difference lies. Essentially, the emphasis in single case studies is on dissociations; however, what is not properly recognized is that group studies are typically concerned with associations and relationships. Shallice is at pains to argue that the value of single cases lies in the information they give concerning dissociation of deficits; he argues consistently against using associations in developing theory (pp. 35; 61; 207). The essential argument given is that an association is not of value because it may later be contradicted by a dissociation, whereas the converse does not hold for a dissociation. The problem, however, is that this is a very general argument. The same logical point was made by Hume (1789) concerning any causal analysis:

Even when experience has informed us of the constant conjunction of two objects it is impossible for us to satisfy ourselves by our reason why we should extend that experience beyond those particular instances which have fallen under our observation. (p. 648)

Shallice's argument thus applies to any observed relationship or association. It is quite true that single cases are a poor and unreliable source of information concerning relationships. This is a well-recognized problem with single cases and one of the main reasons for studying groups of patients. If one looks at the questions posed by studies of groups of patients they typically concern relationships not dissociations. Common questions are the relationship of cognitive impairment to other aspects of psychological impairment and to underlying brain damage. The methodology adopted in conducting and analyzing group studies is typically designed to answer such questions.

Single cases have always provided an important source of neuropsychological evidence and will continue to do so. The strength of the single-case approach is in revealing functional dissociations, but it is a poor guide to associations. On the other hand, group studies are of particular value in investigating relationships and associations. Seen in this light, Shallice's view that neuropsychological evidence should be restricted to dissociations is unsatisfactory. It is unconvincing that there will be long-term progress in a science which does not seek information concerning relationships. The future for cognitive neuropsychology must surely lie with study of both groups and individuals.

1. The focus of the book

Many felt that the empirical base was too narrow (Andrews, Butter & Laeng, Cowan, Grodzinsky, and Smith). Where my discussion of theories of particular functions (e.g., memory [Smith] and executive functions [Butter & Laeng]) ignored evidence from other neurobiological disciplines I consider the criticism justified. With respect to Butter & Laeng's stress on the general importance of lesion localisation in neuropsychology, however, I would agree with Frith that the traditional neuropsychological approach seems much less helpful with functional syndromes analysed by single-case study methods because of the difficulty of localising the functional lesion adequately (see section 9.4; for a concrete example - the short-term memory syndrome - see the discussion in Shallice & Vallar 1990).

Andrews in particular sees the book as being primarily about the general theory of mental organisation, hence far too limited in what information it considers. The book I should have written he argues, would have paid more attention to developmental and cross-cultural data on localisation and recovery of function and to evidence from populations with structural limitations (e.g., the deaf) and special characteristics (e.g., idiot savants) using PET, SPECT, evoked responses, and so on.

Had I been attempting to write a work in the spirit of Fodor's (1983) on modularity, I would agree with Andrews. It is implicit in his view, however, that findings derived by the methods he advocates are theoretically transparent and convergent, applicable without detailed specialized knowledge. As Cowan points out, however, "the full set of tools . . . falls beyond any single investigator's expertise." These diverse data appear on the surface to point in many different directions (see Neuropsychology p. 5). Hence there is a very real danger of "confirmatory bias" in interpreting them from a preexisting theoretical framework derived from another speciality. Smith in fact accuses me of such a bias in my limited use of localisation evidence with respect to executive and memory functions. I see no reason why studies using a technique like PET scanning, which involves very complex normalisation and statistical procedures, should turn out to be any more transparent. For example, in the very interesting PET scanning study of Peterson et al. (1988), certain semantic processing is held to be frontally localised, whereas neuropsychological evidence points to a fairly posterior temporal lobe localisation (e.g., Cappa et al. 1981). How can one assess such a clash of evidence without a detailed knowledge of both methodologies?

Neuropsychology is accordingly concerned with general theoretical questions, but on the (one hopes secure) base of one primary methodology (cognitive neuropsychology), which is examined in depth and compared with a second one (cognitive psychology) that is presumed to be generally better understood. The book is therefore more in the spirit of Posner's (1978) work on mental chronometry than Fodor's (1983) on modularity.

A consequence of this more restricted view is that Neuropsychology does not provide a very good starting point for answering Jerison's provocative questions about evolution. My opinions (e.g., about language) are much the same as his, but most of the methods of neuropsychology discussed are applicable only to humans. Indeed, for

How neuropsychology helps us understand normal cognitive function

Tim Shallice

Department of Psychology, University College London, London WC1E 6BT, England

Electronic mail: ucljitsl@ucl.ac.uk

The commentators on From neuropsychology to mental structure (henceforth Neuropsychology) have in general been gracious in their criticism. The negative parts of their critiques, however, when combined, make a formidable assault on the set of positions I advocated. There were criticisms of the detailed neuropsychological methodology, the general theoretical perspective, and specific views on short-term memory, amnesia, attention and its disorders, and executive functions and consciousness. All these are discussed in turn, but first I would like to make two general points: The first concerns what the book is about.

Author's Response
Chapters 3–7 the underlying functional architecture is assumed to be specific to humans. The topics treated in Chapters 13–16 are, of course, also approachable by general neuroscientific procedures, and in these I have made the standard assumption of a qualitative similarity in functional architecture between humans and the higher mammals. As Jerison points out with respect to blindness and split-brain studies, my treatment in these chapters tends to omit relevant animal studies that were historically important in the development of neuropsychology.

2. Criticisms of accounts of particular areas

2.1. Criticisms of specific theories applied to the neuropsychological evidence. Neuropsychology has two main empirical sections. The substantive arguments of four of the more speculative later chapters (11–16) were criticised in detail. By contrast, only one of the six earlier empirical chapters (3–8) – on short-term memory – was substantively criticised; the chief criticism of these chapters was that the theorising was insufficiently detailed. Thus in nearly all the fields in which I argued that cognitive neuropsychology methodology should be judged, the specific arguments presented were not contested. The disputes over methodology, to which I now turn, were not reflected in a rejection of the specific theoretical conclusions that resulted when the methodology was applied. At worst, the theory was held to be “impoverished” (McCloskey & Caramazza) or “under-detailed” (Grodzinsky) on the basis of a few studies completed since the book was written!

3. Neuropsychological inference

3.1. Single-case and group studies. I considered the section on methodology as perhaps a little dull, but necessary to justify both the many types of empirical study to which I referred and the particular approach I took to single-case studies, which is at odds with much current practice. Yet the reaction to this section was far more negative than to the specific claims made in particular areas. This was also the only section where it seemed that the positions attributed to me by certain critics were the mirror-images of their own positions rather than the ones for which I was attempting to argue.

One set of criticisms, however, was qualitatively different from the others, namely, that of Grodzinsky. For him there was no point in these sections at all. “Neuropsychology . . . is just another method for probing mental operations” and it “does not require unique research strategies; hence all the “methodological” considerations . . . are completely beside the point.” This might mean that anything goes, because the methods of linguistics, for instance differ from those of, say, mental chronometry. Such remarks as “the alleged logical flaws in the argumentation from data acquired from groups of clinically categorized patients are solved . . .” however, suggest that what this commentator means is that the group study methods typical of human experimental psychology should be applied to groups of clinically categorised patients. This seems to imply that I reject group studies, but in fact I defend them (see Chapter 9) if with some reservations (see in particular p. 212). In addition, it seems bizarre to hold that the selection of “clinically defined” groups requires no methods or expertise different from what is needed to select student subjects in normal human experimental psychology. I confess to some difficulty in understanding Grodzinsky’s objections.

Both Hunt and Wilson make a more subtle criticism. They argue that, despite my theoretical position supporting group studies, I in practice emphasise the importance of single-case studies and that this confidence in single-case methodology may be unjustified. In particular, both commentators make the point that single-case methodology routinely assumes qualitative equivalence in functional architecture across patients and that, for some tasks at least, there is evidence that individuals differ qualitatively in the underlying processes used to carry them out. In my view it remains an open question whether further work on individual differences in normal subjects will merely converge theoretically with neuropsychological single-case methodology, as seems to be occurring for reading by sight and by sound, or whether it will undermine it. In any case, I do not see the basis for Wilson’s claim that the evidence of qualitative individual differences makes group studies more useful, although I strongly agree that the multiple single-case is an important design (see Shallice & Vallar 1990).

In complete contrast to the criticisms from Wilson and Hunt (and perhaps Grodzinsky) are those of Caramazza & McCloskey, who argue very forcefully that inferences cannot be made from neuropsychological group studies to normal function. Referring to an argument they have put forward on a number of occasions (e.g., Caramazza & McCloskey 1988; McCloskey & Caramazza 1985) they say “Shallice has obviously misunderstood the argument,” because I referred to an example from normal psychology where it is inappropriate to use group mean data, and they hold inferences from normal and neuropsychological findings to be logically different. My example, however, referred to the case in which what is being tested is the validity of a mathematical function that is not closed under addition, a case that has been discussed in mathematical psychology (e.g., Audley & Jonckheere 1956). This is, of course, quite different from most uses of group data in human experimental psychology. I was attempting to illustrate how, even with normal data, it is the type of inference that is critical. Inferences from normal and neuropsychological data are analogous, in this respect, even though I do agree with the elementary neuropsychological point that “the specific nature of the modification [produced by brain damage] is unknown to us and must be inferred from their [the patient’s] performance” (see p. 219).

My argument was that certain sorts of neuropsychological group data – in particular those indicating certain types of dissociation – allow inferences to normal function to be made just as if they came from individual patients because the pattern exhibited by some member of each group will exhibit generally the same overall pattern as the group mean. An example is the findings of Graf et al. (1984) in which a group of amnesic patients is quantitatively equivalent to the control group on completion but much inferior on recognition. These findings are at
least as useful as a comparable pattern obtained on a single amnesic, and this is so even if there is insufficient data on any individual amnesic in the group to establish reliably the contrast on that patient alone. (See pp. 207–08 for further discussion of the conditions under which this type of statement holds.) Moreover, if a set of group study experiments involves patients selected by the same criteria, then the argument can generalise over studies, as one can assume that if the same set of patients had undertaken all the experiments their mean performance pattern would conform to that of the means of the groups in the separate experiments. What Caramazza & McCloskey do not seem to realise is that my argument does not rely on “the requirement of homogeneity of functional lesion of the group’s members,” which they assume to be necessary for drawing inferences from group data. In many real group studies there are “silent cases” in one or another group, but this does not affect the inferences drawn.

3.2. The role of dissociations. An important criticism made explicitly by Caplan but also implicit in the arguments of Caramazza & McCloskey is that the approach adopted in the book reifies dissociations as a necessary reflection of an underlying damaged subcomponent. This reading is also implicit in Jerison’s provocative suggestion that the flow-diagram models used in cognitive neuropsychology are not really theories at all but complex data descriptions of patterns of dissociations rather analogous to cladograms. This interpretation may arise not only from the general emphasis given to dissociations but also from statements about the use of dissociations for theory-discovery (e.g., p. 220). My remarks on this last topic were much too crude; I did not emphasise it enough that whereas dissociations are valuable in generating hypotheses (say, about alternative reading and writing routes or explicit and implicit memory-processes) many alternative theoretical explanations need to be considered (although Chapter 11 is devoted to this issue). I was actually trying to show “how to select patients and make observations on patients to produce the best chance of developing valid theories” (p. 218) (present emphasis). This was intended to be “a methodological heuristic for obtaining theoretically useful information and not an automatic induction procedure” (p. 36). I therefore agree with Caplan that a double dissociation “does not license the inference that these cognitive components [that underlie performance on the two types of tasks] . . . are specialised mechanisms whose primary functions are to accomplish these tasks.” I would accordingly also reject the analogy with cladograms.

3.3. On associations. My position on inferences from associations as contrasted with dissociations was also criticised by Caramazza & McCloskey (as well as by Wilson, but his premises are in conflict with those of Caramazza & McCloskey and I find his discussion of group studies in this respect insufficiently concrete). Unfortunately, what substantive differences I may have with Caramazza & McCloskey are not illuminated by their argument. They say that if a patient shows a qualitative and quantitative similarity between the deficits manifested on two tasks, the association of deficits can be used to constrain theoretical inferences. I entirely agree. The quotation from Neuropsychology that these commentators take to imply the contrary (p. 35) is qualified in the succeeding paragraph where I point out that “fractionation of a syndrome does not necessarily imply that the more selective of the two disorders is the more pure” (see also pp. 222–24). Moreover, I state explicitly that “the impressive quantitative parallel also means that this is one of the occasions on which it is possible to use an association between deficits as theoretically relevant evidence” (p. 149; see also pp. 155–56). Such examples of productive quantitatively parallel associations were very rare when the book was written. What was far more common, and what my argument about associations in Chapter 10 was directed against, was drawing theoretical conclusions from the existence of impaired performance on two or more tasks, with no quantitative or qualitative parallels between the behavior across the tasks being shown. This situation is not considered by Caramazza & McCloskey.

3.4. On secondary aspects. Yet another example where the methodological differences between the Johns Hopkins group and myself are less clear than McCloskey & Caramazza claim concerns the types of analyses used in neuropsychological research. They argue that “models of reading and spelling . . . have been motivated at least as much by analyses of error types and effects of various stimulus factors (e.g., word frequency, word length) as by patterns of dissociations in performance levels across tasks” – implying that I do not favour such refined analyses. If I did not, it would be a strange rejection of my previous work, as I was involved in the introduction of analyses of word concreteness, word length, regularity, levels of regularity, and typicality of correspondence in the acquired dyslexias (e.g., Shallice & Warrington 1975; 1980; Shallice et al. 1983). In any case, there are extensive discussions of such variables in the three chapters on reading and spelling disorders as well as in parts of Chapter 10. What McCloskey & Caramazza do not seem to appreciate is that according to my approach two tasks contributing to the overall pattern of performance can use stimuli selected to differ on such variables (see, e.g., Chapter 10, pp. 228; 235; 239–40).

Although I stress the importance of the overall pattern of performance for inference to normal function I do not reject more detailed analyses. The latter would provide a more tentative basis for inference, for example, because they require a unique case approach (see pp. 35–37). Even where errors are concerned my frequent references to this type of evidence and my statement that they have provided an important source of evidence (p. 224) are ignored by McCloskey & Caramazza. Where I agree with their criticism is that I now think examining the nature and amount of different error types may be more valuable heuristically than when I wrote Neuropsychology. The recent work of these commentators has been important in changing my mind in this respect, as has some connectionist work I have been doing with Hinton (Hinton & Shallice 1991).

The striking methodological principle that was new to me amongst the commentators was the one put forward by Umilta based on his work with Moscovitch on dementia (Moscovitch & Umilta, in press). This was the idea of treating sparing of a function in the presence of gener-
alised intellectual loss as an index of informational encapsulation. Although there are certain problems with this in the case of strong rather than classical selective preservations, as I will discuss in the next section, the principle seems an important and productive one.

4. The types of theories of normal function to which neuropsychological findings relate

4.1. On “gross, underdetailed” theorising. In Neuropsychology I argued that the types of theories to which neuropsychological findings are most easily related are information-processing ones in which the modes of operation of the individual components are not specified in detail. Two types of objection have been raised to this position. First, it has been argued that this level of theorising is “dangerous” (Bridgeman), “gross [and] underdetailed” (Grodzinsky) and “impoveryished” (McCloskey & Caramazza; see also Jerison). Second, it has been argued that the approach presupposes some form of modularity and that, for at least some mental processes, distributed memory theories may be preferable (Allen, Goshen-Gottstein, Hunt).

First, consider Bridgeman’s argument that psychological models should be realised in a fashion analogous to Mittelstaedt’s (1990) model of head-centric visual localisation, namely, with each component specified by a transfer function. We are far from having such models available for most higher cognitive functions, however. Bridgeman would argue that then we should at least use models that have been implemented, such as Sussman’s (1975) on cognitive skill acquisition. Unfortunately, such a program operates on particular mini-problems far removed from those involved in, say, typical frontal lobe tasks. Moreover, it would be a fairly pointless enterprise to give patients the tasks faced by Sussman’s program, as the patients already have many routine procedures in their cognitive repertory that are unavailable to the program. Even an attempt to simulate the deficits of frontal lobe patients using the most suitable available problem-solving program, SOAR (Laird et al. 1987) would face mammoth problems because SOAR lacks an adequate model of routine performance.

The alternative I adopted was to use underspecified theories as a first step to specifying them further. Bridgeman does not counter the arguments presented in Chapters 3–8 that standard information-processing theories can be applied to neuropsychological evidence more safely now than were the diagrams of nineteenth-century theorists could be then. Moreover, if one takes the Supervisory System model, which is a prototypical of the type Bridgeman criticised, Dehaene and Changeux (in press) have recently put forward a detailed connectionist model of supervisory operations in one frontal lobe task – Wisconsin Card-Sorting. Their model makes a number of specific assumptions that are supported on grounds of neurobiological plausibility but, in addition to the supervisory/routine operation division, it also uses an episodic record to guide the selection of appropriate rules in a given situation in a fashion analogous to the much looser specification of the Supervisory System model.

Another view that gave rise to strong objections was that neuropsychological findings will probably prove less useful for determining the detailed functioning of the mental components than for understanding the overall nature of the functional architecture (Grodzinsky, McCloskey & Caramazza). Such a position cannot be refuted, however, as McCloskey & Caramazza appear to believe, by single examples of the effective use of neuropsychological evidence to support fine-grained theorising. Thus the quotations from Neuropsychology that McCloskey & Caramazza select for criticism refer to “most favourable level” and “most important for” (my emphasis) and not the only level. Indeed, as I discuss some of the earliest fine-grained analyses I know of – those on the operation of the phonological route in reading (see Chapter 5.1) – it would have been ridiculous to claim that neuropsychology could never illuminate finer grained theoretical issues. To demonstrate the incorrectness of the position advocated one would need as wide a range of effective fine-grained analyses as Neuropsychology provides of coarse-grained analyses. We are far from having such analyses available.

My view would be wrong if for most subcomponents of the cognitive system the following two conditions apply. First, for many forms of damage to the subcomponent a particular type of fine-grained behaviour is exhibited. Second, this behaviour does not arise when plausible alternative mechanisms for carrying out the function are impaired. McCloskey & Caramazza provide no argument that this is the case. Indeed, the published one of the two papers referred to by these commentators (Caramazza & Miceli 1990) presents no mechanistic theory of the operation of the graphemic buffer. Any theory based on their impressive evidence would surely be more tentative than the (far from certain) inference that the graphemic buffer itself exists. My argument was simply that a modular system presents neuropsychology with a privileged level for its evidence – the coarse-grained one – and that inferences to other levels of operations are impossible. As for criticisms that such a level of theory is “impoveryished” or “gross [and] underdetailed” (Grodzinsky), such dismissals would make sense if the functional architective were as well known as say the chemical elements. Grodzinsky may hold this to be true of language (although that seems an overoptimistic view of linguistics), but it is certainly not true for much of the rest of the cognitive system.

4.2. On connectionist modelling. Most of the arguments in the book have as a premise the existence of modularity. As Allen, Andrews, Goshen-Gottstein, and Hunt point out, this is a working hypothesis that makes sense of much neuropsychological data rather than something for which the book provides extensive support in the way that, say, Fodor (1983) tries to do. The book is therefore open to the criticism that the neuropsychological evidence could eventually be best explicable in terms of “lesions” to nonmodular architectures.

From critics concerned about this assumption, the most sweeping argument was the one put forward by Goshen-Gottstein. He takes a cognitive psychological analogue of the double dissociation – the short-term memory findings of Watkins and Watkins (1977) that had been held to indicate that two stores are involved in the retention of word span. Lewandowsky and Murdock (1989) have shown, however, that the pattern of data can
be explained with a complex single-store model. Goshen-Gottstein’s general argument seems irrefutable. It remains possible that double dissociations may have analogous explanations in terms of separate changes in two different parameters of a nonmodular architecture. Indeed this was the main message of Chapter 11. There is a world of difference, however, between an abstract possibility and a plausible alternative. Unfortunately, the world of difference, however, between an abstract possibility and a plausible alternative. Unfortunately, Goshen-Gottstein does not consider whether the Lewandowsky/Murdock theory could actually explain the short-term memory patient findings discussed in Chapter 3. The essence of their argument is that connectionist models of cognition, which were clearly independent of the model’s specific mode of operation; in this case, it was a connectionist model using a back propagation algorithm (see Hinton & Shallice 1991). More positively, this particular simulation offers support for another type of neuropsychological inference procedure – the use of the symptom-complex of errors repeated across patients – which is at present out of favour among cognitive neuropsychologists (see Neuropsychology, sec. 2.6). Wherever a lesion was made in this connectionist network, which had a strong attractor structure, and which mapped from a letter representation to a semantic one, the same qualitative pattern occurred with all semantic, visual, and mixed visual or semantic errors.

Hence instead of an error-type’s being a sign of the operation of the component damaged, as in a modular network with all-or-none transmission, the error pattern reflected the architecture of the whole network. This means that if connectionist models become more important than traditional modular ones with all-or-none transmission, then the neuropsychological evidence would still remain relevant. The importance of strong dissociations would decline, however, and that of symptom-complexes may increase. If so, although many specific methodological positions taken in Neuropsychology would need reconsideration, the overall position that the neuropsychological methodology to be adopted should depend on the general theory of the normal cognitive system being assessed would be strengthened.

For nonmodular connectionist models (having no correspondence with any sort of information-processing model) it is quite possible that neuropsychological evidence would be unable to help us understand normal function. The current trend in connectionist modelling seems to be towards greater modularity, not less (Norris 1990; Nowlan 1990), so the prospect does not seem too disturbing for the future relevance of neuropsychological evidence.

5. Criticisms of theories of specific areas

5.1. Short-term memory (Chapter 3). Of the six chapters (3–8) where the approach adopted in Neuropsychology was tested, the one that received the most criticism concerned short-term memory (see Caplan and Cowan). In Caplan’s criticisms of the reification that can be read into my treatment of dissociations (see sect. 3.2 above) he uses my analysis of short-term memory patients as a concrete example. At first glance his discussion of short-term memory reads almost like a paraphrase of the conclusion of Chapter 3. Thus his argument that “if performance on STM tasks reflects the operations of the language processing system applied to these particular tasks, we ought to find evidence that when the structures presented in STM-type tasks have different linguistic features these features affect performance on these tasks” is just what the evidence from normal subjects (Neuropsychology, p. 64) is intended to support. Referring to auditory span tasks, I wrote: “Human experimental psychology has attempted to study phenomena that are essentially in one domain – sentence comprehension – using empirical and theoretical techniques derived from another domain – memory” (p. 67). Where we differ is that I want to reform conceptions of short-term memory whereas Caplan suggests that this concept may be totally useless, merely a reification of a dissociation, and that “performances on STM tasks may be entirely mediated...
by combinations of language processing components, none of which have the properties assigned to STS.” Caplan would have to confront the evidence from normal subjects, I adduce, to support the idea that auditory-verbal span performance (at least for digits and letters) involves primarily a single short-term store (Baddeley 1968; Lyon 1977; Sperling & Speelman 1970; see Neuropsychology, p. 45).

Related arguments also present difficulties for Cowan’s approach to short-term memory from the perspective of attention. He holds that there may be two main contributions to span – a modality-free primary memory and speech-specific rehearsal – with short-term memory patients having a deficit in the rehearsal component or another control process (see Cowan 1988). Yet Cowan (1988) accounts for the auditory modality-superiority effect in short-term memory experiments (Penney 1975) in terms of the “superior encoding of spoken temporal sequences” (p. 168), which suggests the existence of speech specific storage. Moreover, his supporting claim that “suppression of rehearsal has a much more severe impact on speech memory indices...than is generally assumed” is based on the word span. Baddeley and Lewis (1984) report digit span data where the reduction is less impressive, from 7.9 to 5.7 digits. Yet short-term memory patients can have auditory spans of 1–3 digits. Furthermore, Cowan’s argument that short-term memory patients have a problem in the rehearsal component was considered by Shallice and Vallar (1990) and rejected on a number of grounds, only one of which is that the effects of suppression in normal subjects are qualitatively much less severe than the impairment exhibited by some of the patients.

5.2. Disorders of attention (Chapter 13). The criticism by Umiltà of my summary sentence on neglect appears apposite. He points out that later evidence supports the general drift of the chapter, which implicated a problem in shifting the attentional focus to one particular side. Thus Costello and Warrington’s (1987) patient JOH, on whom I placed much stress, neglects both sides – but in different situations – and so he does show that neglect is not just an inability to move the attentional focus to one particular “affected” side. JOH’s difficulties do not conflict, however, with the idea that on any particular occasion the neglect occurs because of a difficulty in moving the attentional focus to one side. Umiltà also points out that a major omission of this chapter was my failure to consider theories of how the movement of an attentional focus might be influenced by activation levels in the contralateral hemisphere (as suggested by Kinsbourne’s (1987) view that activation in a hemisphere is responsible for shifting attention in the contraversive direction). From this perspective JOH’s difficulties could be explained by impairment of attentional control systems in both hemispheres: Carrying out a task that activated one hemisphere more than the other would then leave it more adequately activated than the complementary one.

5.3. Disorders of executive systems (Chapter 14). Three general comments were made about the Supervisory System that was used to model disorders of executive systems. Jerison raises the old argument that a hypothetical subcomponent of this sort is merely a dressed-up version of the homunculus. This is to ignore modern developments in artificial intelligence such as the programs of Fahlman (1974), Sussman (1975), and Laird et al. (1987), which have special parts of their functional architecture responsible for planning, error-correction, and impasse-resolution that are conceptually clearly distinguishable from the part responsible for the realisation of routine activities (Neuropsychology, p. 334). Thus, one of the two major innovations in the most advanced human problem-solving simulation in existence, SOAR (Laird et al. 1987), is that the procedure for resolving conceptual “impasses” is a central feature of its architecture. The function corresponds closely with that of “coping with novelty,” which is a central role of the Supervisory System (see Neuropsychology, pp. 345–50). I am at present engaged in a collaborative attempt to implement parts of the Supervisory System (see also Dehaene & Changeux, in press, discussed in 4.1).

I agree both with Butter & Laeng that further work on the theory should be constrained by neurobiological considerations, and with Bruder that the detailed analyses applied to, say, disorders of perception, memory, reading, and writing also need to be used in this domain. We have recently been analysing more specific impairments of the supervisory system (Shallice & Burgess, in press). As Bruder points out, however, we currently lack the analytical tasks that have proved so useful in other domains. Moreover, for the reasons discussed in Neuropsychology (p. 336), the stable levels of performance required for effective use of analytic tasks may be much more difficult to obtain for investigations of executive functions.

I contrasted a Supervisory-System interpretation of the behaviour of frontal patients with one derived from a classic Rosvold & Mishkin (1961) account of certain findings on the loss of the ability to inhibit “central sets” in frontally lesioned monkeys. This was a somewhat dangerous contrast to make as the inability to inhibit inappropriate ongoing behaviour is in any case seen as one of the functions of the Supervisory System on the Norman-Shallice (1986) theory, and also because one critical experiment, that reported by Knight (1984), was held to discriminate most clearly between the two accounts. Smith objects to my interpretation of the Knight study on the grounds that it is plausible that P300 (a scalp-recorded electrical potential) plays an inhibitory role. [See Verleger: “Event-Related Potential and Cognition: A Critique of the Context Updating Hypothesis and an Alternative Interpretation of P3,” and Donchin & Coles: “Is The P300 Component a Manifestation of Context Updating?” BBS 11(3) 1988; Näätänen: “The Role of Attention in Auditory Information Processing as Revealed by Event-related Potentials and Other Brain Measures of Cognitive Functioning” BBS 13(2) 1990.] If this were the case it would undermine my (inadequate) summary suggestion that “the frontal response was not one of inadequate inhibition” (p. 350). The critical issue, however, concerns the generation of the special P300 to unexpected novel stimuli recorded from fronto-central regions of the brain, which is affected in Knight’s (1984) frontal lobe group, a finding that fits with the Supervisory System account of inappropriate responses to novel situations. Can this particular P300 be said to involve inhibi-
tion of central sets more than the standard P300 to infrequent targets, which was normal in the frontal lobe patients? An explanation could perhaps be developed but it would have to be more specific than just the claim that the P300 plays an inhibitory role. Smith is right however, that it is insufficient to deny the inhibition of central sets account simply on the basis of clinical neurophysiological findings concerning an ERP (event-related potential) component, which is itself poorly understood. Computationally, an adequate response to novel stimuli requires far more than mere inhibition of ongoing procedures (Laird et al. 1987; Sussman 1975). It remains conceivable that this is the only aspect of Supervisory System operations that is frontally localised, but this seems unlikely. (For further relevant discussion, see Shallice & Burgess 1991).

A fascinating potential convergence that was unknown to me when I wrote Neuropsychology is suggested by Frith. He points out that it is possible to characterise "negative" schizophrenia symptoms in theoretical terms related to those I had used to account for the impairments exhibited by frontal lobe patients. Moreover, at least some neuropsychological evidence (Kolb & Whishaw 1983) points in the direction of an impairment of schizoprenic patients on frontal lobe tasks, as well as the PET scanning evidence quoted by Frith. In my view, conceiving of schizophrenic symptomatology in terms of a disorder of the Supervisory System could well prove very fruitful. [See also Gray et al.: "The Neuropsychology of Schizophrenia" BBS 14(1) 1991.]

5.4. Amnesia (Chapter 15). Besides criticising my account of Knight’s (1984) experiments on P300 responses to novelty discussed in the previous section, Smith also objects to my account of classical amnesia, arguing that it is purely in terms of retrieval, which "reflects selective reporting of neurological data and a failure to simplify through reduction." My explanation was not concerned solely with retrieval. I suggested that there were at least three possible forms of disconnection theory (see Chapter 15.5). Of these, only the second would correspond to a retrieval impairment position. I somewhat preferred the third (see p. 373), in which posteriorly localised episodic records need to be connected at presentation with more anterior "E-MOPs" or "headings" (see p. 374) for adequate retrieval to occur later. Even if the hippocampus were no more than a staging post in the formation of such a connection, any impairment of hippocampal function would lead to the connections being inadequately formed. Thus the evidence referred to on transient global amnesia and on electrical stimulation of medial temporal lobe structures would be compatible with the account given.

Why did I stress the posterior-frontal disconnection theory rather than the consolidation theory associated with the hippocampus favoured by Smith? The two theories are not in conflict and obviously in a chapter of this length, given the enormous relevant literature, coverage was inevitably selective, with respect to both theories and findings. The first reason for my selection was the evidence for severe amnesic patients with lesions sparing the hippocampus (p. 369), as in Korsakoff patients with lesions involving the mamillary bodies and thalamus (e.g., Mair et al. 1979), in thalamic stroke patients (e.g., Gentilini et al. 1987; Graff-Radford et al. 1990; von Cramon et al. 1985) and in patients with tumours affecting the fornix (e.g., Rudge & Warrington, in press; Valenstein et al. 1987). Smith alludes to this evidence, but dismisses it on the grounds that there are direct cortical efferents from the hippocampus to the parahippocampal gyrus and also from there to the frontal cortex. Thus he argues that lesions in the Papez circuit could not disconnect posterior from frontal cortices. The region of the frontal cortex involved in the work of Goldman-Rakic (1988b) he cites, sulcus principalis, may not have any relation to long-term episodic memory; Goldman-Rakic (1987), for example, argues that it is specialised for spatial working memory. By contrast, the eventual target of the hippocampal-mamillary-thalamic connections could well be some other anterior region. Moreover, if like Smith, one rejects the disconnection theory, how else are amnesias arising from the lesion sites mentioned above to be explained?

A second and more important reason for being concerned with the disconnection account of amnesia is that a major strand of Chapter 15 was that in its classical form as found, say, in pure Korsakoff patients amnesia involves a very long-lasting retrograde component. Contrary to much current opinion, I argued that the retrograde amnesia was not dependent on the age of the memory but on its nature, that is, as being episodic as opposed to semantic.

If any consolidation process existed it would not relate to this characterisation of the syndrome (although the characterisation would be entirely compatible with theory that the hippocampus indexes episodic traces). Yet Smith does not challenge my arguments about retrograde amnesia. Moreover, Warrington and McCarthy (1988) have reported evidence that, in an amnesic patient, the meanings of words learned during the period affected by a dense retrograde amnesia were retained, which indicates that it is the nature of the material rather than the age of the trace that is critical. Instead, Smith cites the recent paper of Zola-Morgan & Squire (1990), which provides very interesting evidence for a consolidation process lasting a few weeks. It should be noted that the critical evidence favouring a consolidation process, as opposed to just the presence of the permastore type of trace that Bahrick (1984) characterised semantic memory as containing, is the better performance of hippocampal monkeys on older rather than more recently learned items; this evidence is based on selective post-hoc tests, however, and not the more appropriate trend test. Moreover, it could be a characteristic of semantic memory in general that items in a particular domain learned earlier produce somewhat stronger traces than equivalent items learned later. This would fit with the known excessive sensibility of amnesics to proactive interference (Warrington & Weiskrantz 1978; Winocour & Kinsbourne 1978). The most crucial point, however, is that the time-period of the operation of any form of consolidation process of the kind Zola-Morgan and Squire posit—a few weeks—is the wrong order of magnitude for the retrograde components of classical amnesia.

5.5. Consciousness (Chapter 16). Many commentators referred to the functionalist approach to consciousness that I adopted to interpret the neuropsychological syndromes that relate to awareness. Frith differed from the others in proposing an approach similar to mine in an-
other empirical domain — that of schizophrenic experiences such as delusions of alien control. If, as discussed earlier, the cognitive disorders of schizophrenia are to be understood, at least in part, as disorders of the Supervisory System then such patients should also have abnormal phenomenal experience.¹

As various commentators (e.g., Allen, Flanagan) noted, there was an unresolved tension in my position on consciousness between, on the one hand, the idea that no single system or subsystem subserves conscious experience and, on the other hand, the central role ascribed to the Supervisory System and the use of the William James analogy about consciousness as the property of an organ. Allen, Bridgeman, and Flanagan all agreed with me in rejecting the idea that consciousness is attributable to a single component. Others (Baars, Cowan), however, presupposed that consciousness derives from the operation of one single subsystem — a global workspace, or the Supervisory System.

I do not find the latter view very plausible for reasons similar to those discussed by Flanagan. To argue, as Baars does, for the relevance of a single system that “any conscious (or effortful) tasks will compete against any other” is in my view, to overestimate the degree to which two routine but demanding tasks that do not require the same processing structures compete (see e.g., Allport et al. 1972, Shallice et al. 1985), and to ignore the possibility that the interference might result from “cross-talk” rather than demands on a common subsystem. It also oversimplifies the phenomenology of dual task performance (see Shallice 1988). Moreover, I see no reason to assume, as Baars does, that the workspace of complex AI problem-solving systems can be considered isomorphic with the contents of consciousness; this would not be the case, for example, with the contents of SOAR’s short-term memory (Laird et al. 1987). More critically, as I stated (p. 401), no syndrome fits well with the characterisation of damage to a globally distributed database. Umlitá makes a more general point of the same type.

Allen makes two main criticisms of the specific positions I took on consciousness. First, he argues that my somewhat frivolous justification for ignoring the problem of qualia when adopting a functionalist approach is inadequate. However, I fail to see why it is not a reasonable strategy to operate on the assumption that a difficulty that is “conceptual” (Allen) may not be resolved at some future date as scientific ones can be.

Allen’s second objection is that the assumption that the conscious/nonconscious distinction concerns the concepts an organism uses to understand itself is “mysterious unless we know where such understanding lies.” I agree that my theorising focused on what internal representations, and particularly those in what system, are and are not conscious. I barely touched the question of what additional thoughts or actions become available when one becomes conscious of some mental content, and especially what it means for these states to be known to the organism. All three of these issues can in principle be tackled by the methods of cognitive science, however. Hence, although they are now mysterious, I do not see why they need to remain so.

If scientific answers can be provided for such questions it will also resolve certain of the problems I had with the comments of Flanagan. I feel particularly sympathetic to his criticisms. In my first writing on consciousness (Shallice 1972) I tried to show that the information-processing functions related to consciousness did not require the assumption of a single controlling subsystem. The later neuropsychological and artificial intelligence literature led Norman and me to support the idea of a Supervisory System. Such a system would have a less central role in cognition than, say, the controlled processing of Shiffrin and Schneider (1977). Does it then deserve its name and its capital letters? The other control systems (apart from contention scheduling) discussed in Neuropsychology were seen to be evolutionarily dependent on and fully interconnected with the Supervisory System. The existence and typical operation of these interconnections were taken to provide the material basis for the concept of consciousness.

What would happen if the Supervisory System fractionated into separable subsystems? The evidence of Petrides (1987) on separable specific higher-level frontal lobe disorders, together with the tight interconnections of the relevant frontal lobe regions with particular regions of parietal lobe (Goldman-Rakic 1987), might suggest that different modulatory subsystems exist that are fairly independent of each other computationally. This is probably too extreme a view. In any case, among the scientific questions to be addressed by future research are: How close is the interdependence of the control subsystems with each other? and do they conceptually form a system, or a set of intercommunicating but evolutionarily distinct subsystems? Or do they have little or only negative (i.e., inhibitory) communication with each other? In my view, it is the answer to such scientific questions that will determine our future understanding of consciousness (as well as whether Flanagan is right to dismiss the role of individual systems so completely).

NOTE

1. A minor empirical point: Jerison is right that blindsight was conceptually derived from monkey work, at least as far as the Weiskrantz et al. (1974) study was concerned. [See also Campion et al.: “Is Blindsight an Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision?” BBS 6(3) 1983.]

References

References/Shallice: Neuropsychology


---

BBS Associates

Please send BBS your electronic mail ('email') address

BBS is relying more and more on electronic mail to communicate with Associates (especially for circulating abstracts of recently accepted target articles so prospective commentators can nominate themselves).

If you have an email address, please inform BBS right away at:

harnad@clarity.princeton.edu
or
harnad@pucc.bitnet

(If you do not yet have an email address, we urge you to get one!)

---

21st Annual Meeting of the Society for Computers in Psychology
San Francisco, California
November 21, 1991

The 21st Annual Meeting of the Society for Computers in Psychology will be held at the Hyatt Regency Hotel in San Francisco on November 21, 1991, the day before the annual meeting of the Psychonomic Society. The meeting includes presentations, discussions, tutorials, and times for software and hardware demonstrations. The application of computer-based solutions to all areas of psychology are featured, including research, education, clinical practice, and industrial applications.

For further information contact William L. Palya, Department of Psychology, Jacksonville State University, Jacksonville, AL 36265 (BITNET address FWLP@JSUMUS). Phone (205) 782-5641. Fax (205) 782-5680.