the data of abnormal cognitive incapacities to found inferences to the functional architecture of the normal human cognitive system. Contemporary work in neuropsychology involves statistical studies of the correlation of behaviour with physical measures of brain activity in both normal and abnormal subjects, statistical studies of the correlations of behavioural abnormalities in groups of subjects, and studies of behavioural abnormalities in particular individuals, sometimes in conjunction with information about the locations of lesions.[2] The goal of identifying the functional structure of normal cognitive architecture remains as it was in the 19th century.

The fundamental methodological issues about the enterprise of cognitive neuropsychology concern the characterization of methods by which features of normal cognitive architecture can be identified from any of the kinds of data just mentioned, the assumptions upon which the reliability of such methods are premised, and the limits of such methods—even granting their assumptions—in resolving uncertainties about that architecture. These questions have recently been the subject of intense debate occasioned by a series of articles by Caramazza and his collaborators: these articles have prompted a number of responses, including at least one book. As the issues have been framed in these exchanges, they concern:

1. whether studies of the statistical distribution of abnormalities in groups of subjects selected by syndrome, by the character of brain lesions, or by other means, are relevant evidence for determining cognitive architecture;
2. whether the proper form of argument in cognitive neuropsychology is 'hypothetico-deductive'—in which a theory is tested by deducing from it consequences whose truth or falsity can be determined more or less directly—or 'bootstrap testing'—in which theories are tested by assuming parts of them and using those parts to deduce (non-circularly) from the data instances of other parts of the theory;
3. whether associations of capacities, or cases of dissociation in which one of two normally concurrent capacities is absent, or double dissociations in which of two normally concurrent capacities, A and B, one abnormal subject possesses capacity A but not B, while

[2] Neuropsychology has generally made comparatively little use of response times, and I will ignore them here. But see the excellent study by Luce [1986] for a discussion of response time problems related to those considered in this paper.

another abnormal subject possesses B but not A, are the 'more important' form of evidence about normal cognitive architecture.

Bub and Bub [1988] object that Caramazza's arguments against group studies assume a 'hypothetico-deductive' picture of theory testing in which a hypothesis is confirmed by a body of data if from the hypothesis (and perhaps auxiliary assumptions) a description of the data can be deduced. They suggest that inference to cognitive architecture from neuropsychological data follows instead a 'bootstrap' pattern much like that described by Glymour [1980].[3] They, and also Shallice [1988], reassert that double dissociation data provide especially important evidence for cognitive architecture. Shallice argues that if a functional module underlying two capacities is a connectionist computational system of which one capacity requires more computational resources than another, then injuries to the module that remove one of these capacities may leave the other intact. The occurrence of subjects having one of these capacities and lacking the other (dissociation) therefore will not permit a decision as to whether or not there is a functional module required for the first capacity but not required for the second. Double dissociations, Shallice claims, do permit this decision.

The main issue in these disputes is this: by what methods, and from what sorts of data, can the truth about various questions of cognitive architecture be found, whatever the truth may be? There is a tradition in computer science and in mathematical psychology that provides a means for resolving such questions. Work in this tradition characterizes mathematically whether or not specific questions can be settled in principle from specific kinds of evidence. Positive results are proved by exhibiting some method and demonstrating that it can reliably reach the truth; negative results are proved by showing that *no possible* method can do so. There are results of these kinds about the impossibility of predicting the behavior of a 'black box' with an unknown Turing machine inside; about the possibility of such predictions when the black box is known to contain a finite automaton rather than a Turing machine (Gold [1965]); about the indistinguishability of parallel and serial procedures for short-term memory phenomena (Luce [1986]); about which classes of mathematically possible languages could and could not be learned by humans (Osherson and Weinstein [1985]); about whether a computationally bounded system can be distinguished from an uncomputable system by any behavioral evidence (Glymour and Kelly, to appear); about the logical limits of the propositions that can be resolved by any learner (Kelly, submitted) and much more. However

[3] Professor Caramazza informs me that he regards inference in cognitive neuropsychology as having a bootstrap structure, and intended as much in his articles.

abstract and remote from practice such results may seem, they address the logical essence of questions about discovery and relevant evidence. From this point of view disputes in cognitive neuropsychology about one or another specific form of argument are well motivated but ill directed: they are focused on the wrong questions.

From what sorts of evidence, and with what sorts of background assumptions, can questions of interest in cognitive psychology be resolved—no matter what the answer to them may be—by some possible method; and from what sorts of evidence and background assumptions can they *not* be resolved by any possible method? With some idealization, the question of the capacities of various experimental designs in cognitive neuropsychology to uncover cognitive architecture can be reduced to comparatively simple questions about the prior assumptions investigators are willing to make. The point of this paper is to present some of the simplest of those reductions.

2 Theories as functional diagrams and graphs

Neuropsychological theories typically assume that the brain instantiates 'functional modules' that have specific roles in producing cognitive behaviour. In the processes that produce cognitive behaviour, some of the output of some modules is sent as input to other modules until eventually the task behaviour is produced. Various hypothetical functional modules have standard names, *e.g.* the 'phonemic buffer' and accounts of what the modules are supposed to do. Such theories or 'models' are often presented by diagrams. For example, as depicted in Figure 1, Ellis and Young [1988] consider a 'functional model' for object recognition. What do the arrows in the diagrams mean, and what does it mean if one or more of them is missing because of injury? In explaining profiles of normal capacities and abnormal incapacities with the aid of such a diagram, the modules and their connections are understood to be embedded in a larger structure that serves as a kind of *deus ex machina* in producing particular inputs or particular outputs. For example, a subject's capacity to name familiar objects in experimental trials is explained by assuming that the objects are supplied as input to this diagram, and that the subject has somehow correctly processed the instruction 'name the object before you', and this processing has adjusted the parameters of the functional modules and their connections so that the subject will indeed attempt to name the object. None of the instructional processing is represented in Figure 1. Further, it is understood that the modules represented in such diagrams are connected to other possible outputs that are not represented, and with different instructional processing the

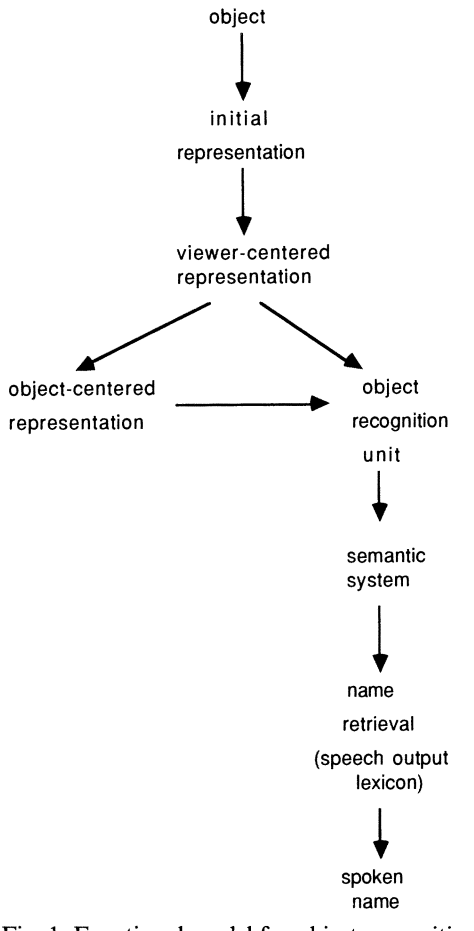


Fig. 1. Functional model for object recognition.

very same stimulus would activate a different collection of paths that would result in a different output. For example, if the subject were instructed 'copy the object before you' and processed this information normally, then the presentation of the object would not bring about an attempt to speak the name of the object but instead to draw it.

In effect, most parts of theories of cognitive architecture are tacit, and the normal behaviour to be expected from a set of instructions and a stimulus can only be inferred from the descriptions given of the internal modules. For example, when Ellis and Young describe an internal module as the 'speech output lexicon' we assume that it must be activated in any process producing coherent speech, but not in processes producing coherent writing or in the processes of understanding speech, writing or gestures. Evidently leaving much of the theory tacit and indicated only by descriptions of internal modules is a great convenience and a practical necessity,

although it may sometimes occasion misunderstanding, equivocation and unprofitable disputes.

The practice of cognitive neuroscience makes a great deal of use of scientists' capacities to exploit descriptions of hypothetical internal modules in order to contrive experiments that test a particular theory. Equally, the skills of practitioners are required to distinguish various kinds or features of stimuli as belonging properly to different inputs, meaning that these features are processed differently under one and the same set of instructions. To address the questions at issue I propose to leave these features of the enterprise to one side, and assume for the moment that everyone agrees as to what stimulus conditions should be treated as inputs to a common input channel in the normal cognitive architecture, and that everyone agrees as to what behaviours should be treated as outputs from a common output channel.

It is also clear that in practice there are often serious ambiguities about the range of performance that constitutes normal, or respectively abnormal, behaviour and that much of the important work in cognitive neuropsychology consists in resolving such ambiguities. I will also put these matters to one side and assume that all such issues are settled, and there is agreement as to which behaviours count as abnormal in a setting, and which normal.

With these rather radical idealizations, what can investigation of the patterns of capacities and incapacities in normal and abnormal subjects tell us about the normal architecture?

3 Formalities

Figure 2 represents another diagram by Ellis and Young. The idea is that a signal, auditory or visual, enters the system, and various things are done to it; the double arrows indicate that the signal is passed back and forth, the single arrows indicate that it is passed in only one direction. The intended reading of the diagram is that if it is intact then spoken and written words will be understood and can produce speech in response that indicates understanding. If, however, any path through the semantic system from the input channel is disrupted while the rest of the system remains intact, then the remaining paths to the phoneme level will enable the subject to repeat a spoken word or pronounce a written word, but not to understand it.

The evidence offered for a diagram consists of profiles of capacities that are found among people with brain injuries. There are people who can repeat spoken words but cannot recognize them; people who can recognize spoken words but can't understand them; people who show parallel

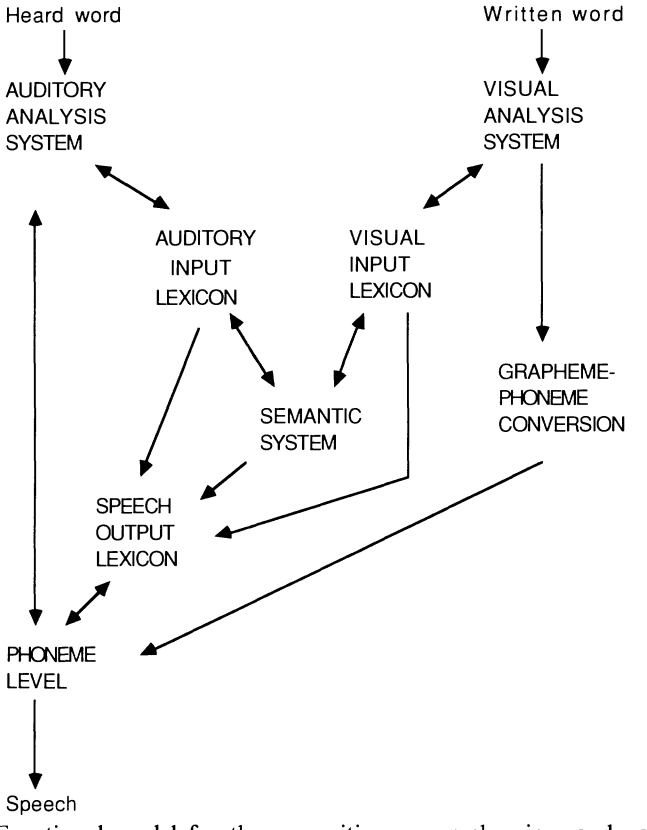


Fig. 2. Functional model for the recognition, comprehension, and naming of written words in reading.

incapacities for written words; people who can repeat, or recognize or understand spoken words but not written, and people with the reverse capacities. What is the logic of inferences from profiles of this kind to graphs or diagrams? To investigate that question I want to consider diagrams that are slightly different from those illustrated.

First, I want the performances whose appearance or failure (under appropriate inputs) is used in evidence to be explicitly represented as vertices in the graphs, and I want the corresponding stimuli or inputs to be likewise distinguished. So where Ellis and Young have an output channel labelled simply 'speech' I want output nodes labelled 'repeats', 'repeats with recognition', 'repeats with understanding'. Anywhere that a psychologist would identify a normal capacity I want a corresponding set of input nodes and output nodes. This convention in no way falsifies the problem for such relations are certainly implicit in the theory that goes with the conventional diagram; I am only making things a bit more

explicit. Second, I want to consider only the identification of pathways that are *essential* for a normal capacity. So if we were considering only the structure associated with the capacity to repeat a spoken word with understanding, the existence of pathways from the heard word to speech that do not pass through the 'semantic system' would be irrelevant. There are certainly examples in the literature of capacities that have alternative pathways, either of which will produce the appropriate outputs. I will ignore this complication. The justification for this second assumption is that I want to explore limitations on any possible strategy for identifying cognitive structure from normal and abnormal profiles of capacities. Restricting ourselves to identifying essential pathways and ignoring the possibility of alternative pathways that are sufficient for a capacity makes the problem of distinguishing one graph from others easier rather than harder. Limitations that hold for easier problems will hold as well for more difficult problems.

The system of hypothetical modules and their connections form a *directed graph*, that is, a set V of vertices or nodes and a set E of ordered pairs of vertices, each ordered pair representing a *directed edge* from the first member of the pair to the second. Some of the vertices represent input that can be given to a subject in an experimental task, and some of the vertices represent measures of behavioural response. (We count instructions to subjects as part of the input.) Everything in between, which is to say most of the directed graph that represents the cognitive architecture, is unobserved. Each vertex between input and behavioural response can represent a very complicated structure which may be localized in the brain or may somehow be distributed; each directed edge represents a pathway by which information is communicated.

Such a directed graph may be a theory of the cognitive architecture of normals; the architecture of abnormals is obtained by supposing that one or more of the vertices or directed edges of the normal graph has been removed. Any individual subject is assumed to instantiate some such graph. In the simplest case, we can think of the output nodes of such a directed graph as taking values 0 and 1, where the value 1 obtains when the subject exhibits the behaviour expected of normal subjects for appropriate inputs and instructions, and the value of 0 obtains for abnormal behavior in those circumstances.

One of the ideas of cognitive neuropsychology is that one and the same module can be involved in the processing of quite different inputs related to quite different outputs. For example, a general 'semantic system' may be involved in using knowledge in speech processing, but it may also be involved in using knowledge in writing or in non-verbal tasks. Some of the input channels that are relevant to a non-verbal task that accesses the

'semantic system' may not be input channels for a verbal task that accesses the 'semantic system'. Although there is in the diagram or graph a directed graph from input channels particular to non-verbal tasks to the output channels of verbal tasks, those inputs are none the less irrelevant to the verbal task. Formally, the idea is that in addition to the directed graph structure there is what I shall call a *relevance* structure that determines for a given output variable that it depends on some of the input variables to which it is connected in the directed graph but not on other input variables to which it is so connected. The relevance structure is simply part of the theory the cognitive scientist provides. One and the same output variable can have several distinct relevant input sets. I will call a *capacity* any pair $\langle I, U \rangle$, where U is an output variable (or vertex) and I is a set of input vertices, such that in normals the set I of inputs is relevant to output U.

Between input and output a vast number of alternative graphs of hypothetical cognitive architecture are possible a priori. The fundamental inductive task of cognitive psychology, including cognitive neuropsychology, is to describe correctly the intervening structure that is common to normal humans.

To begin with I make some simplifying assumptions about the direct graph that represents normal human cognitive architecture. I will later consider how some of them can be altered.

A1. Assume the graph is *acyclic*. That is, the internal process that results in a subject's exhibiting a normal cognitive competence on any particular occasion in response to any particular set of inputs is such that for each functional module X activated in the process, there is no sequence of modules $X_1, X_2, \ldots X_n$, such that some output of X goes to X_1, some output of X_1 goes to X_2, ..., and some output of X_n goes to X.

A2. Assume that the behavioural response variables take only 0 or 1 as values, where the value 1 means, roughly, that the subject exhibits the normal competence, and the value 0 means that the subject does not exhibit normal competence.

A3. Assume that all normal subjects have the same graph, *i.e.* the same cognitive architecture.

A4. Assume that the graph of the cognitive architecture of any abnormal subject is a *subgraph* of the normal graph—*i.e.* is a graph obtained by deleting either edges or vertices (and of course all edges containing any deleted vertex) or both in the normal graph.

A5. The default value of all output nodes—the value they exhibit when they have not been activated by a cognitive process—is zero.

A6. If any path from a relevant input variable to an output variable

that occurs in the normal graph is missing in an abnormal graph, the abnormal subject will output the value 0 for that output variable on inputs for which the normal subject outputs 1 for that variable.

A7. Every subgraph of the normal graph will eventually occur among abnormal subjects.

These assumptions are in some respects unrealistic; input and output are not clear 0, 1 valued functions, for example, and undoubtedly there is feedback among modules. These complications do not affect negative results below, but they make suspect the application to practice of positive formal results. Further, one might object to the assumption that all pathways in a graph between input and output must be intact for the normal capacity. An alternative explored by Bub and Bub [1988] is that just one pathway need be intact. It turns out, however, that this interpretation only makes identification of structure more difficult, but does not change the essential results. I have assumed, in keeping with what seems to be theoretical practice, that the architectural diagrams do not include directed edges representing connections that *inhibit* an effect. If such edges were allowed, injuries could present new capacities not present in normals; from a formal point of view the possibility is interesting and should be investigated.

4 Discovery problems and success

We want to know when, subject to these assumptions, features of normal cognitive architecture can be identified from the profiles of the behavioural capacities and incapacities of normals and abnormals. It is useful to be a little more precise about what we wish to know, so as to avoid some likely confusions.

I will say that a *discovery problem* consists of a collection of alternative conceivable graphs of normal cognitive architecture. So far as we know a priori, any graph in the collection may be the true normal cognitive architecture. We want our methods to be able to identify the true structure, no matter which graph in the collection it is, or we want our methods to be able to answer some question about the true structure, no matter which graph in the collection it is. Whichever graph may actually describe normal architecture, the scientist receives examples—subjects—who instantiate the normal graph and who instantiate various subgraphs of the normal graph. For each subject the scientist obtains a *profile* of that subject's capacities and incapacities. So, abstractly, we can think of the scientist as obtaining a sequence of capacity profiles, where the maximal profiles (those with the most capacities) are all from the true but unknown

normal graph, and other profiles are from subgraphs of that normal graph.

Because of A7 eventually the scientist will see every profile of capacities associated with any subgraph of the normal graph. Let us suppose, as is roughly realistic, that the profiles are obtained in a sequence, with some (perhaps all) profiles being repeated. After each stage in the sequence let the scientist (or a method) conjecture the answer to a question about the architecture. No matter how many distinct profiles have been observed at any stage of inquiry, the scientist cannot be sure that further distinct profiles are not possible. We cannot (save in special cases) be sure at any particular time that circumstance has provided us with every possible combination of injuries, separating all of the capacities that could possibly be separated. Hence, if by success in discovering the normal cognitive architecture we mean that after some finite stage of inquiry the scientist will be able to specify that architecture and know that the specification will not be refuted by any further evidence, success is generally impossible. We should instead require something weaker for success: the scientist should eventually reach the right answer by a method that disposes her to stick with the right answer ever after, even though she may not know when that point has been reached.

I will say that a method of conjecturing the cognitive architecture (or conjecturing an answer to a question about that architecture) *succeeds* on a discovery problem provided that for each possible architecture, and for each possible ordering (into an unbounded sequence) of the profiles of normals and abnormals associated with that architecture, there is a point after which the method always conjectures the true architecture or always answers the question correctly. If no method can succeed on a discovery problem, I will say the problem is *unsolvable*.

5 Some examples

Consider the graphs in Figure 3. The discovery problem posed by this collection of alternative graphs can be solved: whichever graph should describe the true cognitive architecture, one can eventually conjecture the correct graph from a sequence of profiles of normal and abnormal capacities and stick with that conjecture. All of these graphs allow the same normal profile: $N = \{\langle I1, U1\rangle, \langle I1, U2\rangle, \langle I2, U1\rangle, \langle I2, U2\rangle\}$. With each of these graphs there is associated the subgraphs that can be formed by deleting one or more edges or vertices. Each normal graph entails constraints on the profiles that can occur in abnormals. Graph (1), for example, entails the empty set of contraints; every subset of N is allowable as an abnormal profile if (1) represents the normal architecture. Graph (2) imposes strong contraints: If an abnormal has two intact

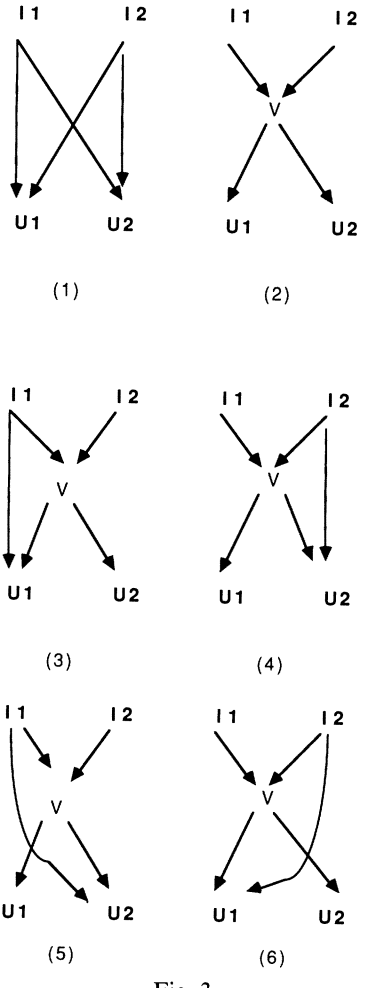


Fig. 3.

capacities that together involve both inputs and both outputs, then he must have all of the normal capacities. Graph (3) permits that an abnormal may be missing ⟨I1, U1⟩ while all other capacities are intact. Graph (4) allows that an abnormal may be missing the capacity ⟨I2, U2⟩ while all other capacities are intact. We have the following inclusion relations among the sets of allowable (normal and abnormal profiles) associated with each graph: The set of profiles allowed by graph (1) includes those allowed by (3) and (4). The set of profiles allowed by (4) is not included in and does not include the set of profiles allowed by (3). The sets of profiles allowed by (3) and (4) both include the set of profiles allowed by (2). And so on.

To make matters as clear as possible, I present a list of the profiles that

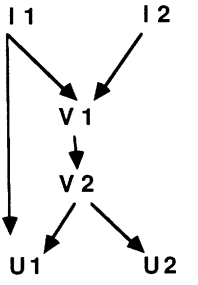
the six graphs permit, where a profile is a subset of the four capacities, and the capacities (Ii, Uj) are identified as ordered pairs i, j. The set of all possible profiles is as follows:

N:	1,1	1,2	2,1	2,2
P1:	1,1	1,2	2,1	
P2:	1,2	1,2		2,2
P3:	1,1		2,1	2,2
P4:		1,2	2,1	2,2
P5:	1,1	1,2		
P6:	1,1		2,1	
P7:		1,2	2,1	
P8:	1,1			2,2
P9:		1,2		2,2
P10:			2,1	2,2
P11:	1,1			
P12:		1,2		
P13:			2,1	
P14:				2,2
P15:				

Graph 1: Abnormals with every profile occur.
Graph 2: Abnormals with P5, P6 and P9–P15 occur.
Graph 3: Abnormals with P4, P5, P6 and P9–P15 occur.
Graph 4: Abnormals with P1, P5, P6 and P9–P15 occur.
Graph 5: Abnormals with P3, P5, P6 and P9–P15 occur.
Graph 6: Abnormals with P2, P5, P6 and P9–P15 occur.

The following procedure solves the discovery problem: *Conjecture any normal graph whose set of normal and abnormal profiles includes all of the profiles seen in the data and having no proper subset of profiles (associated with one of the graphs) that also includes all of the profiles seen in the data.*

One can think of the inference procedure in this case as embodying a kind of simplicity principle. This does not mean that every discovery problem posed by a collection of possible cognitive architectures and assumptions A1 through A7 is solvable. There are at least three ways in which indistinguishable structures can occur: The edges coming into a vertex v can be pinched together at a new vertex v′ and a directed edge from v′ to v introduced: the edges coming out of a vertex v can be moved so that they are out of a new vertex v′ and an edge from v to v′ introduced; and, finally, a vertex v can be replaced by a subgraph G such that every edge in v is replaced by an edge into G, every edge out of v is replaced by an edge out of G, and every input to G has a path in G to every output of G. Each of



(7)
Fig. 4.

these operations results in a graph that is indistinguishable from the original graph in the normal and abnormal profiles it allows. The first two operations are really only special ways of thinking about the third.

For example, graph (7) (Figure 4) is indistinguishable from graph (3). If it is added to the preceding set of six graphs, the corresponding discovery problem cannot be solved. Whenever two capacities have the same output variable, we can 'pinch' any subset of their paths and obtain an indistinguishable graph (Figure 5). Of course, the possibilities are not restricted

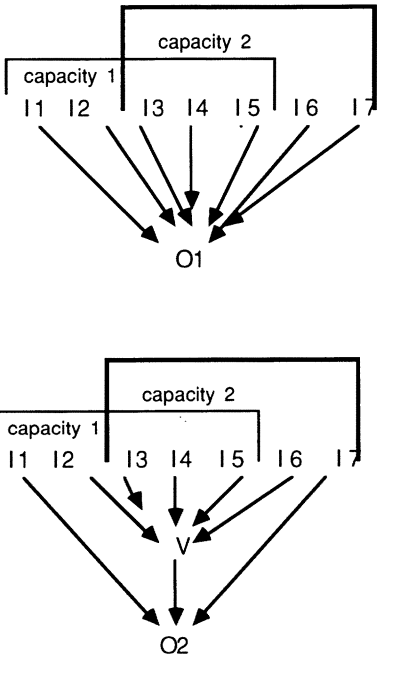


Fig. 5.

to a single pinch. There can be any grouping of lines, and there can be hierarchies of intermediate nodes. The space of possibilities is *very* large. The number of ways of introducing extra vertices that are immediately between the inputs and a single output is an exponential function of size of that set. And, of course, directed edges between intermediate vertices at the same level can be introduced. One possible view about such indeterminacies is of course that they represent substructure that is not to be resolved by cognitive neuropsychology. Bub and Bub [1991] have suggested that if there is for each internal module an input/output pair specific to that module then the entire graph structure can be identified, and that seems correct if extraordinarily optimistic.

The conclusion seems to be that under the assumptions A1 through A7 a good many features of cognitive architecture can be distinguished from studies of individuals and the profiles of their capacities, although a graph cannot be distinguished from an alternative that has functionally redundant structure. Under assumptions A1 through A7, several of Caramazza's claims are essentially correct: he is correct that the essential question is not whether the data are associations, dissociations, or double dissociations; the essential question is what profiles occur in the data. He is correct that from data on individuals one can solve some discovery problems. In any particular issue framed by assumptions of this kind, an explicit characterization of the alternatives held to be possible a priori, and clear formulation in graph theoretic terms of the question at issue would permit a definite decision as to whether the question can be answered in the limit, and by what procedures.

Unfortunately, when the framework is modified to include other, plausible theoretical assumptions that seem to have a hold in cognitive neuropsychology, the prospects are less bright. The assumptions made by Shallice [1988], in particular, while substantively plausible, reduce the possibility of using abnormal data to identify properties of normal cognitive architecture.

6 Resource/PDP models

A picture of the brain that has some currency supposes that regions of the brain function as parallel distributed processors, and receive inputs and pass outputs to modules in other regions. Thus the vertices of the graphs of cognitive architecture that we have thus far considered would be interpreted as something like parallel distributed processing networks (McClelland *et al.* [1986]). These 'semi-PDP' models suggest a different connection between brain damage and behavioural incapacities than is given in A1 through A7. A familiar fact about PDP networks is that a

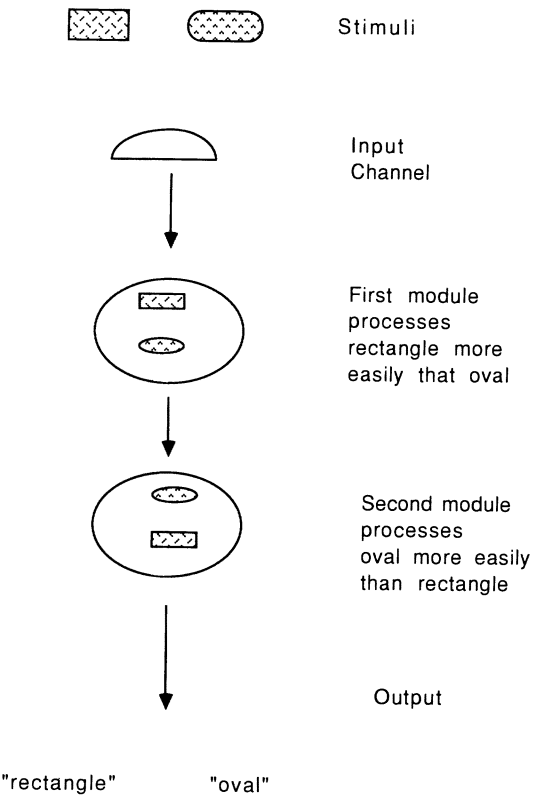


Fig. 6.

network trained to identify a collection of concepts may suffer differential degradation when some of its 'neurons' are removed. With such damage, the network may continue to be able to make some inferences correctly but be unable to perform others. Thus a 'semi-PDP' picture of mental functioning argues that damage to a vertex in a graph of cognitive architecture is damage to some of the neurons of a network and may result in the elimination of some capacities that involve that vertex, but not others. Shallice, for example, has endorsed such a picture, and uses it to argue for the special importance of double dissociation phenomena in cognitive neuropsychology. He suggests that some capacities may be more difficult or computationally demanding than others, and hence more easily disrupted. Double dissociations, he argues, show that of two capacities, at least one of them uses some module not involved in the other capacity.

On reflection, it seems clear that Shallice's point could be made about connections between the PDP modules; some capacities may place greater demands on an information channel than do other capacities that use that

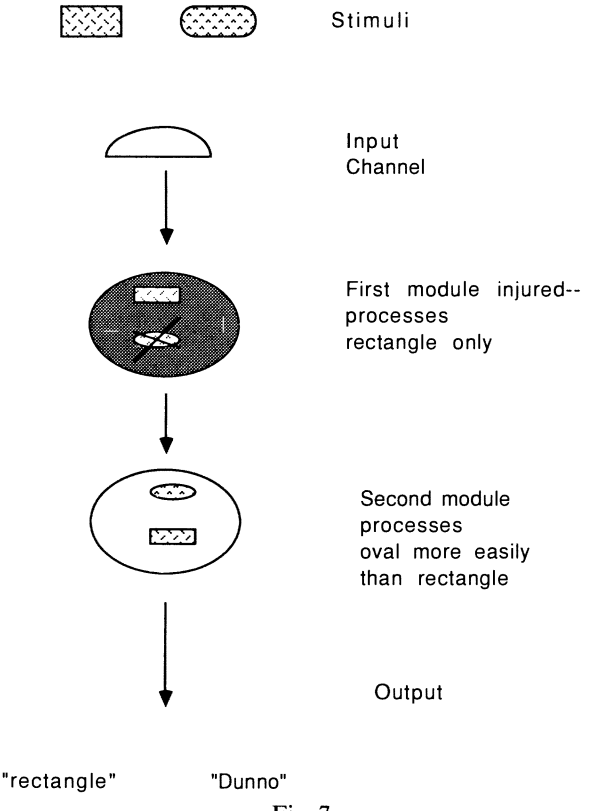


Fig. 7.

same channel. Further, of two capacities that use in common two PDP modules (or channels), one capacity may be the more demanding of one of the modules, and the other the more demanding of the other module. If, in fact, two capacities use exactly the same channels and internal modules, and involve at least two distinct internal modules, then double dissociation may occur provided one capacity uses more of the resources of one of the internal modules while the other capacity uses more of the resources of another, distinct, internal module. Consider the contrived example in Figure 6. Suppose the first module is injured, but only enough to prevent processing the oval (Figure 7). Suppose, now, that the second module is damaged, but only enough to prevent processing the rectangle (Figure 8).

With semi-PDP models, double dissociations thus support the inference that there exists a module m(A) involved in capacity A and there exists a distinct module m(B) involved in capacity B, but double dissociations *do not* support any inference to the conclusion that module m(A) is unnecessary for capacity B or that module m(B) is unnecessary for capacity A.

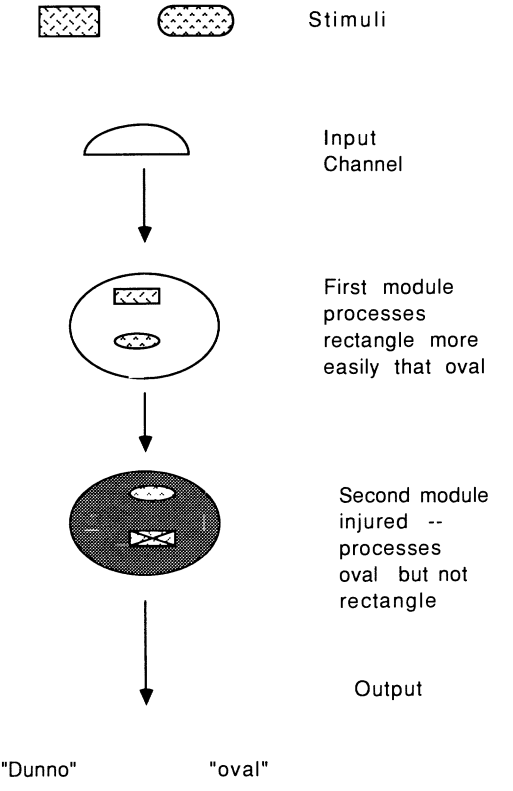
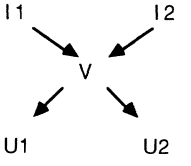


Fig. 8.

Consider next whether under the same hypothesis information about profiles of capacities and incapacities permits us to discover anything at all about cognitive architecture. Shallice's assumptions amount to replacing A6 with a more complicated condition, and altering slightly the character of discovery problems.

With each vertex or edge of the normal graph we should imagine a *partial ordering* of the capacities that involve that edge or vertex. That capacity 1 is *less than or equal to* capacity 2 in the partial ordering indicates that any damage to that edge or vertex that removes capacity 1 also removes capacity 2. If capacity 1 is less than or equal to capacity 2 and capacity 2 is less than or equal to capacity 1, then any injury to the module that removes one capacity will remove the other. If capacity 1 is less than or equal to capacity 2 for some edge or vertex, but capacity 2 is not less than or equal to capacity 1 for that edge or vertex, then capacity 1 is *less than* capacity 2 for that edge or vertex, meaning that capacity 2 can be removed by damage to that element without removing capacity 1. If capacity 1 is not less than or equal to capacity 2 for some edge or vertex, and capacity 2 is



If (I1,U1) is more demanding than all other capacities, profile P4 is added.

If (I1,U2) is more demanding than all other capacities, profile P3 is added.

If (I2, U1) is more demanding than all other capacities, profile P2 is added.

If (I2, U2) is more demanding than all other capacities, profile P1 is added.

Fig. 9.

also not less than or equal to capacity 1 for that edge or vertex, then they are *unordered* for that graph element, meaning that some injury to that graph element can remove capacity 1 without removing capacity 2, and some injury to that graph element can remove capacity 2 without removing capacity 1. A degenerate case of a partial ordering leaves all capacities unordered. I will call a graph in which there is attached to each vertex and directed edge a partial ordering (including possibly the degenerate ordering) of the capacities involving that graph element a *partially ordered graph.*

The set of objects in a discovery problem are now not simply directed graphs representing alternative possible normal cognitive architectures. The objects are instead partially ordered graphs, where one and the same graph may appear in the problem with many different orderings of capacities attached to its edges and vertices. The presence of such alternatives indicates an absence of background knowledge as to which capacities are more computationally demanding than others. I will assume that the goal of inference remains, however, to identify the true graph structure.

Rather than forming abnormal structures by simply deleting edges or vertices, an injury is implicitly represented by *labelling* a directed edge or vertex with the set of capacities involving that edge or vertex that are assumed to be damaged. The profile of capacities associated with such a damaged, labelled graph excludes the labelled capacities. Depending on whether or not there is a partial ordering of capacities or outputs attached to graph elements, there are restrictions on the possible labellings. When partial orderings are assumed a discovery problem is posed by a collection of labelled graphs.

On these assumptions alone the enterprise of identifying modular structure from patterns of deficits is hopeless, as a little reflection should make evident. Even the simplest graph structures become indistinguish-

able. An easy illustration is given by six graphs in the discovery problem of the previous section. Consider what happens when the discovery problem is expanded by adding to graph 2 some possible orderings of the computational demands placed on the internal module v by the four capacities considered in Figure 9.

Thus in addition to the profiles allowed by graph (2) previously, any one of the four profiles characteristic of graphs 3–6 may appear, depending on which capacity places the greatest computational demands on the internal module. If all capacities are equally fragile, the set of profiles originally associated with graph 2 is obtained; still other profiles can be obtained if orderings of the internal module of graph 2 are combined with orderings of the directed edges in that graph. Similar things are true of graphs 3–6. Thus unless one has strong prior knowledge as to which capacities are the most computationally demanding (for every module), even simple discovery problems appear hopeless.

7 Conclusion

The conclusion I draw is not that cognitive neuropsychology is in vain; quite the contrary. My conclusion is that even the smallest formal analysis makes clear some weak points in the project, and emphasizes where argument and inquiry ought to be focused. I regard computational neuropsychological models as interesting and even plausible in many respects, but it should be apparent that any attempts to identify modular functional structure on the assumptions such theories incorporate will depend almost entirely on making good cases about the comparative processing demands of different capacities.

Department of Philosophy
Carnegie Mellon University
and
Department of History and Philosophy of Science
University of Pittsburgh

References

Bub, J. and Bub, D. [1988]: 'On the Methodology of Single-case Studies in Cognitive Neuropsychology', *Cognitive Neuropsychology*, **5**, pp. 565–82.

Bub, J. and Bub, D. [1991]: 'On Testing Models of Cognition Through the Analysis of Brain-Damaged Performance', preprint.

Caramazza, A. [1984]: 'The Logic of Neuropsychological Research and the Problem of Patient Classification in Aphasia', *Brain and Language*, **21**, pp. 9–20.

Caramazza, A. [1986]: 'On Drawing Inferences About the Structure of Normal

Cognitive Systems from the Analysis of Patterns of Impaired Performance: The Case for Single-Patient Studies', *Brain and Cognition*, **5**, pp. 41–66.

Ellis, A. and Young, A. [1988]: *Human Cognitive Neuropsychology*. New Jersey: Lawrence Erlbaum.

Glymour, C. [1980]: *Theory and Evidence*. Princeton: Princeton University Press.

Glymour, C. and Kelly, K., 'Why You'll Never Know if Roger Penrose is a Computer', *Behavioral and Brain Sciences*, to appear.

Gold, E. Mark [1965]: 'Limiting Recursion', *Journal of Symbolic Logic*, **30**, pp. 28–48.

Kelly, K. 'General Characterizations of Inductive Inference Over Arbitrary Sets of Data Presentations', *Journal of Symbolic Logic*, submitted.

Luce, R. D. [1986]: *Response Times*. Oxford: Oxford University Press.

McClelland, J., Rumelhart, D. *et al.* [1986]: *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

McCloskey, M. and Caramazza, A. [1988]: 'Theory and Methodology in Cognitive Neuropsychology: A Response to Our Critics', *Cognitive Neuropsychology*, **5**, pp. 583–623.

Osherson, D., Stob, M. and Weinstein, S. [1985]: *Systems That Learn*. Cambridge, MA: MIT Press.

Shallice, T. [1988]: *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.