PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Dissociating Executive Functions of the Prefrontal Cortex [and Discussion]

T. W. Robbins, D. Weinberger, J. G. Taylor and R. G. Morris

Phil. Trans. R. Soc. Lond. B 1996 **351**, 1463-1471
doi: 10.1098/rstb.1996.0131

Dissociating executive functions of the prefrontal cortex

T. W. ROBBINS

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, U.K.

SUMMARY

An analysis is provided of three distinct paradigms that have been used to study executive functions of the prefrontal cortex involving planning, self-ordered memory or attentional set-shifting. Psychological and anatomical dissociations are sought from the perspective of studies of patients with frontal lobe lesions, functional neuroimaging, psychometric studies in normal volunteers and experimental studies in non-human primates. Particular attention is paid to attempts to dissociate mnemonic from other executive capacities. Thus, patients with frontal damage are shown to have deficits in their (1) use of strategies to improve performance in a spatial working memory task and (2) capacity to make an extra-dimensional shift due to a high-order failure of inhibition in an attentional set-shifting paradigm. These results are discussed in terms of anatomical and neuropharmacological dissociations of different aspects of executive function within the prefrontal cortex shown in monkeys.

1. INTRODUCTION

The term 'executive functioning' generally refers to those mechanisms by which performance is optimized in situations requiring the operation of a number of cognitive processes (Baddeley 1986). Executive functioning is required when effective new plans of action have to be formulated, and appropriate sequences of responses must be selected and scheduled. Components may include the enhancement of information held temporarily or 'on line' (cf. Goldman-Rakic's 1987 concept of 'working memory'), the marshalling of attentional resources (Shallice 1982), the inhibition of inappropriate responses in certain circumstances (Shallice & Burgess 1993) and the 'monitoring' of behaviour with respect to affective or motivational state (Damasio 1994; Petrides 1996). Some insight into the nature and organization of executive functions can be gleaned from a psychometric approach (e.g. Duncan *et al.* 1995), but it is still not established to what extent they represent emergent properties of interactions between specialized cognitive subsystems or the operation of a single central executive (Baddeley 1986), possibly with dissociable components (Shallice & Burgess 1993).

As many of the deficits shown by patients with frontal lobe dysfunction are of an executive nature, a neural perspective may be useful. Equivalence between the prefrontal cortex and executive functioning cannot of course be assumed, especially in view of the occurrence of the 'dysexecutive syndrome' in patients with damage to other brain regions. However, it may be profitable to consider possibly distinct forms of executive dysfunction in the context of the considerable anatomical connectivity existing between the prefrontal cortex and other regions of the cerebral cortex, as well as the subcortical brain. Four particularly important forms of neural interaction involving the prefrontal cortex are with: (1) dedicated processing modules of the posterior cortex such as the parietal and temporal lobes; (2) limbic structures such as the amygdala and hippocampus; (3) the output of the striatum, which targets the frontal lobe; and (4) the ascending monoaminergic and cholinergic systems of subcortical origin, which exert potentially diverse effects on forebrain functioning. The fact that the prefrontal cortex itself is not a homogeneous structure, having several distinct cytoarchitectonic regions, also has implications for the nature and organization of executive functions (Pandya & Yeterian 1995).

2. VALIDATION AND STANDARDIZATION OF TESTS SENSITIVE TO FRONTAL LOBE DYSFUNCTION

Our overall aim is to seek illuminating functional dissociations of executive and non-executive mechanisms, via comparative studies of brain-damaged humans and monkeys, psychometric analyses and studies of functional neuroimaging in intact human subjects. To this end, we have devised tests with executive as well as non-executive components for testing patients with frontal lobe damage.

For these studies, groups of patients with frontal lobe excisions, together with age- and IQ-matched controls, have been used (for full clinical details see Owen *et al.* 1990, 1993, 1996*a*). These patients ($n = 40$) include cases resulting from neurosurgery undertaken generally for the relief of epilepsy or the removal of tumours. The majority of cases are unilateral with a rough equivalence of numbers according to the side of the lesions. The group of frontal lobe lesioned patients is heterogeneous with respect to the exact location and extent of the lesion (confirmed on the basis of magnetic resonance imaging (MRI) or computerized tomographic (CT) scans and the notes of the neurosurgeon C. E. Polkey). Groups of patients with other cortical lesions or neurodegenerative diseases have been studied as

positive controls, including temporal lobe excisions, amygdalo-hippocampectomy and basal ganglia disorders. Such controls may help to identify the nature of possible interactions with other components of the distributed neural networks in control of performance (see above), as well as helping to define the precise contribution of the prefrontal cortex itself.

The three main types of test probe aspects of executive functioning involving planning, working memory, response control and attentional shifting, and have been found to be sensitive to damage of the prefrontal cortex. The test of planning is based on the Tower of London task (Shallice 1982) but modified in several ways to obtain several independent measures of performance (Owen *et al.* 1990) and to maximize its 'look ahead' requirements (Owen *et al.* 1995*b*). The attentional set-shifting paradigm (Roberts *et al.* 1988; Owen *et al.* 1991) represents the decomposition of a test much used in clinical assessments of frontal lobe damage (the Wisconsin Card Sorting Test or WCST), according to the principles of learning theory. The third test, of working memory (Morris *et al.* 1988; Owen *et al.* 1990), is based loosely on Passingham's (1985) adaptation of a food searching test for monkeys with dorsolateral prefrontal lesions, and on the 'self-ordered' memory tasks of Petrides & Milner (1982). The latter test not only measures memory performance, but is also amenable to analyses of the strategies adopted by subjects performing the task.

Data have been obtained from a large group of normal volunteers using these tests, as well as others from the Cambridge Neuropsychological Test Automated Battery (CANTAB), and subjected to factor analysis (Robbins *et al.* 1994, 1996). This analysis indicates that measures from some of the three main tests load on common factors (e.g. spatial working memory and planning), but some do not (attentional set-shifting). Further analyses show that tests of visual recognition memory or learning, or other tests of frontal lobe function such as verbal fluency, load on an independent factor. Moreover, a test of non-verbal 'fluid intelligence' (see Duncan *et al.* 1995), AH 4–2, loads across virtually all factors, including those that capture tests sensitive to frontal lobe dysfunction and those that do not (see Robbins *et al.* 1996).

This analysis indicates that these tests measure dissociable aspects of executive functioning which contribute to task performance over and above more basic cognitive functions, such as visual perception and short term storage. Moreover, the analysis indicates possible dissociations between basic memorial requirements of tasks and their executive components which control response selection, such as the adoption of an overall strategy or plan, or the utilization of specific attentional inhibitory mechanisms. This general hypothesis can be tested further by dissecting performance on these tasks at a neural level, for example by studies of humans and animals with defined lesions, or using functional neuroimaging.

3. STRATEGIC VERSUS MEMORIAL FACTORS IN PERFORMANCE

(*a*) *Self-ordered working memory*

Successful performance in control subjects on the self-ordered spatial working memory task often exemplifies a searching strategy which essentially retraces the 'route' previously employed by the subject in searching through the spatial array of boxes, but 'monitors' or 'edits' it to avoid previously reinforced locations. This strategy can be captured by an index which is demonstrably uncontaminated by overall mnemonic performance (Owen *et al.* 1990, 1996*a*) and yet which correlates highly with such performance (Owen *et al.* 1990; table 1). Use of this strategy can thus markedly reduce the load on memory caused by interference from previous unreinforced choices.

About 70% of frontal patients we have tested are markedly impaired (> 1 standard deviation from the control mean) on the self-ordered spatial working memory task, as shown by large increases in responses to locations that were previously reinforced ('between-search errors'). However, this apparent deficit in 'working memory' is accompanied by impairment in the use of an effective strategy (Owen *et al.* 1990, 1996*a*), suggesting that at least part of their deficit on the task arises from executive failure. Several other patient groups have equivalent deficits in overall performance on the spatial working memory tests, as measured by the between-search error score, but no significant deficits in the use of strategy. These groups include patients with early-in-the-course Alzheimer's disease (Sahgal *et al.* 1992), Parkinson's disease (Owen *et al.* 1992) and temporal lobe excisions or amygdalo-hippocampectomy (Owen *et al.* 1995*a*, 1996*a*). While patients with Huntington's disease (Lawrence *et al.* 1996) and Korsakoff's syndrome (Joyce & Robbins 1991) also exhibit marked impairments in strategy, as well as memory performance, these deficits might be expected to arise from the disruption of neural circuitry intimately associated with the prefrontal cortex, namely the striatum and mediodorsal thalamus, respectively. It would seem most unlikely that the deficit in strategy formulation is secondary to an impairment in the ability to remember the previous sequence of choices, as (1) a measure of a computerized form of the Corsi spatial span task showed no significant differences between frontal lesioned patients and controls; and (2) correlations between this span measure and the adoption of strategy are generally low and non-significant in frontal patients (table 2; Owen *et al.* 1996*a*), as well as in the normal population, especially when the stronger correlations between span and spatial working memory performance, and spatial working memory and strategy scores are partialled out (see table 1). This pattern of correlations remains robust when the effects of age, verbal and 'fluid' IQ (as measured by AH4–2) are partialled out, and suggests that performance on the spatial working memory task is governed by two major factors, one related to short-term spatial memory and the other to strategic factors.

Our data provide empirical support for suggestions made from previous studies (Passingham 1985; Pet-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Inter-correlation matrix for task performance; normal subjects* ($n = 200$)

(The tasks and asscociated measures are described in detail in references cited in the text.)

	extra-dimensional shift	span	spatial working errors	memory strategy	Tower of London
extra-dimensional shift (errors)	–	−0.227**	0.160	0.141	0.160
spatial span		–	−0.435**	−0.230**	0.199*
spatial working memory (errors)			–	0.518**	−0.379**
spatial working memory (strategy score)				–	−0.326**
Tower of London (perfect solution)					–

* $P < 0.01$; ** $P < 0.001$.

Table 2. *Inter-correlation matrix for task performance; frontal patients* ($n = 19$)

(The tasks and associated measures are described in detail in references cited in the text. NART IQ and age partialled out.)

	fluency	extra-dimensional shift	span	spatial working errors	memory strategy	Tower of London
verbal fluency (FAS)	–	0.252	−0.492*	−0.557*	0.193	−0.040
extra-dimensional shift (errors)		–	−0.299	0.284	0.310	−0.483*
spatial span			–	−0.435*	−0.120	0.098
spatial working memory (errors)				–	0.446*	−0.336
spatial working memory (strategy score)					–	−0.406*
Tower of London (perfect solutions)						–

* $P < 0.05$.

rides & Milner 1982) of an executive contribution to the 'frontal' deficits on such self-ordered tasks over and above their memorial requirements. If this is true there should be a smaller deficit in frontal patients on similar tasks where the strategic component is less prominent. This prediction has been confirmed using two tasks designed by A. M. Owen. They are direct verbal and visual analogues of the spatial working memory task described above. However, instead of searching spatial locations, subjects are required to search through changing arrays of abstract visual patterns or words without semantic content (based on surnames taken from the London telephone directory) (Owen *et al.* 1996*a*). For controls, there is considerably less evidence of an effective strategy based on the 'updating' of a repetitive sequence in either case. And there is no significant deficit on either task in the frontal group, although both tasks are considerably more difficult than the spatial working memory analogue. There is no obvious explanation of this difference in terms of the exact sites of the patients' lesions in the prefrontal cortex. However, performance is significantly impaired in the visual working memory task for a group of temporal lobe lesioned patients (Owen *et al.* 1996*a*), and for both tasks in a group of patients with Parkinson's disease (unpublished observations).

Therefore, it appears that the executive deficit in the spatial working memory task in frontal patients has a considerable degree of selectivity, but is not simply produced by damage to a short-term memory storage system. It also seems likely that performance on the task is a product of interactions between posterior cortical structures (including the hippocampus). Moreover, it is evident that frontal patients are not invariably impaired on difficult tasks that might be expected to depend on high levels of general intelligence. Rather, it appears that deficits are more evident when there are important strategic components to the task that might reflect the operation of an executive or 'supervisory attentional' system.

(*b*) *Planning*

The finding of impaired use of strategy in frontal patients is consistent with the clinical impression of their inability to plan in everyday life. However, there has been only limited empirical evidence for this view (Shallice & Burgess 1993). Our modifications of the Tower of London (or Stockings of Cambridge) test, in two separate formats, emphasize different aspects of planning, namely the reproduction of the appropriate motor sequence and the manipulation of appropriate mental imagery. In the first format the subjects solve problems by rearranging a configuration of stimuli in the bottom half of the screen to resemble a 'goal' configuration in the top half. A 'yoked motor control' condition enables thinking time to be estimated by subtracting motor latencies to make each move of the sequence. Using this test format, we have shown that frontal patients (irrespective of lesioned side) were less efficient in generating accurate solutions than controls. They also had longer 'thinking times' after the first move, suggesting that they had initiated solutions without fully 'thinking them through' (Owen *et al.* 1990). By contrast, patients with temporal lobectomies were not significantly impaired in any aspect of performance (Owen *et al.* 1995*a*).

It is important to distinguishing the three disc Tower of London puzzles from the four disc Tower of Hanoi tasks which depend on trial and error learning. Goel &

Grafman (1995) criticize these latter tests for assessing planning functions because of the contaminating requirements to inhibit prepotent responses. As they concede, these difficulties are mainly avoided in the Tower of London problems. The difficulties are avoided to an even greater extent by our second version of the Tower of London problems, which can be solved by only a single overt response. For these problems, the subject inspects an array with two different arrangements of coloured balls. He or she is required to indicate the number of 'moves' it would take to transform the bottom arrangement into the top by touching an appropriate response panel labelled from 1 to 5. Both the accuracy and latency of responding can be measured. Frontal patients are again less accurate than controls, but do not take significantly longer to arrive at their correct choices, unlike patients with Parkinson's disease (Owen *et al.* 1995*b*). Therefore, it is difficult to argue that the effects arise merely from non-specific deficits in response control that produce impulsive forms of responding.

(*c*) *Functional neuroimaging in normal volunteers*

The design of both forms of the Tower of London test makes them suitable for studies of functional neuroimaging in normal subjects. When the 'mental' format was used changes in regional cerebral blood flow (rCBF) in both 'easy' (2–3 move) and 'difficult' (4–5 move) problems were compared using positron emission tomography (PET) with ^{15}O. A further comparison condition was with an array which required the subject to sustain attention to it before making a similar motor response (Baker *et al.* 1996). Both types of comparison indicated that the 'mental' task produced increases in rCBF in a distributed cortical network that included the superior occipito-parietal cortex and three main zones in the frontal cortex: the premotor area, a band of activation including the dorsolateral prefrontal cortex (Brodmann's areas 9/46) bilaterally and the frontopolar (Brodmann's area 10) cortex on the right. Similar areas were activated in a parallel study performed with the 'motor' format of the task, except that activations were greater on the left side (Owen *et al.* 1996*b*). This was conceivably because of the different requirements of the two tasks on spatial working memory and action production, although the differences might have arisen in part from a different type of comparison condition, in the latter case involving a memory sequencing task.

An important conclusion is that a complex planning task in humans activates discrete regions of the prefrontal cortex, predominantly in the dorsolateral and polar regions, in conjunction with posterior cortical systems implicated by other studies in the manipulation of mental imagery. The activation of the prefrontal cortical sites presumably reflects a functional interaction with this posterior system that leads to the generation and monitoring of candidate sequences of moves to solve the problems. There are striking parallels with the self-ordered memory task described above, where the goal is less clearly specified, but there is a more obvious memory load. This is seen not only from the correlations between performance on the memory and planning tasks shown in table 1, but also from recent functional imaging studies of a close analogue of our self-ordered task (Owen *et al.* 1996*c*). These studies indicated two major sites of activation in the dorsolateral and ventrolateral prefrontal cortex. Further data suggest that the ventrolateral area (Brodmann's area 47) may play a role in the 'passive' receipt of items for memory (as in spatial span performance), whereas the dorsolateral prefrontal cortex is important for the 'monitoring' role by which candidate sequences are compared with the goal sequence. This dissociation of two distinct factors contributing to the self-ordered spatial working memory task thus accords with the psychometric evidence shown in table 1. Further studies are required to elucidate the role of the right frontopolar area (Brodmann's area 10) in the planning task.

4. ATTENTIONAL SET SHIFTING VERSUS MEMORIAL FACTORS IN PERFORMANCE

Novel circumstances often dictate that the processes governing response selection in predictable situations must be countermanded to allow the expression of new intentions. A classical example is the fixation with particular ways of solving problems (or 'Einstellung' of the Gestalt psychologists) which interferes with the production of flexible strategies for problem solving. In the clinical context, perseveration on the WCST has often been used as an indication of frontal lobe dysfunction (Milner 1963), although there has been considerable disagreement about the psychological and neural specificity of this test in recent years. Some authors have argued that it provides yet another test of 'working memory' in which the subject has to keep the currently valid sorting principle constantly in mind. There has also been considerable disagreement over the extent to which it represents a 'pure' frontal test (Anderson *et al.* 1991), although the notion of a 'pure' test in itself is misleading, as it is apparent that neuropsychological tests consist of many interactive components that depend on the outputs of different neural systems. The obvious possibility that the test depends on discrete regions within the prefrontal cortex has never been resolved; there is no agreement concerning either the possible laterality of the effects, or their critical zone. An informative recent functional imaging study employing PET identified several neocortical regions that were activated by the WCST task, including both dorsolateral and orbitofrontal regions, as well as both temporal and parietal lobe structures (Berman *et al.* 1995) .

We have also undertaken a functional analysis of the WCST, which recognizes that it has several cognitive components, at the core of which is the extra-dimensional shift test, derived from human and animal learning theory (Downes *et al.* 1989; Roberts *et al.* 1988). In an extradimensional shift, attention to compound stimuli is transferred from one perceptual dimension (e.g. colour) to another (e.g. form), on the basis of changing reinforcement or feedback. Control tests include the ability to shift performance on the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

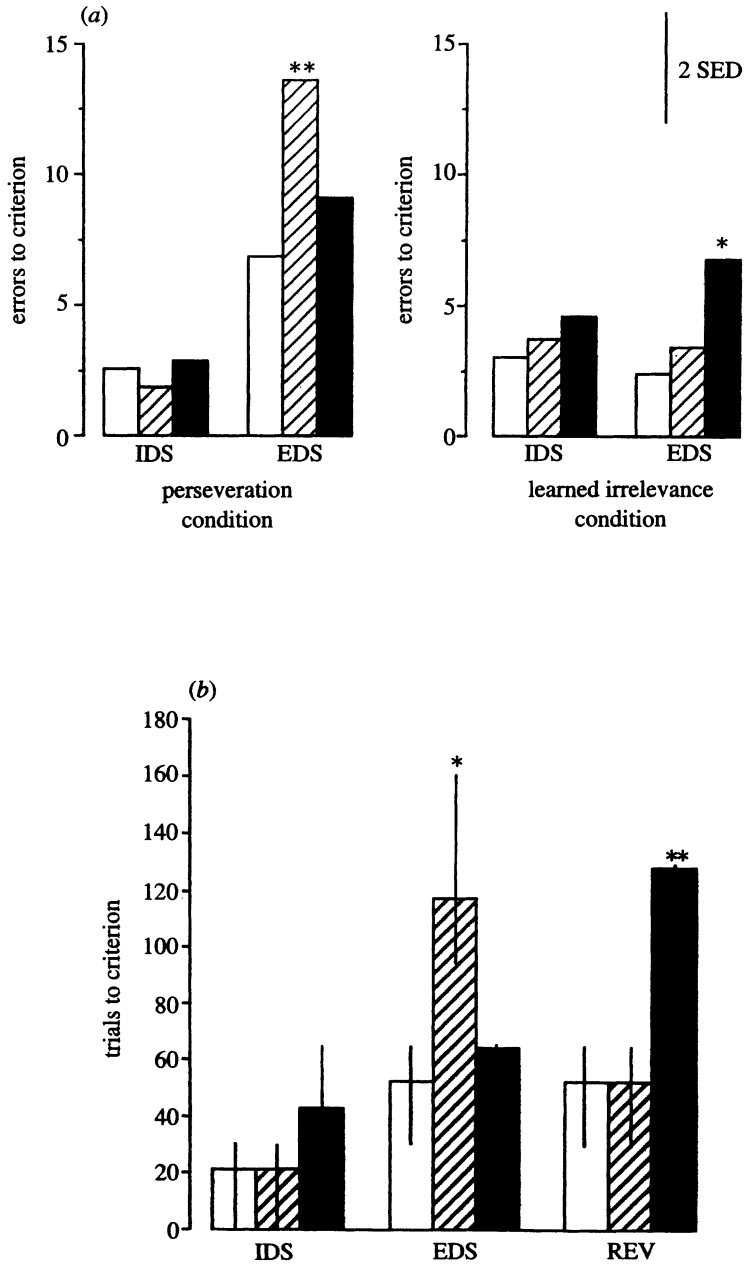

Figure 1. (*a*) Errors to criterion for the intradimensional shift (IDS) and extradimensional shift (EDS) stages of the attentional set-shifting paradigm for the perseveration and learned irrelevance conditions by groups of patients with frontal excisions (hatched bars) or Parkinson's disease (PD) (medicated and unmedicated groups combined) (shaded bars), as well as age- and IQ-matched control subjects (open bars). Replotted from Owen *et al.* 1993. ** $P < 0.01$; * $P < 0.05$, compared with controls. 2SED = 2 standard errors of the differences between means.(*b*) Trials to criterion for intradimensional shift (IDS), extradimensional shift (EDS) and reversal (REV) stages of the attentional set-shifting paradigm for marmosets with lesions of the lateral (hatched bars) or orbital prefrontal cortex (solid bars), or with sham surgery (open bars) (data replotted from Dias *et al.* 1996). Vertical lines refer to the range of scores. ** $P < 0.01$; * $P < 0.05$ compared with the other groups.

basis of altered feedback within a dimension (reversal learning) and the ability to shift performance to novel exemplars of the same dimension (intra-dimensional shift) (see Roberts *et al.* 1988; Owen *et al.* 1991).

About 50% of patients with frontal lobe damage have been shown to exhibit selective difficulties in reaching a learning criterion at the extra-dimensional shift, whereas patients with temporal lobectomies or amygdala-hippocampectomies were unimpaired (Owen *et al.* 1991). However, both of these temporal lobe lesions had some impact on performance at the extra-dimensional shifting stage, where the latencies to respond were significantly lengthened (Owen *et al.* 1991). Thus, this is the third of our tasks that requires processing in the posterior (in this case, temporal) cortex, presumably in conjunction with prefrontal mechanisms.

Patients with basal ganglia disorders also exhibit major deficits on the attentional set-shifting task, although these may differ qualitatively from those seen after frontal damage, in two main ways. There is a greater incidence of failures in patients with striatal

dysfunction at those stages when the set is being acquired (Owen *et al.* 1992). They may also fail the extra-dimensional shift in a different way than frontal patients. This can be shown by changing one of the perceptual dimensions at the extra-dimensional stages with a novel dimension that replaces either the previously reinforced or non-reinforced dimension. Subjects are required either to shift responding to the novel dimension, ignoring the previously reinforced dimension, or to shift to the previously non-reinforced dimension, ignoring the novel one. These two conditions can be termed 'perseveration' and 'learned irrelevance', respectively. In simple terms, the subject has to inhibit responding to the previously reinforced dimension in the perseveration condition and to overcome inhibition to respond to the previously non-reinforced dimension in the learned irrelevance condition. The frontal patients have a selective deficit in the perseveration condition, whereas patients with basal ganglia disorders make more errors in both conditions (Owen *et al.* 1993; see figure 1*a*).

These results identify the exact nature of the extra-dimensional deficit in frontal patients as a specific failure of response inhibition leading to perseveration. The results also show that it is difficult to explain these patients' shifting impairment simply in terms of 'holding stimuli on-line in memory'. There was no deficit in the learned irrelevance condition which would appear to provide a similar load for memory as the perseveration condition. This argument also holds for the intra-dimensional stages of the task, where similar compound stimui were employed and no deficits observed in the frontal group for these forms of associative learning. Thus, it can be argued from these data that frontal patients have major deficits in response inhibition that are not secondary to problems in short-term memory. These conclusions are substantiated by findings of intact extra-dimensional shift performance in the face of major short-term memory deficits in patients early in the course of Alzheimer's disease (Sahakian *et al.* 1990).

The failure of attentional shifting is probably importantly determined by the novelty, and hence learning requirements, of the test situation. Some evidence for this is provided from a recent study by R. D. Rogers in this laboratory in which subjects have been trained to shift continuously from one dimension ('letters of the alphabet') to another (digits), on the basis of learned cues. A group of patients with frontal excisions showing deficits on the perseveration component of the attentional set-shifting task failed to show any impairments on this pretrained shifting task, although considerable and equivalent costs of shifting were entailed in both the frontal lesioned and normal subjects.

5. ANATOMICAL AND PHARMACOLOGICAL DISSOCIATIONS OF EXECUTIVE DYSFUNCTION

The pattern of deficits shown by individual frontal patients in our sample (see also table 2) suggests the possibility of functional dissociations with anatomical correlates. However, interpretation of such data is compromised by the inconsistency and transient nature of frontal deficits, as well as the variability in the size and extent of the damage within the prefrontal cortex, which often fails to respect anatomical boundaries. Such considerations justify the use of tasks that can be used in experimental animals, as well as in humans, to study the effects of well-defined experimental lesions (cf. Goldman-Rakic 1987; Petrides 1996).

Our attentional set-shifting paradigm is one such task which can be used to study the neural basis of different aspects of response inhibition in monkeys. Thus, Dias *et al.* (1996) have studied the effects of excitotoxically-induced, localized lesions of the prefrontal cortex of marmosets on performance of the intra-dimensional and extra-dimensional shifts, as well as on the reversal of associations for the compound stimuli following the extra-dimensional stage. A double dissociation of effects was found for lesions to the lateral prefrontal cortex (designated as area 9 by Brodmann) and to the orbitofrontal cortex (Brodmann's areas 11, 12 & 13). Specifically, the lateral lesion selectively impaired extra-dimensional shifting, whereas the orbitofrontal lesion selectively impaired reversal learning (see figure 1*b*). Neither lesion affected learning at the intradimensional stage.

These results suggest that distinct processes of response inhibition are recruited to control extra-dimensional shifting and associative reversal learning, represented in separable processing domains within the prefrontal cortex. Other evidence for the dissociation is provided by a demonstration of opposite patterns of effects of cortical cholinergic depletion, following lesions of the basal forebrain and of prefrontal dopamine depletion, produced by 6-hydroxydopamine, in marmosets (Roberts *et al.* 1992, 1994). Cholinergic depletion selectively impaired serial reversal learning, whereas prefrontal dopamine loss actually selectively enhanced extra-dimensional shifting. Furthermore, it is important to emphasize that these effects in monkeys, like those in humans, cannot be reduced to deficits in 'holding stimuli on line', as shown by the fact that marmosets with prefrontal dopamine loss that exhibited enhanced attentional shift performance were significantly impaired in the classic delayed response task (Roberts *et al.* 1994). Analogous dissociations between the test of spatial working memory described above, and other 'executive' functions have recently been reported following treatment of patients with dementia of the frontal lobe type with the alpha-2 adrenoceptor receptor antagonist idazoxan (Coull *et al.* 1996). The involvement of these subcortical systems may represent ways in which arousing or rewarding signals influence prefrontal function; these signals may potentially disrupt, as well as enhance processing in their terminal domains, depending on the test situation.

At a theoretical level the results show that these identical compound stimuli can be processed at more than one site in the prefrontal cortex, perhaps simultaneously. At one of these sites, processing allows a 'reward' tag to be shifted from one stimulus to another ('affective shifting'). At the other site, shifts

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

are effected between responding to different dimensions (e.g. shapes rather than colours) of complex stimuli, rather than to particular exemplars, a seemingly higher-order cognitive process. On the basis of previous experience, a 'set' is established for the control of responding, perhaps on the basis of processing in the posterior (e.g. inferotemporal) cortex. However, in novel situations, it is adaptive to relax that control, in order to allow the development of an alternative 'response set'. These data are not necessarily incompatible with hierarchical models of prefrontal cortex that emphasize the serial flow of information from sectors of the prefrontal cortex in receipt of basic perceptual and associative information to other, possibly superordinate regions (such as the dorsolateral prefrontal cortex) (Petrides 1996). However, the present double dissociation is more consistent with the parallel processing of information by relatively independent sectors of the prefrontal cortex (Pandya & Yeterian 1995).

6. CONCLUSIONS

Neuropsychological evidence obtained from subjects with frontal lobe lesions, as well as psychometric and functional neuroimaging data obtained from normal volunteers, have been presented that are consistent with the hypothesis that the prefrontal cortex plays a major and specific role in response selection processes, particularly when these have a possible function in the strategic control of responding. These response selection processes are sensitive to inhibitory influences at several functional levels, including mechanisms by which the effects of particular associations of stimuli with reward, and the superordinate effects of stimulus categories or dimensions, are attenuated, hence facilitating the expression of voluntary behaviour in novel circumstances. These dissociable aspects of executive function appear to be mediated by distinct neural systems that engage different regions of the prefrontal cortex.

I thank B. J. Everitt, A. M. Owen, A. C. Roberts, R. D. Rogers and B. J. Sahakian for discussion, as well as other colleagues for their collaboration and assistance, in particular, North East Age Research (P. Rabbitt and L. McInnes) and the Wellcome Dept. of Cognitive Neurology. The work was supported by the Wellcome Trust.

REFERENCES

Anderson, S. W., Damasio, H., Dallas Jones, R. & Tranel, D. 1991 Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *J. Clin. Expt. Neuropsych.* **13**, 909–922.

Baddeley, A. D. 1986 *Working memory*. Oxford: Clarendon Press.

Baker, S. C., Rogers, R. D., Owen, A. M. *et al.* 1996 Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia* **34**, 515–526.

Berman, K. F., Ostrem, J. L., Randolph, C. *et al.* 1995 Physiological activation of a cortical network during performance of the Wisconsin Card Sorting Test: a positron emission tomography study. *Neuropsychologia* **33**, 1027–1046.

Coull, J. T., Sahakian, B. J. & Hodges, J. R. 1996 The alpha-2 antagonist idazoxan remediates certain attentional and executive forms of dysfunction in dementia of the frontal type. *Psychopharmacology* **123**, 239–249.

Damasio, A. 1994 *Descartes' error*. New York: Putnam Press.

Dias, R., Robbins, T. W. & Roberts, A. C. 1996 Dissociation in prefrontal cortex of attentional and affective shifts. *Nature* **380**, 69–72.

Downes, J. J., Roberts, A. C., Sahakian, B. J., Evenden, J. L. & Robbins, T. W. 1989 Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. *Neuropsychologia* **27**, 1329–1344.

Duncan, J., Burgess, P, & Emslie, H. 1995 Fluid intelligence after frontal lobe lesions. *Neuropsychologia* **33**, 261–268.

Goel, V. & Grafman, J. 1995 Are the frontal lobes implicated in 'planning' functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia* **33**, 623–642.

Goldman-Rakic, P. S. 1987 Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In *Handbook of physiology. The nervous system* (ed. F. Plum), pp. 373–417. Bethesda, MD: American Physiological Society.

Joyce, E. M. & Robbins, T. W. 1991 Frontal lobe function in Korsakoff and non Korsakoff alcoholics: planning and spatial working memory. *Neuropsychologia* **29**, 709–723.

Lawrence, A. D., Sahakian, B. J., Hodges, J. R., Rosser, A. E., Lange, K. W. & Robbins, T. W. 1996 Executive and mnemonic functions in early Huntington's disease. *Brain*. (In the press.)

Milner, B. 1963 Effects of different lesions on card sorting. *Archives of Neurology* **9**, 100–110.

Morris, R. G., Downes, J. J., Sahakian, B. J., Evenden, J., Heald, A. & Robbins, T. W. 1988 Planning and spatial working memory in Parkinson's disease. *J. Neurol. Neurosurg. Psychiat.* **51**, 757–766.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E. & Robbins, T. W. 1990 Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia* **28**, 1021–1034.

Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J. & Robbins, T. W. 1991 Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excision, temporal lobe excision or amygdalo-hippocampectomy in man. *Neuropsychologia* **29**, 993–1006.

Owen, A. M., James, M., Leigh, P. N. *et al.* 1992. Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain* **115**, 1727–1751.

Owen, A. M., Roberts, A. C., Hodges, J. R., Summers, B. A., Polkey, C. E. & Robbins, T. W. 1993 Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain* **116**, 1159–1179.

Owen, A. M., Sahakian, B. J., Semple, J., Polkey, C. E. & Robbins, T. W. 1995*a* Visuospatial short term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdala hippocampectomy in man. *Neuropsychologia* **33**, 1–24.

Owen, A. M., Sahakian, B. J., Hodges, J. R., Summers, B. A., Polkey, C. E. & Robbins, T. W. 1995*b* Dopamine-dependent fronto-striatal planning deficits in early Parkinson's disease. *Neuropsychology* **9**, 126–140.

Owen, A. M., Morris, R. G., Sahakian, B. J., Polkey, C. E. & Robbins, T. W. 1996*a* Double dissociations of memory and executive functions in a self-ordered working memory tasks following frontal lobe excision, temporal lobe excisions or amygdala-hippocampectomy in man. *Brain*. (In the press.)

Owen, A. M., Doyon, J., Petrides, M. & Evans, A. C. 1996*b* Planning and spatial working memory examined with positron emission tomography. *European J. Neurosci.* **8**, 353–364.

Owen, A. M., Evans, A. C. & Petrides, M. P. 1996*c* Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cerebral Cortex* **6**, 31–38.

Pandya, D. P. & Yeterian, E. H. 1995 Morphological correlations of human and monkey frontal lobe. In *Neurobiology of human decision-making* (ed. A. E. Damasio, H. Damasio & Y. Christen), pp. 13–46. New York: Springer.

Passingham, R. E. 1985 Memory of monkeys (Macacca mulatta) with lesions in prefrontal cortex. *Behavioral Neuroscience* **9**, 3–21.

Petrides, M. P. 1996 Lateral frontal cortical contribution to memory. *Seminars in the Neurosciences* **8**, 57–63.

Petrides, M. P. & Milner, B. 1982 Deficits on subject-ordered tasks after frontal- and temporal lobe lesions in man. *Neuropsychologia* **20**, 249–262.

Robbins, T. W., James, M., Owen, A., Sahakian, B. J., McInnes, L. & Rabbitt, P. M. 1994 Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia* **5**, 266–281.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L. & Rabbitt, P. 1996 A neural systems approach to the cognitive psychology of ageing: studies with CANTAB on large samples of the normal elderly population. In *Methodology of frontal and executive function* (ed. P. Rabbitt) Hove, East Sussex: L. Erlbaum Associates, (In the press.)

Roberts, A. C., Everitt, B. J. & Robbins, T. W. 1988 Extra- and intra-dimensional shifts in man and marmoset. *Q. J. Exp. Psychol.* **40B**, 321–342.

Roberts, A. C., Robbins, T. W., Everitt, B. J. & Muir, J. L. 1992 A specific form of cognitive rigidity following excitotoxic lesions of the basal forebrain in monkeys. *Neuroscience* **47**, 251–264.

Roberts, A. C., De Salvia, M. A., Wilkinson, L. S. *et al.* 6-Hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analogue of the Wisconsin Card Sorting test: Possible interactions with subcortical dopamine. *J. Neurosci.* **14**, 2531–2544.

Sahakian, B. J., Downes, J. J., Roberts, A. C., Philpot, M., Levy, R. & Robbins, T. W. 1990 Preserved attentional function and impaired mnemonic function in dementia of the Alzheimer type. *Neuropsychologia* **28**, 1197–1213.

Sahgal, A., Lloyd, S., Wray, C. J. *et al.* 1992 Does visuospatial memory in Alzheimer's disease depend on the severity of the disorder? *Int. J. Geriatric Psychiatry* **7**, 427–436.

Shallice, T. 1982 Specific impairments of planning. *Phil. Trans. R. Soc. Lond.* B **298**, 199–209.

Shallice, T. & Burgess P. 1993 Supervisory control of action and thought selection. In *Attention: selection, awareness and control* (ed. A. Baddeley & L. Weiskrantz), pp. 171–187. Oxford: Clarendon Press.

Discussion

D. Weinberger (*Clinical Brain Disorders Branch, National Institute of Mental Health, Neuroscience Center at St. Elizabeths, 2700 Martin Luther King Jr Avenue, SE, Washington D.C. 20032*). I agree that your attentional set-shifting paradigm, like the WCST involves response inhibition but clearly it also has a strong working memory component. Therefore how can you be so sure that the deficits on your set-shifting paradigm following frontal lobe damage relate solely to response inhibition?

T. W. Robbins. Our attentional set-shifting paradigm clearly has several cognitive components. Common to each of the visual discrimination tasks is the requirement to respond to reinforcing feedback and to hold 'on-line' the results of preceding trials (assuming that this is the key aspect of 'working memory' to which the question refers). However, the visual discrimination tasks differ in several other aspects including their differing requirements for response inhibition. For example, in the reversal tasks the subject simply has to inhibit responding to the previous exemplar, and shift responding to the alternative one, both exemplars having the same perceptual dimensions. By contrast, in the extra-dimensional shift task, the subject has to inhibit responding to an attentional dimension such as shape, and shift attention ton to an alternative dimension, in our case, superimposed lines. As we have indicated, these different types of shifting for reversal and extra-dimensional set-shifting appear to depend on different regions of prefrontal cortex in the marmoset. The reason we think it unlikely that these regions also mediate the memorial aspects of the task is because similar discriminations without this response inhibition component, but with similar working memory demands, for example, in the case of the intra-dimensional shift, are not impaired by prefrontal lesions. Furthermore, in our studies with human patients, we have designed different versions of the extra-dimensional shift task which vary in their requirements for inhibition, but which have similar associative and working memory components. In this study only that version of the task requiring inhibition of the previously reinforced attentional dimension exhibited deficits following frontal lobe excisions. Therefore, the most parsimonious account of these data is that the frontal lesions impair response inhibitory processes over and above any possible additional contribution of short-term memory and associative learning.

J. G. Taylor (*Department of Mathematics, King's College, Strand, London, WC2R 2LS, U.K.*). To what extent do the inhibitory mechanisms that you describe operating at the level of the prefrontal cortex also depend upon parallel circuitry in the striatum, given as you say, that patients with Huntington's disease also apparently exhibit perseverative responding?

T. W. Robbins. It is important to realize that the prefrontal cortex has important neuroanatomical connections with the basal ganglia which are indeed impaired in Huntington's disease. According to Alexander *et al.*(*Ann. Rev. Neurosci* **9**, 357–381, 1986), there are a number of parallel corticostriatial loops whose functions are segregated, to some extent. Thus, for example, orbitofrontal and dorsolateral regions of prefrontal cortex project to different regions of the striatum (i.e. to ventromedial and dorsal caudate nucleus, respectively). In recent studies we have found that patients early in the course of Huntington's disease have specific deficits in extra-dimensional shifting, whereas more severely affected patients also have very large deficits in reversal learning which are characterized by perseverative responding to the originally reinforced exemplar. The staging of these two forms of shifting deficit is consistent with the proposed neuropathological progression of Huntington's disease from dorsal to relatively ventral structures in the caudate nucleus (Hedreen & Folstein, *J. Neuropath. exp. Neurol.* **54**, 105–120, 1995). Our data are thus consistent with the hypothesis that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the two forms of shifting may depend upon different fronto-striatal systems.

R. G. Morris (*Institute of Psychiatry, University of London, de Crespigny Park, Denmark Hill, London, SE5 8AF, U.K.*). In the spatial working memory studies, if the strategy measure was covaried out in the analysis, was there still a spatial working memory deficit in the frontal lobe group.

T. W. Robbins. If the strategy measure is used as a covariate in an analysis of covariance of the memory measure for the self-ordered working memory task, there is a significant reduction in the size of the difference for the memory score between the frontal patients and control subjects, but the group difference remains highly significant. This is consistent with the notion that the task has several cognitive components, more than one of which may be affected by damage to the prefrontal cortex. However, a note of caution is that the index of strategy is, of course, only approximate, and so it is possible that it under-estimates the true contribution of the strategic deficit to the impairment in self-ordered memory in frontal patients.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES