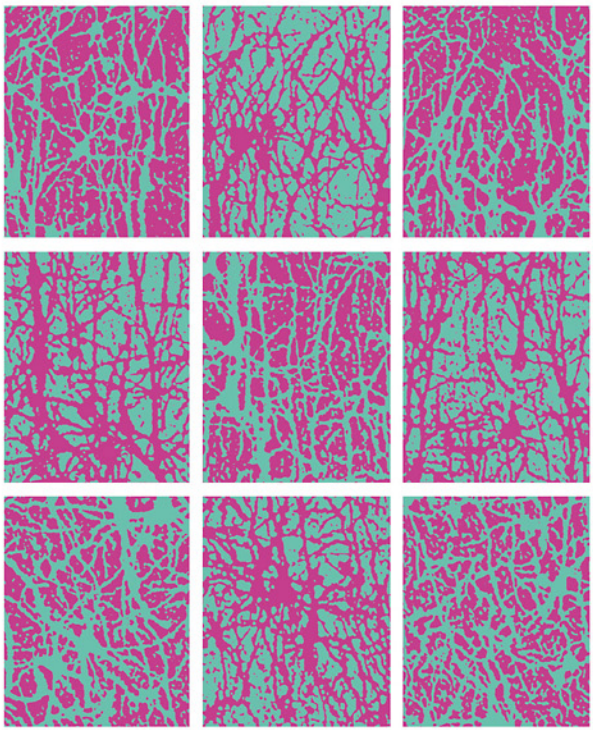
Methodology of Frontal and Executive Function



Edited by Patrick Rabbitt

Methodology of Frontal and Executive Function

edited by

Patrick Rabbitt
University of Manchester, UK
and
University of Western Australia, Perth, WA

This edition published in the Taylor & Francis e-Library, 2005.

Psychology Press is part of the Taylor & Francis Group

www.psypress.co.uk

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

ISBN 0-203-34418-9 Master e-book ISBN

ISBN 0-86377-485-7 (pbk)

Contents

List of Contributors

Nick Alderman, The Kemsley Unit, St. Andrews Hospital, Northampton NN1 5DG, U.K.

Alan Baddeley, Department of Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, U.K.

S.E.Black, Department of Medicine, Research Program in Aging and Cognitive Neurology, Sunnybrook Health Science Centre, 2075 Bayview Avenue, Toronto, Canada.

Michael J.Bronskill, Medical Imaging Research, Sunnybrook Health Centre, 2075 Bayview Avenue, Toronto, Canada.

Brian H.Buck, Cognitive Neurology, Sunnybrook Health Science Centre, 2075 Bayview Avenue, Toronto, Canada.

Paul W.Burgess, Department of Psychology, University College London, Gower Street, London WC1E 6BT, U.K.

Irene Daum, Clinical Psychology Unit, Faculty of Psychology, Ruhr-University of Bochum, 44780 Bochum, Germany.

Sergio Della Sala, Department of Psychology, University of Aberdeen, Old Aberdeen AB24 2UB, Scotland, U.K.

Hazel Emslie, MRC-APU Rehabilitation Research Group, Box 58, Addenbrookes Hospital, Hills Road, Cambridge CB2 2EF, U.K.

Jonathan J.Evans, Oliver Zangwill Centre, Princess of Wales Hospital, Ely, Cambs, U.K.

Jonathan K.Foster, Department of Psychology, University of Manchester, Oxford Road, Manchester, U.K.

Colin Gray, Department of Psychology, University of Aberdeen, Old Aberdeen AB24 2UB, Scotland, U.K.

M.James, Department of Psychology, National Hospital, Queen Square, London, U.K.

Christine Lowe, Age and Cognitive Performance Research Centre, University of Manchester, Coupland I, Oxford Road, Manchester M13 9PL, U.K.

Tom Manly, MRC-APU Rehabilitation Research Group, Box 58, Addenbrookes Hospital, Hills Road, Cambridge CB2 2EF, U.K.

Andrew R.Mayes, Department of Clinical Neurology, N Floor, Royal Hallamshire Hospital, Glossop Road, Sheffield S10 2JF, U.K.

L.McInnes, North-East Age Research Panel, Department of Psychology, Newcastle University, Newcastle-upon-Tyne, U.K.

A.M.Owen, Department of Psychiatry, University of Cambridge, School of Clinical Medicine, Cambridge CB2 2QQ, U.K.

Costanza Papagno, Third Neurology Department, University of Milan, S.Paolo Hospital, Milan, Italy.

Alan J.Parkin, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, U.K.

Louise H.Phillips, Psychology Department, Aberdeen University, Old Aberdeen AB24 2UB, Scotland, U.K.

Patrick Rabbitt, Age and Cognitive Performance Research Centre, The University of Manchester, Coupland I, Oxford Road, Manchester M13 9PL, U.K.

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

Ian H.Robertson, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, U.K.

B.J.Sahakian, Department of Psychiatry, School of Clinical Medicine, University of Cambridge, Cambridge CB2 2QQ, U.K.

Hans Spinnler, Third Neurology Department, University of Milan, S.Paolo Hospital, Milan, Italy.

Barbara A.Wilson, MRC-APU Rehabilitation Research Group, Box 58, Addenbrookes Hospital, Hills Road, Cambridge CB2 2EF, U.K.

CHAPTER ONE
Introduction: Methodologies and Models in the Study of Executive Function

Patrick Rabbitt *Age and Cognitive Performance Research Centre, University of Manchester, U.K.*

A common theme of all contributors to this book is that, because tests of executive function have low test-retest reliability and uncertain validity, it has been hard to draw clear empirical distinctions between "executive" and "non-executive" tasks. It has also proved difficult to relate particular behaviours which have become regarded as being paradigmatic exemplars of "executive" function to specific neuroanatomical areas or neurophysiological systems. We argue that some of these difficulties occur because executive behaviour has been discussed, simultaneously, at more than one level of description. Accounts of purposeful and conscious behaviour, in the language of everyday subjective experience, map uneasily onto clinical taxonomies of the modes of dysexecutive function observed in brain-damaged patients. Performance indices empirically measured in laboratory tasks are often treated as being directly equivalent to the hypothetical system performance characteristics specified in functional models of working memory and selective attention. As a result, it has passed unrecognised that hypothetical components of "executive behaviour," such as "inhibition," "planning," "monitoring" and "control" which are, in fact, simply descriptions of task demands, may have very poor construct validity because although these demands appear logically different they can be met by identical production system architectures. Confusions have also arisen because both task performance indices and system performance characteristics have been treated as equivalent to statistical constructs, derived to explain individual differences in performance on intelligence tests, such as Spearman's (1927) index *gf* (see review by Rabbitt, 1996).

Individually and collectively, the contributions to this book show why it is unsurprising that all empirical psychologists find it very difficult and some philosophers (Fodor, 1983) think it impossible to give a satisfactory functional account of the central executive system (CE). Classical descriptions of executive impairment associated with prefrontal lesions have immediacy and richness, and are strikingly similar to definitions of "willed" "purposeful" or "voluntary"

behaviour that have preoccupied philosophers and theologians for more than two millennia. For example, contemporary catalogues of the functions of the hypothetical "central executive" are strikingly similar to the formal criteria for commission of mortal sin given by Roman Catholic theologians: Beasts are soulless automata, and so always blameless. Humans also do not transgress while performing even very complex vile acts, provided that they do so involuntarily and automatically without conscious awareness of themselves as perpetrating agents. The minimal functional processes involved in the commission of a mortal sin are awareness of the self as the intending perpetrator of the act; recognition of the unpleasant implications of the act for others by possession of a theory of mind; recognition of its moral repulsiveness by possession of a theory of the mind of God; an ability simultaneously to represent alternative acts and their possible outcomes in working memory in order efficiently to choose between them; conscious formulation of a well-articulated plan to perform the act successfully; self-initiation and execution of sequences of appropriate actions to consummate this plan during which recognition of personal culpability is maintained by continuous monitoring; recognition of attainment of the vile goal state and an intention to use what has been learned in its pursuit to perform it again if opportunity occurs.

Clearly only the central executive (CE) can sin. Do we yet have functional models that improve on this already comprehensive theological description?

Both theologians and neuropsychologists seem to be describing the same ghost, and it is natural that a class of investigators whom Burgess (this book) labels "ultra-cognitive neuropsychologists" should try to deduce its properties from the results of selective damage to the machine that it haunts. An impressive, but unlikely, discovery would be that whenever a particular, functionally circumscribed, brain system is damaged the patient becomes capable only of brutish automaticity. An even more interesting finding would be that different lesions disable one, but not others, of a set of logically distinct and tightly specifiable functions such as "planning," "sequence initiation," "monitoring" or "inhibition of alternative responses." This would reveal the "double disassociations" of effects prized by "ultra-cognitivists." Unfortunately all the chapters in this collection, and a voluminous background literature excellently reviewed by Fuster (1989), Reitan & Wolfson (1994) and others, point out that attempts to relate any of the capabilities that are agreed to be most characteristic of "executive" behaviour to particular brain systems or even to diffuse changes in volume of brain tissue have been unsuccessful (see especially Foster et al. this book). Failures to offer convincing neurophysiological descriptions of the bases of most complex cognitive skills may, as Burgess (this book) clearly anticipates, sap our faith in the methodology of "ultra-cognitive neuropsychology." It has been suggested that some of these difficulties may be avoided if the problem is restructured as a search for empirical distinctions

between "executive" and "dysexecutive" behaviours and the question of where, or what, their neurophysiological bases may be is deliberately, and indefinitely, postponed (e.g. Baddeley & Wilson, 1988). While this ploy frees us to carry out behavioural experiments without worrying about how their results bear on what we know about the brain, it also deprives us of an important source of evidence against which to check the plausibility of our models. In the difficult task of understanding executive function we need all the help that neuropsychology can offer.

Nevertheless, attempts to distinguish behaviourally between "executive" (E) and "non-executive" (NE) functions at least force us to move beyond the richly allusive language of everyday subjective experience and try to develop a functional taxonomy distinguishing performances, skills, or behaviours that are characteristic of executive function from those that are not. Collectively, our contributors generate an exhaustive list:

First, executive control is necessary to deal with novel tasks that require us to formulate a goal, to plan, and to choose between alternative sequences of behaviour to reach this goal, to compare these plans in respect of their relative probabilities of success and their relative efficiency in attaining the chosen goal, to initiate the plan selected and to carry it through, amending it as necessary, until it is successful or until impending failure is recognised. It is crucial to note that this prospectus does not merely boil down to a statement that executively controlled behaviour is much more complex than NE behaviour. Even choices between very complex behaviour sequences (aka plans) can be carried out "automatically" under NE control. Examples of this might be skilled drivers negotiating complex traffic while carrying on non-trivial conversations, or chess grand masters rapidly selecting among extremely difficult lines of play by recognising generically familiar positions during "lightning chess" tournaments. The key distinction seems to be between situations in which a person must, for the first time, recognise, evaluate, and choose among a variety of *alternative* options and those in which a *single* effective behaviour sequence, which has been previously identified, and instantiated by practice, is run off without the need to propose and evaluate alternatives.

In contrast to E behaviour, NE behaviour tends to be initiated by, and to continue automatically in response to, changes in environmental input. A possible way to envisage this is that even very complex NE behaviours are externally driven and controlled by sequences of environmental events or by previously learned "plans" or "programs" held in long-term memory. In contrast, E behaviours can be initiated and controlled independently of environmental input, and can retain the adaptive flexibility to rescue plans even when the environment does not behave as anticipated and no guidance is available from previous experience.

It follows that, should the rules suddenly change, as in the Wisconsin Card-sorting Task or its variants, "executive" function will be necessary to appreciate from feedback that a change has occurred and to formulate and test new plans or rules until successful. Some authors stress that part of this process is the "inhibition" or "suppression" of previously used and so habitual rules in order to project, select, and use alternatives.

Second, a study reported by Burgess (this book) shows that deliberate retrieval of structured information from memory in order to answer a question or to elaborate a partial description may involve the conscious, and reportable, formation and prosecution of well-articulated memory "search plans." Burgess points out that CE control is not only necessary to manage transactions with the external world but also with the "internal information environment" of long-term memory. Thus CE control extends beyond the *current* internal or external information environment to restructure interpretation of the past as well as to attempt active control of the future. Burgess's excellent insight is timely because it allows us to distinguish between NE retrieval of information from memory in response to learned associations and environmental cues and active, "strategic" or "planned" memory search carried out under voluntary control. Closely related suggestions by Phillips (this book) are that individuals with inefficient executive function apparently generate categories of words by simple associations between successive items, or by empirically inscrutable random processes, while the better performance of individuals with intact executive function may be partly attributable to their discovery and use of memory search strategies. These suggestions of a strong link between CE control and efficiency of organised recall of material from memory are particularly interesting in view of Mayes' and Daum's (this book) neat demonstrations that lesions in prefrontal areas that are often associated with losses of executive function also impair learning, recall, and recognition. They raise the further question whether prefrontal lesions may impair *strategic control* of recall and recognition to a greater extent than do the lesions of the mediotemporal lobes and of the hippocampus and limbic system that are more usually associated with the "classic" amnesic syndrome. In this case lesions might disassociate strategic control of memory from memory efficiency.

Third, E functions are necessary to initiate new sequences of behaviour and also to interrupt other ongoing sequences of responses in order to do so. A logically related property is that they can suppress, or inhibit and replace, automatic and habitual responses with task-appropriate responses. E functions can also check involuntary perseveration by appropriately "switching" attention to new sources of information. In this sense "inhibition" of habitual responses in order to make other responses, or "switching" between one sequence of responses or one aspect of the environment and another, appear to be logically, and operationally, very similar concepts (see Lowe & Rabbit this volume, and Baddeley, Maylor, &

Godden, in press). Whether by strategically switching attention from one source to another over time, or by simultaneously and strategically allocating attentional resources to more than one input, E processes control the allocation of attention, particularly in complex tasks in which a variety of different demands must be met simultaneously.

Fourth, E functions are necessary to prevent responses that are inappropriate in context. This rather poorly articulated aspect of dysexecutive function must be logically distinguished from the suppression of habitual or unnecessary responses. It seems first to have been documented by clinical observations of inappropriate social behaviour following frontal lesions. Such lapses of social judgment may perhaps have logical analogues in empirical findings, such as the increases in numbers of bizarre responses that patients with frontal lesions make to questions in the Shallice Cognitive Estimates Test (Shallice & Evans, 1978) and in reports, such as that by Parkin (this book) that older, and so potentially more "executively challenged," individuals make markedly more false-positives in recognition memory experiments than do younger adults. It is hard to suggest any single, comprehensive description of the cause of these and similar lapses but "responding without taking all of the available information into consideration," "responding without considering alternatives," "responding on the basis of insufficient information" or "failing to monitor for plausibility" may all capture something that they have in common. It is also difficult to disassociate these ideas from findings such as Parkin's of an increase in false-positives that may, generically, also be described as "loss of strategic control of memory." Note also that these problems of response production can also be logically, if not functionally, distinguished from failures to monitor responses in order to detect whether they are "correct" or "wrong" in the sense that they have or have not succeeded in achieving a particular intention. They seem, rather, to be failures to judge when an *intention,* as well as the response in which it results, is implausible or unacceptable in a particular context.

Fifth, dual-task performance has become a test bed for exploration of functions of planning and control because the simultaneous performance of two tasks may require strategic allocation of attention and synchronisation of responses in order to service both. This may be achieved by rapidly switching from one task to the other and back again as their relative demands fluctuate over time, or by other forms of resource management such as are envisaged in models of "M space" (Case, 1992; Case & Sandieson, 1992) or quantified in terms of the index mu (see Baddeley, della Salla et al. this book). Simultaneous performance of two E tasks is seen as being especially demanding because, in effect, it may involve coping with all of the demands intrinsic to task 1 and task 2 and also, in addition, with the demand of strategically controlling allocation of attention between tasks. As many elegant experiments by Baddeley (1986) have shown, mutual interference between simultaneous tasks is particularly severe if they

make simultaneous demands on the same sensory modality or representation system. In contrast, when two tasks are both under NE control, even if both are complex they may be simultaneously performed with less mutual interference and with less need for strategic scheduling of their conflicting demands. However, as recent work by Bourke, Duncan, and Nimmo-Smith (1993) suggests, there may always be some loss of efficiency in dual-task performance.

Sixth, while carrying out plans to cope with novel demands, E functions are necessary to monitor performance in order to detect and correct errors, to alter plans when it becomes clear that they are becoming unlikely to succeed, or to recognise opportunities for new and more desirable goals and to formulate, select among, initiate, and execute new plans to attain them. Even complex sequences of NE behaviours do not seem to be monitored at this level. This distinction requires elaboration because it has long been known that error correction in simple tasks such as serial choice reaction time is very fast and "automatic" (Rabbitt, 1968), so that continuous monitoring for discrepancies between intentions and acts does occur, but at speeds (e.g. < 40ms) which are much too fast to be attributable to "conscious", or "executive" intervention. In contrast, E monitoring is envisaged as a higher level activity that not only detects when actions have not been completed as planned, but can also predict when failure is imminent, or an even more attractive goal may be missed, unless a plan, which is currently being carried out precisely as intended, can be rapidly changed.

Seventh, Manly and Robertson (this book) suggest that another characteristic of E control is that it can enable attention to be sustained continuously over long periods. Among other benefits, this allows prediction of outcomes of long, complex sequences of events. This property seems to have first been suggested by Wilkins and Shallice (1987). Perhaps an example of this property of E behaviour is Duncan's (1995) clever demonstration that dysexecutive individuals show "goal neglect" in a novel task in which they are, at random intervals, signalled either to "switch" from one pattern of responding to another or to "stick" with the response pattern in which they are currently engaged. Another example might be the "lift" or "elevator" monitoring task, designed by Robertson and his associates, in which participants watch a changing display which signals a series of upward and downward movements of a hypothetical conveyor. By monitoring these changes patients have to identify the floor that the conveyor has currently reached. We have already discussed how E control is necessary to formulate plans. Manly and Robertson's observations suggest that we should add to this the ability to hold intended sequences of actions in mind over long periods of time so as to carry them out as soon as appropriate circumstances occur. Another general description of such activities might be "remembering to do necessary things at the right times" which is currently investigated as "prospective memory" in experiments such as those by Einstein and McDaniel

(1990) or Maylor (1990; see also a useful collection of papers by Brandimonte, Einstein & McDaniel, 1996).

Finally, one last taxonomic distinction is that E behaviours are, while NE are not, accessible to consciousness. In view of the difficulty of empirically investigating conscious states it is not surprising that this distinction has been much less stressed than any of the others we have discussed.

DISTINGUISHING BETWEEN EXECUTIVE AND NON-EXECUTIVE FUNCTIONS

When these criteria are listed together it becomes clear that they all describe the *outcomes,* but not the *functional aetiology* of the cognitive activities that we hope to understand. They are consensus descriptions, similar to descriptions of what constitutes "intelligence" elicited from laymen and "experts" that tell us *the things that people do* that allow us to label them as more or less "intelligent" but do not attempt to define the processes responsible for "intelligent behaviour" (e.g. Sternberg & Powell, 1982; Sternberg & Detterman, 1986). Such descriptions are intuitively appealing, and even illuminate our subjective experience, but are untrustworthy guides as to how to investigate function. For example, when an unpractised rather than a more habitual response is chosen we may describe this as "inhibition", when a complex sequence of successive acts is initiated we call this "planning" and when an impending failure of intention is noticed and corrected we may describe this as successful "monitoring" of the progress of an ongoing plan. However, increasing sophistication in modelling processes in terms of production systems or of connectionist networks should make us wary of assuming that choice between alternative responses is best described in terms of "inhibition" or "suppression" of one and selection of the other, or that the detection of impending, as well as completed, errors demands the "attentional resour ces" or a "higher order system" that, as it were, re-processes or reexamines the current content of other, ongoing, "lower order" systems. The practical question is whether terms such as "planning", "inhibition" or "monitoring" have any "construct validity". It is clear that each of these hypothetical processes can be ostensively defined in terms of a *different set of exemplary situations or tasks,* but it is not at all clear that each of them involves quite *distinct functional processes,* or that failures of processes underlying some of these tasks can be empirically shown to be independent of failures of different processes underlying others.

Life would be simpler if there were generally agreed paradigmatic "executive" tasks available for rigorous empirical analysis. Unfortunately, as our contributors amply document, no single task among those listed has gone unchallenged as a prototypic exemplar of "executive function". This is partly because, as most of our other contributors point out, individuals suffering from "dysexecutive syndromes" can nevertheless successfully carry out "any number" of putatively

"executive tasks" (Burgess, this book). Further, as Burgess complains, and Duncan (1995) has also pointed out, tasks cease to be effective tests of executive function as soon as they are performed more than once. Thus tasks can be categorised in terms of what appears to us to be their "common logical content" but not, necessarily, in terms of their "common functional support systems" in respect of the effects of particular loci of brain damage. At this point it is interesting to note how heavily previous investigators have depended on the converse assumption, that cognitive skills that are, supposedly, affected by different loci of brain damage must necessarily be supported by different *kinds* of functional processes. This assumption must be suspect because, in terms of abstract models, it is entirely plausible that functions that are impaired by lesions at different sites in fact depend on identical system architectures. One of the many challenges of interpreting the effects of frontal lesions is that they do not allow this *post hoc,* or *"post trauma,"* taxonomy of behavioural processes in terms of lesion sites, on which many colleagues, including those whom Burgess (this book) terms "ultra cognitive neuropsychologists," have depended.

This distinction between the logical status of the "task performance indices" that we obtain from diagnostic tests or laboratory experiments, the "system performance characteristics" which are specified in functional models of cognitive processes and the functional neurophysiology and neuroanatomy of the central nervous system (CNS) explains some disappointing failures of statistical analysis. It seems easy to assume that terms such as "planning," "inhibition," and "concept shifting" are not merely descriptions of different task demands, whose effects can be quantified in terms of the indices we measure in laboratory experiments, but are also labels marking valid qualitative distinctions between the functional processes by means of which these demands are met. If we do assume this, it seems natural to expect that, when many different "executive tasks" are given to the same large groups of normal healthy adults, scores on subsets of tasks that demand related activities such as "planning," or "inhibition," or "monitoring" should be strongly associated. In contrast, if we do not assume that a taxonomy of task demands is also, *necessarily,* a taxonomy of underlying processes, we will have no such expectation.

A number of studies have put this to the test, with rather mixed results (e.g. Spinnler et al., cited by Baddeley, Della Salla et al., this book; Rabbitt, 1993; Rabbitt & Yang, 1996; and Robbins et al., 1994, and this book). In these studies, large numbers of individuals have been given the same sets of "executive" and other tasks and principal components analyses have been made to test whether or not tasks that have been assumed to depend on identical, or closely related, functions do in fact load on the same factors. A further option has been to enter individual difference variables, such as participants' ages or intelligence test scores, into analyses to discover whether these, also, are represented more strongly in some task-specific factors than in others. Robbins et al. (1994; and

this book) report factor separation between scores from tests of executive function that are assumed, on logical grounds, to be mediated by different processes. Baddeley, della Salla et al. (this book) cite a finding by Spinnler and associates that separations between subsets of frontal tasks were unclear when data from intact normals were analysed. However, analyses of batteries given to patients with frontal lesions found better evidence for separation of variance between tasks in conceptually related categories. This raises the possibility that, while there are no marked clusterings of associations between scores obtained from people with intact brains, in patients with damage to the frontal cortex we may find meaningful clusterings of deficits suggesting some localisation of function,

In our laboratory large batteries of different tests of executive function have been given to very large groups (>600) of normal elderly aged from 60 through 86 years. While variances in scores on tests of information-processing speed and in scores on memory tests are clearly best represented by separate factors, we find that subgroups of supposedly similar "executive tasks" do not cluster in any obviously meaningful way. Correlation matrices are informative since associations between scores on tests of allegedly very similar executive functions are significant but modest (i.e. r=.3 to .45), but in almost all cases reduce to non-significance after variances associated with scores on intelligence tests, such as the Cattell and Cattell (1960) culture fair test, have been partialled out. In other words, as far as we can yet tell, in a population of normal, healthy, active community residents aged from 60 to 85 years, associations between scores on "executive" tests can be almost entirely accounted for in terms of their common loadings with intelligence test scores; as Duncan (1995) might put it, with their common loadings on Spearman's (1927) single factor of general intelligence, *g*.

One reason for such failures of differentiation between tasks may be that tasks that have been classified as separate in terms of their demands are, in fact, carried out by similar functional processes. A different possibility is that different kinds of "executive task demands" are met by distinct functional processes, but all these processes make demands on a common pool of cognitive resources. In this case correlations between task scores will reflect common levels of demands on resources rather than commonality of underlying functional processes. This may be particularly the case in studies that examine the effects of diffuse brain changes brought about by "normal" ageing, by closed-head injuries or by Alzheimer's disease in which diffuse brain damage may be supposed to result in a global reduction of resources rather than in circumscribed losses of particular functions. It seems still to be moot whether observed failures on "executive" tasks result from damage to specific, functionally localised subsystems or whether all that the executive tasks have in common is that they are "harder," "more demanding," require a higher information-processing/working memory/attentional load or are simply more "stressful" than the "non-

executive" tasks with which they are compared, and so are earlier compromised by any reduction in the level of "global" resources.

DIFFICULTIES IN EMPIRICAL STUDIES OF "EXECUTIVE" FUNCTION

Poor Test/retest Reliability

All contributors insist that only novel tasks can pick up deficits, and because they can only be novel once, the test/retest reliability of "executive" tasks is disappointingly low. This unreliability is a source of measurement error that is likely to weaken correlations between test scores and to contribute to failures to find clear patterns of associations between levels of performance on tests with logically similar demands.

The Uncertain Validity of "Executive" Tests

All contributors also point out that "executive" tests have questionable validity in the sense that they are rather unreliable diagnostic indices for "frontal" damage, or dysexecutive behaviour. The possible reasons for this seem central to theoretical discussions as to what executive competence is, and how it can be assessed.

(1) There is a problem of measurement error. Since executive tasks typically have poor test/retest reliability, in the sense that they only modestly predict performance on themselves, the resulting measurement error must also blur correlations between scores on executive tasks and on any other performances against which we try to validate them.

(2) A potential problem is that the performances that we hope to examine may be intensely task-specific. For example, Kimberg and Farrah (1993) suggest that activities such as planning and executing motor sequences, managing category shifts in Wisconsin card-sorting performance, inhibiting incorrect responses in the Stroop test, or retrieving contextual information from memory may each be each managed by a separate working memory system which is dedicated to a single, situation-specific function. In this case, any one of these working memory systems might, in theory, be damaged without effect on the others, resulting in precisely the lack of associations between performance deficits and lesions to which contributors draw attention.

The possibility that executive subsystems are extremely domain-specific also suggests a reason for the insightful complaint by Wilson and Evans (this book) that executive tasks offer little help with the problems of rehabilitation of patients with frontal damage, and do not answer questions such as "how do patients' problems affect their function in everyday life?", "which difficulties are

likely to cause the most distress"?, and "can the patient return to a previous lifestyle?". We cannot hope that domain-specific performance on particular diagnostic tasks will predict competence across a broad range of complex everyday skills.

(3) When particular "executive" skills are highly practised they become remarkably robust to injuries that severely disable other functions of the cognitive system. They also become extremely domain-specific (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In addition to the other difficulties discussed above this could account for the comment by Burgess that particular patients, who have marked frontal damage, can nevertheless successfully carry out any number of executive tests, but may yet fail on others.

(4) The neurophysiological and neuroanatomical processes underlying executive task performance may be diffusely, rather than focally, represented in the brain, so that a taxonomy of task demands cannot be mapped on to a corresponding taxonomy of lesion sites.

(5) Some investigators have hoped to validate "executive" tests in terms of their ability to detect particular losses of performance by individuals suffering from diffuse brain damage, with probably particular involvement to frontal cortex, caused by Alzheimer's disease, by closed-head injuries, or by the progress of "normal" ageing (Lowe & Rabbitt, this book; Robbins et al., 1994, and this book). This is reasonable if we suppose that the dysexecutive syndrome is caused by loss of a "global" resource which may, perhaps, also be behaviourally quantified by reduction in *g* (Duncan, 1995) or by slowing of information-processing rate (Salthouse, 1985) or by loss of working memory capacity (Salthouse, 1991). However, some methodological difficulties need to be borne in mind. Alzheimer's patients are known not only to suffer changes to prefrontal cortical tissue, but also to experience sensory losses and often also suffer from a variety of localised lesions that interfere with memory, language, and other specific cognitive skills. Closed-head injuries often present similar problems of combinations of diffuse with focal damage. It must always be remembered that "normal ageing" is a process that is only loosely timebound, in the sense that the rates at which biological changes in the brain occur, and so the rates at which cognitive competence is altered, differ widely between individuals. Thus when we compare different age groups, without the help of neurophysiological data, such as may be provided by brain scans, all that we can be sure of is that more individuals in the older than in the younger cohorts have experienced the brain changes in which we are interested. We must not forget that we are comparing populations in which the incidence of a particular CNS condition is more or less frequent, rather than comparing populations in which all individuals suffer from an identical condition which, in all of them, is increasingly more strongly expressed as they age. Incidence of particular age-related brain changes may remain relatively low until a step-function rise in their

frequency occurs within a particular, critical age range. The paper by Robbins et al. (this book) illustrates this point, showing abrupt and marked declines in scores for individuals over 75, following a relative plateau without change for all younger groups.

(6) The terms that we use in order to classify the common demands made by subsets of diagnostic tests may have poor construct validity. An example is the term "inhibition," the ability to attend selectively to goal-related signals and to ignore, and withhold ("inhibit"), habitual, or over-practised responses to other stimuli, which is classically demonstrated as particular difficulty with the "interference" condition of the Stroop test (see review by McLeod, 1991). The idea that "inhibition" is a particular, empirically delimitable functional property that is gradually lost in old age has supported an extensive literature suggesting that older individuals are poorer at Stroop-type tasks in which they are required to inhibit some practised responses in order to make others (e.g. see Hasher, Stoltzfus, Zacks, & Rypma, 1991; West, 1996).

It is certainly the case that the difference between the times taken to complete "conflict" and "no conflict" conditions of Stroop-like tasks does increase with age, and Rabbitt (1996) has shown that this does not occur simply because decision times on conflict conditions are always longer than in other conditions, and all decision times are scaled up by the same simple multiplier as meta-analyses by Cerella (1985) would suggest. However, in our laboratory, we have been unable to find any commonality of individual differences in "inhibition" between each of a wide variety of logically identical but superficially dissimilar Stroop-like tasks. That is, we can find no evidence that the ability to inhibit responses across a range of different tasks is consistently greater in some individuals than in others. We have also failed to demonstrate correlations between individual differences in the amount of slowing in the conflict conditions of Stroop tasks and in the critical versions of other, putatively, exemplary tests of executive function such as the "Trails task." In other words we have not been able to show that the term "inhibition" has empirically demonstrable construct validity, even when assessed by tasks that have been taken as operational definitions of its use. This is, perhaps, not surprising given that those who have used the term "inhibition" as a component of executive behaviour have tended to be somewhat promiscuous in their choice of definitions of its aetiology, i.e. as an observable property of single neurones, as a theoretical construct in connectionist simulations of learning networks, or as a property of particular information-processing "modules" in the prefrontal cortex, or as an important general property of the "attentional system." A pervasive difficulty seems to be that executive tasks are, necessarily, very complex, and that attempts to fit them into linguistic categories borrowed from everyday discourse such as "inhibition," or "planning," or "monitoring" are necessarily Procrustean. It may not be possible to give useful, generic descriptions to particular "kinds" or "styles" of executive task, and we may have

to accept, as scientists modelling behaviour in terms of production systems have shown, that a very wide range of apparently dissimilar behaviours may, in fact, all derive from the same small number of assumed system properties (Anderson, 1993).

(7) A related concept, that of "task purity," has been cited as a difficulty in all of neuropsychology, and perhaps especially in the study of executive function (Weiskrantz, 1992). That is to say, the tasks available to assess executive function are "impure" in the sense that they may also make demands on a variety of other cognitive skills or functions that are supported by brain structures that are quite independent of frontal cortex. An associated problem is that when laboratory tasks are used to investigate individual differences in cognitive function it has become conventional to measure only a single performance index such as "reaction time" or "percentage correct" and to use this as the sole standard of comparison between older and younger or more or less "intelligent" individuals. This restricts what we can learn, because these single task performance indices necessarily represent the pooled outcomes of many distinct functional processes. Experiments by Nettelbeck, Rabbitt, Wilson, and Batt (1996) and Rabbitt and Yang (1996) show that it is usually possible to obtain not merely one but several different and independent performance indices from any given task. When this is done, best-fitting Lisrel models show that indices reflecting information-processing speed are independent of indices reflecting accuracy of retrieval of information from memory, or rate of learning. These techniques of task decomposition, exploring relationships, and independence between different task performance indices, and so of the functional processes that they reflect, may be one way to explore problems of "task impurity." Unfortunately the problem is deeper, and our attempts at assessing executive function may be balked by what we may term "process impurity" rather than "task impurity."

(8) Explorations of executive function have used the classical methodology of human experimental psychology: to try to develop tasks in which we can control as many demand variables as possible in the hope of isolating, quantifying, and measuring the effects of some single critical variable that, speculatively, taxes one single, hypothetical functional process and not others. It may be that this venerable strategy is entirely inappropriate for analysing executive function because an essential property of all "executive" behaviour is that, by its nature, it involves the simultaneous management of a variety of different functional processes. For example, to the extent that we can design experiments to test the performance parameters of each of these subprocesses on its own, as when studying Baddeley and Hitch's "articulatory loop" or "scratch pad" systems, we succeed in studying lower order "automatic" function rather than "executive control." In order to study how these processes are controlled we need to develop not less, but more, complex tasks and to find ways of simultaneously measuring

a variety of different, independent, task performance indices and comparing them between executively intact and executively challenged individuals.

(9) It is possible that "executive" tests may be validated in terms of their correlations with scores on tests of general fluid intelligence that load heavily on the single factor of general ability, *g*. This would be consistent with empirical findings by Duncan (1995) that initial performance on his "goal-neglect" task is predicted by scores on the Cattell "culture fair" intelligence test, which does load highly on g_f, but is not predicted by scores on the WAIS-R, a test with a much lower g_f loading. We shall consider the relationship between intelligence test scores and frontal function in more detail below. For example, consider the use of these assumptions in experiments by Baddeley, Maylor and Godden (in press) and Lowe and Rabbitt (this book), who gave large groups of elderly individuals aged from 60 to 80 years tests of "executive function" such as the Stroop test, Baddeley's stimulus dimension switching task and, in the Lowe and Rabbitt study, also a version of the Trails test. All participants were also scored on the AH 4 (1) and the Cattell culture fair IQ tests, and on other measures of information-processing speed, such as 4-choice RT. As expected, higher IQ test scores and information-processing speed improved, and old age significantly reduced efficiency of performance on all executive tasks. The aim was to discover whether all changes in executive task performance that occur with increasing age are accounted for by concomitant changes in information-processing speed and in g_f. In tasks identified as making demands on "executive" function, but not in NE tasks, significant age-related variance in performance still remained even after speed and intelligence test scores had been taken into consideration. One possible account of these findings is that changes in executive function associated with age are partly, but not entirely, accounted for by age-related slowing of information processing and decline in performance on tests of g_f.

This argument now seems uneasy because it begs the question of intelligence test purity. For example, if an even purer measure of g_f than the culture fair test were available, or if composite scores from several different intelligence tests were entered into the regression equations as predictors, it seems very likely that all, rather than only some, of the age-related variance in executive tasks would be accounted for. Such findings would be evidence that, when a suitably stringent measure of g_f is used, all age-related changes in executive function can be explained in terms of a reduction in a singletask performance parameter, g_f. The logical difficulty with these attempts to show that age-related changes are due to changes in executive function, rather than in g_f alone, is similar to that with Salthouse's (1996) attempts to show that all age-related variance in memory test scores can be accounted for by age-related change in information-processing rates. Salthouse showed that, while age-related variance in memory test scores is only slightly reduced when times scores from a single, simple task such as 4-

choice RT are entered as predictors, as increasing numbers of measures of "information-processing speed" obtained from increasingly complex tasks are added, the remaining age-related variance is progressively reduced and finally disappears.

Salthouse concludes from this that all age-related variance in memory tasks can be shown to result from a "global" slowing of information-processing speed if scores from sufficiently many and sensitive tests of information-processing speed are used as predictors in regression equations. This can be criticised as a tautology because he achieved this result by progressively including in the regression equation scores from increasingly complex tasks, including ones such as letter/letter substitution coding which are known to have a high memory component. As an armchair experiment it is certainly possible to envisage a similar "test battery for age" on which scores correlated strongly with people's dates of birth and which, when gross unadjusted scores were entered into a regression equation, accounted for all age-related variance in most other tasks. A demonstration that particular subsets of task performance indices do, and others do not, detect all of the variance in individual differences in performance associated with an individual difference variable, such as age, would not get us very far. The particular difficulty when g_f is included as one of these performance indices is that it is a statistical, and not a functional, construct. It represents that proportion of variance in individual performance that is common between a very large number of different intelligence tests, each of which incorporates demands on a very wide range of different cognitive functions: among these information-processing speed, working memory capacity, planning, control of selective attention, and, as Duncan has elegantly demonstrated, the ability to keep goals in mind, to monitor performance, and to switch goals (or S-R codes) during a task. In other words g_f has a statistical definition in terms of that large proportion of variance that is shared between complex, intellectually demanding tasks which, because they correlate modestly with performance in many everyday life situations, we call "intelligence tests." It has no functional definition because it represents a commonality of prediction by tasks that all make as wide a range of demands as possible, and so involve a correspondingly wide range of functional processes. It follows that scores on intelligence tests that load highly on g_f will modestly predict performance on simple tasks, such as "inspection time" or "choice reaction time" which share some of the demands that the problems make, but will much more strongly predict performance on complex tasks, such as other intelligence tests, that share an even wider range of the demands that they make. In other words, g_f cannot be reified in terms of any particular cognitive process, or set of processes, precisely because it is *a measure of the amount of overlap in functional demands* made by complex intelligence tests that have developed to make simultaneous demands on as wide a range of different processes as possible.

(10) A quite different problem, which does not seem to have been considered by any of our contributors, is that there may be no functional distinction between "executive" and "non-executive" processes, in the sense that they are not discrete steady states of the cognitive system but arbitrary points along a continuum determined by the level of practice attained. We consider this possibility below.

EXECUTIVE AND NON-EXECUTIVE PERFORMANCE?

Because the terms "controlled" and "automatic" processing are very widely used to make the same distinction, many readers may have found the terms "E"and "NE" processing coy and clumsy. However the terminology has been deliberately chosen to signal the point that, although all investigators agree that the terms "controlled" and "automatic" derive from classic studies by Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977), the stringent empirical criteria on which these authors based their distinction seem to have been forgotten. These classic empirical studies and Schneider's (1993) more recent theoretical discussions make it clear that the distinction between "controlled" and "automatic" processing marks qualitative, rather than merely quantitative, changes in performance that occur when tasks are very highly practised.

(1) Controlled processing is very sensitive to variations in information load. Specifically, when unpractised people search for some target symbols among others they become either slower, or less accurate, as the number of different items for which they search is increased from two to six or more. After prolonged practice *with identical target sets*, increases in target set size that initially slowed speed and reduced accuracy no longer affect performance.

Shiffrin and Schneider's (1977) original and indispensable insights illuminate earlier findings by Rabbitt (1959) that when people first encounter a task in which they have to discriminate between two different sets of symbols or words, their decision times become slower as the number of items in each of the two response classes is increased. As practice continues, the slopes of functions relating decision times and accuracy to set size steadily reduce until no rises in latency or errors occur, even when the number of items in each symbol set is increased from 2 to over 16 items. Experiments by Rabbitt (1965) made the additional point that, early in practice, people discriminate between target symbols by using many more than the minimum necessary number of features, cues, or points of difference between them. As practice continues the number of features that are selectively identified and compared approaches the theoretical minimum needed to make the discrimination. Given Shiffrin and Schneider's (1977) insights these findings can be reinterpreted as evidence that when people first attempt tasks, they attend to, and process, much more than the theoretical minimum amount of data that they need in order to make the discriminations required of them. Unless they can do this they cannot hope to optimise their performance because it is only by first considering and evaluating as wide a

range of all features as possible that they can they can hope eventually to discover the most effective minimum subset. A consequence of this might be that reductions in system capacity (as perhaps by old age, or diffuse brain damage in closed-head injuries, or less permanently by acute ingestion of alcohol) will impose a resource bottleneck on the complexity of initial, controlled, processing that is not felt on highly practised, and so fully "automated," processes that have come to make much smaller informational demands. A corollary speculation is that a reduction in the efficiency of initial controlled processing is likely to delay, and may even entirely prevent the eventual attainment of automaticity with practice.

(2) As a task is practised and becomes increasingly automated, the amount of information that has to be processed in order to carry it out steadily reduces and it begins to interfere less with other activities that are carried on at the same time. As a result, highly automated tasks can even tually be carried out with no apparent mutual interference (Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977). However, later statements by Schneider and Detweiler (1988) emphasise that optimisation of dual-task performance cannot be achieved by practising each task, separately, on its own. The two tasks must be practised together in order to achieve the qualitative changes in performance that are necessary to minimise mutual interference.

(3) Once a task is performed automatically, it can apparently also be performed unconsciously. That is, people can seldom remember and report the last decisions that they made before being suddenly interrupted while carrying out a very highly practised task. During controlled performance people can almost always remember their last decisions before unexpected interruptions.

(4) Authors who have adopted Shiffrin and Schneider's distinction between controlled and automatic processing as synonymous with that between E and NE function have usually neglected their most important empirical distinction: automatic performance does not generalise from one task to another. Indeed automated performance is so intensely task, or situation, specific that even apparently trivial alterations to task demands result in an immediate regression to "controlled" behaviour and a consequent, abrupt, loss of speed and accuracy. Moreover, if task demands are constantly varied, even in the minor sense that the subsets of target and background letters in a visual search task are slightly altered from session to session, "automatic" performance is never attained. An exception to the general neglect of this issue is an observation by Duncan (1986) which raises the interesting point that the size of the information unit employed in the task may vary so that, under some circumstances, quite wide changes in the physical characteristics of a display may be made without loss of automaticity so long as higher order constructs remain constant.

These objective empirical criteria clarify the distinction between controlled and automatic information-processing modes in two ways: First, they show that

it is misleading to discuss "controlled" and "automatic" behaviour as if they were discrete steady states. They are, rather, only labels designating arbitrary points along a seamless continuum of change with practice. When a novel task is first encountered processing is mainly, or entirely, controlled, and then immediately begins to pass through a continuum of empirically indistinguishable intermediate states between which there are, apparently, no abrupt disjunctions. Second, the Shiffrin and Schneider criteria show that it is misleading to speak of either of these processing modes as being more or less "efficient" in any absolute sense. They are all optimally efficient for the particular level of practice at which they become possible. Early in practice it is highly efficient to consider, and to use, a very wide range of sensory information and other data in order to be able to discover, as quickly as possible, what are the best procedures to achieve the desired outcomes. The broad range of information that must initially be used steadily narrows as optimal procedures for making decisions are gradually discovered and employed. As the procedures employed approach the minimum necessary to carry out the task efficiently performance becomes increasingly fast and accurate, but also increasingly inflexible. Thus, as practice continues, generalisation to other, similar, tasks gradually becomes less possible and, eventually, even slight changes in task demands bring about a regression to a "controlled" mode in which novelty can begin to be assimilated and the process of optimisation can begin anew.

Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977) carefully defined their terms "controlled" and "automated" in terms of very precise empirical criteria, ostensively defined by many, thoroughly replicated, empirical studies. If we choose to use these terms as synonyms for "executive" and "non-executive" processes we must carefully consider how much of this valuable conceptual and empirical legacy we wish to buy into. To do this, it is helpful to consider some formal models for the attainment of automaticity, and for the disruption of "controlled" or "executive" function. As a heuristic device it may be useful to consider how well we can account for the range of characteristics in terms of which our authors have distinguished E from NE processing in terms of a very simple model adapted by the author from a proposal by Crossman (1958).

A DESCRIPTIVE MODEL

The model envisages an information-processing network within which an indefinitely large number of alternative pathways for stimulus processing and response choice are available between the input end, I, and the output end, R.

When the network first encounters a novel task or visual display, there is no selection of input so that many nodes at I are likely to be activated. These will include those nodes that can encode aspects of displays which are optimal for the decisions required, as well as others nodes that encode aspects of displays that

are irrelevant, or even misleading, for this purpose. The same will be true of nodes intervening between I and R, which can be envisaged as a large number of activated information-processing pathways, or procedures. Most of these will be useless, or even counterproductive, for the decisions required. Others will be able to carry out task-appropriate information-processing transactions with widely varying degrees of redundancy, and so efficiency. The model makes the further assumption that, as the network is increasingly trained on a particular task, efficient pathways are increasingly selected and inefficient or irrelevant pathways are increasingly disused. This gradually reduces the number of activated nodes in the network towards the minimum necessary to make the required decisions and also increasingly liberates other pathways to cope with other, concurrent, information-processing activities.

Although this model is conceptually very simple, it can meet all of the criteria used by Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) to distinguish controlled from automatic processing in visual search. The first criterion, that once automatic processing is attained, wide variations in information load have no measurable effect on speed or accuracy, can be explained by selection of increasingly optimal pathways allowing the performance of the network to converge on its theoretical limit for parallel information processing. At this point variations in task information load that do not exceed the limit for parallel processing do not cause corresponding changes in speed or accuracy of task performance.

Incidentally, Crossman (1958) proposed this model to predict another characteristic feature of improvement in performance with practice, with which Shiffrin and Schneider were not concerned. He formally showed that gradual discard of less efficient and selection of more efficient pathways not only generates the power function learning curves for average decision times that seem to be invariably observed for tasks of all kinds (see Crossman, 1958; Newall & Rosenblum, 1981), but also predicts the concomitant, logarithmic, reduction in trial-to-trial variance in decision times which seems to be observed in all tasks as practice proceeds.

The second criterion of automatic processing, the increasing ability to carry on the task without interference from other concurrent activities, can be explained by progressive disuse of inefficient pathways as automaticity develops. This gradually makes information-processing capacity available to meet other, concurrent, demands.

The third criterion, that events before sudden interruptions can be recalled during “controlled” but not during automatic processing, can be explained by assuming that, early in practice, inputs result in activation of entire sheaves of pathways and so also set up more extensive, and perhaps correspondingly more persistent, memory traces. In contrast, late in practice, the activation of a few uniquely efficient pathways soon becomes indistinguishable from system

background noise. Admittedly this is an *ad hoc* speculation that does not explain "reduced consciousness," but only "reduced persistence" of the after-effects (memory traces) of events. In the absence of other empirical criteria for consciousness than remembering that an event occurred, perhaps this does not matter. Another way to think about this might be that, early in practice, detailed monitoring of decisions is necessary and this cannot occur unless much, potentially redundant, information about the input that is being categorised and the decision that is being made can persist in the system. As practice reduces the necessity for detailed monitoring, persistence of information in the system becomes less helpful.

The crucial defining characteristic of automatic task performance is its intense task specificity. The model accounts for this by supposing that the number of different information-processing pathways employed to carry out a task gradually reduces until they approach the minimum necessary to carry it out successfully. At this point specificity is so intense that generalisation is impossible and even slight changes in task demands make use of previously optimal pathways impossible and necessitate reversions to a "controlled" mode of processing in which changes can be identified and new optimal pathways can be discovered to accommodate them.

This model is a simplistic example of a particular style, or genre which may be called "metaphorical connectionist." Nevertheless it can account for many of the methodological difficulties encountered in studying executive function. For example, executive tasks are bound to have low test/retest reliability because practice brings about qualitative changes in the way in which they are performed. Since functional processes become intensely task-specific we would not expect performance on one task to strongly predict performance on others, even if they are apparently very similar. These and other predictions are also made by the much better-developed connectionist and production system models for executive control developed during the past decade.

CONNECTIONIST AND PRODUCTION SYSTEM MODELS

During the late 1980s and early 1990s many clever and illuminating attempts were made to devise connectionist models to account for some of the clinically observed effects of frontal lobe damage. In relation to previous discussions an interesting characteristic of all of these models is that they each undertake to describe performance only in an intensely specific task domain. Investigators selected particular tasks that have been widely accepted as diagnostic tools to assess behavioural changes consequent on damage to prefrontal cortex. For example, Levine and Prueitt (1989) proposed a connectionist network that handled novelty and which, when degraded, showed response perseveration similar to that observed in some patients. Dehaene and Changeux (1989, 1991)

proposed connectionist models for normal and impaired performance on delayed response tasks and on the Wisconsin Card-Sort Test. One of the most comprehensive and widely cited of these attempts is by Kimberg and Farrah (1993), who undertook to model the effects of frontal damage on four tasks: motor sequencing (following clinical observations by Jason, 1985; Kimura, 1977; and Milner, 1982); the Wisconsin Card-Sort Test (following many studies, e.g. Anderson et al., 1991); the Stroop test (see review by McLeod, 1991); and memory for context, or source memory (Shallice & Burgess, 1991).

Kimberg and Farrah (1993) used the Act R production system architecture developed by Anderson (1993) to construct separate, independent models for each of these systems and to compare the behaviour of these simulations when links between the representations of productions and representations of operations were either optimally weighted or weakened. As they expected, they found that weakening of these links caused increases in errors of all kinds but, interestingly, in all tasks, increases in perseveration errors were especially marked.

Apart from its intrinsic interest and ingenuity, the Kimberg and Farrah (1993) model has the important feature that it accepts that, while different kinds of behaviour may, indeed, be controlled by logically identical "rules" or "procedures," these procedures may be incorporated in independent, and intensely task-specific, production systems which are developed to cope with different situations. Kimberg and Farrah call these "information-specific working memory systems." The model is instructive in two ways. First, it gives a well-articulated account of how performance in quite different cognitive skills may be controlled by logically identical production architectures, and how damage to these production systems can have logically similar effects on the performance of very different cognitive skills. That is, it shows how changes in a single, low-level, system performance characteristic, such as connection weights, can produce similar effects (perseveration errors) in quite different tasks. Second, Kimberg and Farrah's "information-specific working memory systems" may, indeed, be quite independent of each other so that damage to some can leave others unaffected. As they point out, this characteristic could account for disassociations between performance on different, though logically similar, tasks that are often reported in clinical practice, and in recent neuroanatomical studies (e.g. by Goldman-Rakic, 1987, and others). In Kimberg and Farrah's models different working memories, which may be supported by different neuroanatomical systems, perform procedurally identical operations on quite different inputs.

There has been recent interest in applying models for changes in frontal function to conditions such as closed-head injury, "normal ageing" (see Parkin, and Robbins et al., this book) and Alzheimer's disease (see Baddeley, Logie, Bressi, Della Sala, & Spinnler, 1986; Baddeley et al., this book) in which brain

changes are usually diffuse, rather than focal. An alternative way of putting this is that the diffuse brain changes, such as the widespread loss and alteration of neural tissue in "normal" old age described by Brody (1994), Flood and Coleman (1988), and Parashos & Coffey (1994) and many others, are likely to result in a degradation of performance of all of the variety of distinct, and separate, "working memories" that Kimberg and Farrah's and other similar models envisage. In this case we might indeed expect to see similar age-related performance deficits in superficially very dissimilar tasks. It would be possible to misinterpret these similar qualitative changes in performance on a variety of different tasks as evidence of damage to, or loss of capacity by, some higher order, "central executive" system that applied the same set of rules or control processes to a variety of lower order, subordinate systems. Kimberg and Farrah's model saves us from the belief that this, rather clumsy, conceptualisation is, in any way, obligatory.

An inconspicuous, but crucial, feature of all of these models is that they envisage systems that learn, by appropriate practice, how to carry out particular tasks optimally well. The elementary learning network model, derived from Crossman (1958), makes this point by using a simplistic assumption that novel tasks are extremely demanding of information-processing resources because they involve activations of redundant and unnecessary units and pathways and that increasing use of critical, and discard of redundant, units and pathways not only brings about quantitative but also qualitative changes in efficiency of task performance. One of the most dramatic and intriguing of these changes is the increasing ability to carry out simultaneously other tasks without interference. This is important in "frontal studies" because dual-task performance has been an empirical tool of choice both in definitions of component subsystems of the Baddeley "working memory" system, but also in recent studies of changes in executive control (see Baddeley, Della Salla et al., this book). A recent model by Schneider and Detweiler (1988) provides a useful corrective to the idea that "dual-task performance" can be completely understood in terms of a progressively more economical sharing of a "common resource" which can be provisionally defined in terms of the number of available units and connections between them, and quantified and compared across different pairs of possible simultaneous tasks in terms of a single empirical index such as "M-space" (Case & Sandieson, 1992) or mu (Baddeley et al., this book). Computations of these and other global indices of joint resources very usefully allow comparative analyses of the joint demands of different kinds of tasks. However, this misses the point that in order for dual-task performance to become possible qualitative as well as quantitative changes must occur. Whether we can find empirical tests capable of detecting transitions between the successive, staged, qualitative changes specified in models such as that by Schneider and Detweiler (1988) remains an interesting question.

SCHNEIDER AND DETWEILER'S MODEL

Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) have shown that it is eventually possible for two tasks to be carried out without mutual interference, but this only happens if they are practised together, rather than separately. Schneider and Detweiler (1988) formally propose a number of successive, qualitative changes which must take place before dual-task performance becomes maximally efficient. The model envisages that even single-task learning involves a transition between successive, qualitatively distinct, information-handling stages. When a task is first encountered it initially involves a very heavy memory load because it is necessary to maintain in working memory, and to load as necessary, a variety of memory vectors, representing almost all possible connections between representations of information and operations that have to be carried out. In a subsequent processing stage, that of context-controlled comparison, the representation of information is translated into a more economical set of memory vectors, which are maintained and used. Further practice enables Goal—state maintained controlled comparison in which target memory vectors are directly transmitted from input to produce output; finally, automatic single-task processing allows even more economical use of memory vectors, and so of what might be termed, in the context of such quantification as M-space, or mu, maximally economical use of processing space. Dual-task performance requires at least seven further changes. In order of their use during practice on dual tasks these are: (1) shedding the tasks of pre-loading information for each task, and of time-scheduling its access and use; (2) letting go of higher workload strategies that are necessary to cope with complex tasks; (3) utilising non-competing resources, as far as possible; (4) multiplexing transmissions over time (i.e. increased parallel processing of information from both tasks); (5) reducing redundancy to a minimum by shortening necessary transmissions; (6) converting interference from concurrent transmissions; and, finally, (7) achieving economies of information processed by chunking transmissions.

In distinction to the simplistic network model with which we began, this sophisticated model does not merely assume that individuals improve by carrying out fewer and fewer of a set of operations that are, essentially, of the same kind but rather that they proceed through a decalage of successive, independent "stages" in each of which the production system carries out a different pattern of operations. The challenge is to invent experimental tasks in which identifying characteristics for each of these successive processing stages may be observed. Until this is possible empirical neuropsychologists may have to be content with attempts to improve the sensitivity of global quantitative indices of dual-task information load, such as mu (Baddeley, Della Salla et al., this book), which at least allow comparative statements of the relative amounts of capacity consumed by each of two, simultaneous and competing, tasks.

It is also worth noticing that this model, involving progress through successive different stages of information processing, does specify a higher order controller. The controller, of course, need not necessarily be a single, supervisory system that organises these transactions in any and all possible pairs of tasks. Though this is not specified in the model, it seems likely that it would be necessary to have different controlling systems that are dedicated to the operation of different pairs of tasks. In other words, intense task specificity will remain a characteristic of the system.

GENERAL ISSUES

The most striking point is that any task, however complicated, is, as Burgess remarks, "only novel once" and it is only on initial presentation that we can be sure that it works as a diagnostic tool. In his discussion of his clever "goal-neglect" task Duncan (1995) also stresses that only initial trials are diagnostic of executive deficits. In other words, the main operational distinction between "controlled" or "E" and "automatic" or "NE" processing is amount of practice. This is explicitly recognised by models designed to model the transition, with practice, from controlled to automatic behaviour such as Schneider and Detweiler's (1988) model for development of automaticity of dual-task performance and in the model derived from Crossman (1958); it is not overtly recognised in the Kimberg and Farrah (1993) model, which is not concerned to describe how task-specific working memory systems are developed but only to simulate the effects of damage on "working memory systems" that have matured to a steady state. If we accept that prolonged practice on any task brings about continuous, rather than disjunctive, qualitative changes in the ways in which information is processed, E and NE processing are best defined as the hypothetical limiting points of this continuum. Controlled and automatic processing are arbitrary, rather than empirically definable, limiting points on this continuum because, at least for adults, no task can be entirely novel, and so unprepared by any previous experience whatever and it is difficult to find examples of tasks that have been so highly practised that we may confidently assume that no further improvement may occur. The idea that we must conceptualise the systems that we study as constantly engaged in empirically elusive transitions between states, in a condition of "becoming rather than being," could account for the general findings by clinicians, clearly articulated by Burgess, that so-called "executive" tasks tend to have low test/retest reliability and consequent high measurement error.

All of these models envisage systems that are intensely task-, or domain-, specific in the sense that they have developed to deal only with particular classes of information, or to control some particular activities and not others. This is made particularly clear in Kimberg and Farrah's (1993) speculation that there may be a multiplicity of independent working memory systems, each of which

has been especially developed to deal with particular tasks such as the Stroop test, the Wisconsin Card-Sort Test, or with programming of sequences of responses or retrieval of contextual information from memory. It is also a basic assumption of the Schneider and Detweiler (1988) model that maximally efficient dual-task performance requires continued practice on a given pair of tasks, and cannot be transferred if the demands of even one of these tasks are changed. Inferentially, focal brain damage may impair some, but not all, of these highly specialised "dedicated" task-specific systems and we should expect to find that, as Burgess and other clinicians point out, the diagnostic value of individual tests for executive function may be quite limited because patients who fail on some executive tasks may, nevertheless, often successfully perform many others.

Nevertheless these models also suggest how exceptions may occur to this general rule that the effects of focal damage are likely to be specific to some and not general to all tasks. Some models predicate single higher order control systems that may organise subordinate systems which each carry out one of a pair of simultaneously performed tasks (Schneider & Detweiler, 1988) or that each deal with a different type of information (Schneider, 1993, 1994). As these authors point out, it is also possible to envisage hierarchies of control systems in which higher order systems control other domain-specific subsystems. If such hierarchical systems exist it is possible that damage that specifically affects a higher order system might allow its associated lower order systems to carry out automatic information processing but only in limited, and uncoordinated, ways. If damage affects some, but not other, specific dedicated subsystems (Kimberg & Farrah, 1993) and also some, but not other, superordinate controlling systems, we would expect the confusing disassociations of effects that are sometimes reported in the clinical literature.

These models also provide frameworks in which to discuss the effects on "executive" function of diffuse damage to the cortex, such as may occur in "normal" old age (Brody, 1994; Flood & Coleman, 1988; Parashos & Coffey, 1994) or in closed-head injuries. It follows from Kimberg and Farrah's (1993) elegant demonstrations that it would be possible to model "global" impairment of the cortex as a simultaneous "weakening of connections" in all of a wide range of dedicated production systems, with consequent impairment of performance in all of the different dedicated "working memory" systems responsible for a wide variety of tasks. In models such as those described by Schneider (1993, 1994), diffuse damage may be supposed to impair selectively the efficiency of higher order systems in a control hierarchy, because these make the greatest demands on resources. This would disrupt scheduling and control of complex behaviour sequences, with preservation of simple sequences and of single responses. Thus in laboratory experiments we might expect "global" losses of efficiency to be reflected in loss of speed and accuracy in solving complex problems, such as those included in intelligence tests, with consequently reduced

scores and estimates of levels of general fluid intellectual ability or g_f (Duncan, 1995). Loss of global efficiency may also be expressed as loss of speed and accuracy in very simple tasks, such as choice reaction time or tachistoscopic recognition thresholds (Eysenck, 1986; Jensen, 1980, 1985, 1987; Salthouse, 1985) or as a reduction of working memory capacity measured in laboratory experiments (Just & Carpenter, 1992; Kyllonen & Krystal, 1990; Salthouse, 1991).

All of the models that we have reviewed assume that a controlled mode of information processing is essential to cope with initial attempts to carry out complex novel tasks. They all illustrate that controlled processing is "resource expensive," in contrast to the "resource thrifty" automatic performance which prolonged practice on the same tasks eventually allows. A plausible corollary is that "global" reductions in information processing and working memory capacity are likely to impair initial performance of complex, novel tasks that require resource expensive controlled processing more than performance on simple tasks that never make severe demands on information-processing resources, or of complex tasks in which automaticity, achieved by extended practice, has reduced demands. The different effects of "global resource reduction" on controlled and automatic tasks would explain observation that cognitive skills that disrupt performance on complex novel tasks may have no effect on highly practised tasks that, apparently, make very similar, and equally complex, demands (e.g. Luria & Tsvetskova, 1964; Walsh, 1978). It would also explain cases in which performance on tasks that make quite different functional demands may be jointly impaired. If we add the possibility that even relatively focal cortical damage may also reduce global efficiency we would expect the confusing picture observed in clinical practice, in which patients may entirely lose some highly specific skills but be apparently unimpaired on other, apparently very similar, skills and also have a generally reduced ability to cope with complex novel task demands. This might also explain why, when we give large groups of normal adults batteries of tests which have been developed to reveal losses of specific "executive functions" in brain-injured patients, we may often, or even usually, fail to find any patterns of association between scores that are consistent with our current assumptions of differences between the various neurophysiological processes that the tests are supposed to reflect. In intact brains performance on disparate skills correlate because they all reflect the general level of available cognitive resources.

These models also suggest that, even when brain damage or old age causes losses of global information-processing resources that are not so severe as to prevent performance on any novel tasks, they may nevertheless be sufficient to reduce the maximum level of complexity of tasks that can be dealt with, to slow the rate at which tasks can be automated and, perhaps, to limit the level of automation that even very extended practice can bring about. For example, in

Schneider and Detweiler's (1988) model, global degradation of the system would not only compromise the resource expensive processing that is initially necessary to carry out complex novel tasks but also impair the ability to progress, with practice, through the seven successive processing stages necessary to attain maximally efficient dual-task performance. Investigators seem to have neglected the possibility that dysexecutive behaviour may reduce the rate, or restrict the level, at which automaticity can be attained. The obvious theoretical demands for further research on the extent to which failures to cope with executive tasks can be ameliorated by practice are paralleled by the urgent concerns of practitioners such as Wilson and Evans, (this book) who try to find ways in which training can improve the abilities of their clients to cope with their everyday lives. The rate of improvement with practice, and the levels of performance that can ultimately be attained, are major concerns in rehabilitation. Wilson and Evans emphasise the further concern that levels of performance on the tests developed by clinicians to diagnose probable locus of damage are poor at predicting either performance in everyday life, or the likelihood of long-term improvement. Most of the models we have considered would suggest that, while levels of performance on initial testing on "executive" tasks may give relatively weak predictions of subsequent rate of improvement, the observed rate of improvement with practice would strongly predict the ultimate level of performance that can eventually be attained. If this were so, the techniques that are currently used by clinicians to predict the likely benefits of rehabilitation therapy could be much improved and would gain a useful theoretical framework.

Some of the models that we have reviewed also suggest that global losses of efficiency may also degrade the performance of higher order components in a hierarchy of production systems which we might suppose to be more resource expensive than others. We might expect this to reduce the efficiency with which the activities of lower order, automated, subsystems can be sequenced and integrated to perform complex tasks. Thus a global reduction in information-processing resources might cause similar failures of organisation and control of behaviour in a variety of superficially dissimilar tasks.

A final point is that all of these models, particularly those by Kimberg and Farrah (1993), show that performance of tasks that make very different demands can nevertheless be successfully simulated by production systems with identical architectures. This suggests good reasons for the low construct validity of terms such as "planning," "inhibition," "impulsivity" and "memory of context" which have been borrowed from the language of everyday subjective experience in order to act as labels for hypothetically distinct "components", of executive behaviour. Within the context of models such as that proposed by Kimberg and Farrah (1993) it is clear that terms such as "inhibition" or "perseveration" may provide a convenient taxonomy of task demands but not of the functional processes by means of which the tasks are carried out. In short, the poor construct

validity of the terms we use is often the result of a confusion between the levels of description of task performance indices and system performance characteristics, and can explain why our data analyses often fail to meet our expectations.

None of these models was intended to account for individual differences in the initial performance of novel "executive" tasks, or for variance in the rates of transition from controlled to automatic performance. Nevertheless, all of them suggest that the level of initial performance of complex novel tasks, the rate at which automaticity is attained, and the maximum degree of automaticity that is possible to a system may all vary with the overall level of "global system resources. This may be affected, for example, by the processes of "normal ageing," by the progress of dementia, or by both focal and diffuse brain damage. This raises the question whether we have any empirical way to compare individual differences in overall levels of resources that may determine corresponding differences in efficiency of attentional control, of information-processing rate and of working memory. Recently investigators have become intrigued by reliably replicable findings that scores on pencil and paper "intelligence tests" modestly, but very reliably, predict performance across a wide range of different laboratory tasks assessing information-processing speed (Eysenck, 1986; Jensen, 1980, 1985, 1987), working memory capacity (Carpenter, Just, & Shell, 1990; Just & Carpenter, 1992; Kyllonen & Krystal, 1990) and the efficiency with which attention can be allocated to subtasks demanded by a complex skill such as driving (Duncan et al., 1993). More recently Duncan (1995) has suggested that individual differences in errors on a "goal-neglect" task, which he proposes as a particularly sensitive index of dysexecutive behaviour, are well predicted by intelligence tests such as the Cattell and Cattell (1960) culture fair, which are known to load highly on Spearman's (1927) common factor of general fluid mental ability, g_f, but, in contrast, not by scores on other tests, which have relatively low g_f loadings, such as the WAIS—R. Within the framework of the models that we have discussed this opens the larger question of whether individual differences in global cognitive resources that determine level of central executive competence may be empirically assessed by intelligence tests, theoretically interpreted as differences in global central nervous system efficiency and quantified and compared in terms of a single index, g_f.

"EXECUTIVE TESTS" AND "INTELLIGENCE TESTS"

Phillips (this bopk) shrewdly points out that in many respects the history of discussions of executive function closely parallels that of discussions of general intelligence. Psychometricians draw a clear distinction between performance on novel and complex tasks or tests that require "fluid general mental ability", or g_f, and performance on tasks that depend on learned procedures and acquired

knowledge of the world which can be described as "crystallised intelligence" (Horn, 1982; Horn & Cattell, 1966). This distinction is supported by observations that the progress of "normal ageing" which is associated with diffuse neuronal loss and change, particularly to the frontal lobes (Brody, 1994; Flood & Coleman, 1988; Parashos & Coffey, 1994) brings about marked declines in scores on tests of fluid intelligence *(g_f)* but relatively little decrement on tests of learned procedures and information, such as social skills, knowledge of the world or vocabulary, that have become "crystallised" by a lifetime's practice. This precisely parallels the distinction made by Luria and Tsvetskova (1964) and Walsh (1978) between "executive or controlled" functions, which are vulnerable, and "automatic" skills, which are relatively resistant to brain damage.

Like tests of "executive function", tests of "fluid intelligence" must be novel in order to be valid. Training allows even people of modest general ability to solve difficult test questions rapidly and accurately and so to reach ceiling scores on most intelligence tests. To adapt Burgess's useful phrase "no problem can be an intelligence test twice" and, alike for intelligence tests and for diagnostic tests of dysexecutive functioning, the fact that procedures necessary to support very complex activities rapidly become "automated" by practice means that overall level of g_f, or executive function, cannot be assessed by any single task or behaviour, even if it is very complex. For example so-called "idiots savants" brilliantly execute extremely complicated procedures that they have spent years perfecting, such as playing back note-perfect renditions of complicated pieces of music that they have heard only once, or flawlessly carrying out very elaborate mental calculations, but may be entirely at a loss with other, unpractised, tasks that seem trivial to most other humans. Attempts to define intelligence and executive function encounter an identical difficulty: no, single, exemplary task, or even subset of tasks provides an adequate ostensive definition. It is often necessary to fall back on consensus definitions drawn from the common sense of the "man in the street" (in the case of Sternberg and Powell's illuminating empirical assay (1982), of people on a railway station) or poll the collective wisdom of "distinguished experts in the field." Like definitions of "executive" function these tend to be wide-ranging catalogues of examples of intelligent behaviour and to avoid entirely discussions of underlying process.

Duncan's (1995) and Duncan et al.'s (1993, 1995) important papers speculate that intelligence tests, such as the Cattell and Cattell (1960), that are good measures of g_f are also the most sensitive predictors of levels of executive function. Wilson and Evans (this book) eloquently complain that though single diagnostic tests developed by clinicians may detect specific dysexecutive problems and, to some extent, define probable sites of brain damage (although see the caveats by Burgess, Foster et al., and other contributors), they are not good at predicting the extent to which patients are likely to be able to meet the

diverse demands of their daily lives. If Wilson and Evans' argument is turned around, it implicitly makes the point that failures to meet complex everyday demands are, as it were, more sensitive diagnostic indicators of general loss of executive competence than are many of the tests developed by clinicians.

Let us consider, as a working hypothesis, Duncan's (1995) suggestion that g_f, as measured by scores on the Cattell and Cattell (1960) culture fair test (CCF), is one of the most sensitive indices of executive efficiency that we have. In this case, as Duncan points out, since only novel tasks involve "executive" function, we would expect CCF scores to predict performance particularly well on initial trials on "executive" tasks such as his "goal-neglect" paradigm. In this case we might expect that CCF scores will predict performance progressively more weakly as practice increasingly shifts performance along the continuum from "controlled" to "automated" mode. Another way of conceptualising this is that as a task is practised the way in which it is carried out progressively changes becoming increasingly different from the way that it was performed on initial trials. In this case we would expect performance on trial 1 to predict performance on trial N=2 robustly, but increasingly less well on each further practice trial. The converse conclusion is that as practice continues so the way in which a task is performed becomes increasingly stereotyped until, when optimal "automatic" performance is finally attained, the task is, eventually, performed in exactly the same way on every trial. In other words, correlations between levels of performance on any trial N_i and on the immediately preceding trial, $N_i - 1$, should steadily improve as trial practice continues.

Figure 1.1 plots these three types of correlations for data from 10 successive practice sessions on a "one back" four-choice serial reaction time task in which 90 volunteers responded, not to the signal currently on the display but to the immediately preceding signal. When the first signal of a series appeared they pressed any response key. The display immediately changed and they then responded by pressing one of four keys that was appropriate, not to the current, but to the immediately previous signal. This process continued through a run of 200 successive signals and responses which constituted a testing session. This task has features characteristic of "frontal" or "executive" task demands. To do it volunteers need to alter and update continually information held in their working memories and, on each trial, to suppress a tendency to respond to the current signal in order to respond to the memory of the previous one. In order to obtain an independent measure of information-processing speed, in addition to this task all volunteers were tested on a straightforward four-choice reaction time task in which they responded, as fast and accurately as possible, to each of four different signals as soon as they appeared.

As training continued, performance on the first practice trial became an increasingly weak predictor of performance on later trials. This suggests that practice does not simply increase the speed and accuracy with which a task is

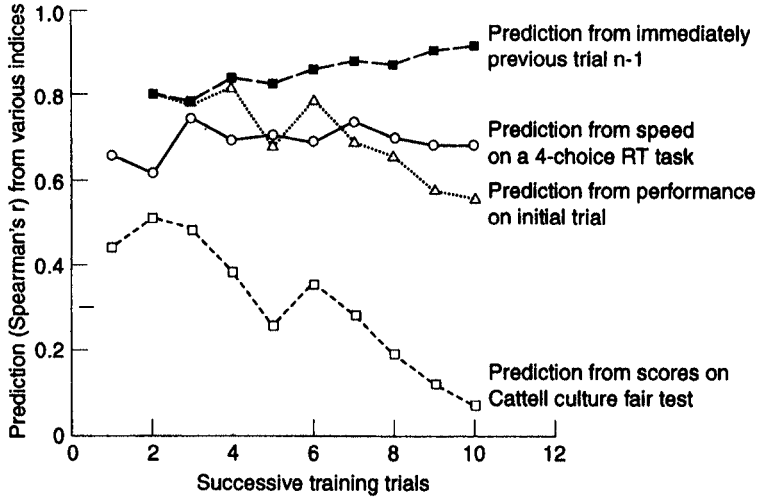


FIG. 1.1. Changes in correlations between performance on a letter/letter substitution coding task over 10 practice sessions, with performance on the initial trial of the same task, with performance on the immediately preceding trial (N–1), with performance on a 4-choice reaction time task, and with scores on the Cattell "culture fair" intelligence test.

performed, but also qualitatively changes the way in which it is carried out. As training continued performance on any trial is increasingly well predicted by performance on the immediately preceding trial. This supports the idea that, as a task is increasingly practised, it begins to be carried out in an increasingly stereotyped manner. In contrast to these two, "task-internal," performance measures, predictions from a task-external measure of information processing speed, the straightforward four-choice RT task, remained roughly constant across all 10 practice trials.

Cattell scores are moderately strong and statistically significant (r=.437 to .481) predictors of performance during the first three sessions but after this correlations steadily decline and, after session 8, they are no longer significant. This is nicely consistent with Duncan's (1995) suggestion that indices of g_f, such as scores on the Cattell culture fair test (CCF), pick up variance in individuals' ability to deal with a challenging novel task that involves running updating of information in working memory and deliberate choice between competing responses. In contrast CCF scores do not pick up individual differences in the efficiency with which the same task is carried out once it has been automated by practice. It should be noted that these findings are closely consistent with those obtained by Ackerman (1989) in other, very different tasks. It seems that intelligence test scores can, indeed, be good predictors of executive or controlled but not of automated procedures. However, we argue that a conclusion that

"executive function" is, in some sense, equivalent to g_f does not solve the problems of functional description of executive processes.

Considered as a simple empirical fact, it is surely unsurprising that scores on pencil and paper intelligence tests, given under time pressure, should predict speed and accuracy on demanding novel tasks. It would be much more surprising if they did not. The enterprise that Burgess terms "ultra-cognitive" neuropsychology has been supported by the expectation that brain lesions are most informative of function when they have very clear-cut and tightly defined effects. As a consequence the diagnostic tasks developed by its practitioners have been deliberately developed to be domain-specific. In contrast, the driving force for development of "intelligence tests" has been the possibility of predicting everyday competence across as wide a range of situations as possible. In this respect intelligence tests may be considered to have been shaped by evolutionary pressures to predict complex everyday activities as much as by practical intuitions and theoretical insights about the nature of "intelligence." The progressive deletion during test development of items that do not, and the selective retention of items that do predict real-life performance have been evolutionary pressures that have ensured that only those tests that consistently make valid predictions of performance of complex novel tasks in everyday life have survived and spawned offspring. Successive generations of tests have gradually become modest, but effective, predictors of "successes in the real world" such as academic achievement, fitness for promotion in the armed services, job status, and lifetime income level. For all their notorious limitations they are still the best practical predictors of everyday competence that we have.

The question is whether intelligence tests make relatively good predictions of executive function because they pick up variance between individuals in a single factor, g_f, or because they include problems that, individually and collectively, make demands on a very wide range of distinct and separate cognitive abilities such as, for example, information-processing speed, working memory capacity, and learning rate, each of which can be measured in isolation and, possibly also more sensitively, by other more specialised tests. The same question must, of course, be asked of putative "executive" tests and tasks. Does a task qualify as a test of executive function only in respect of, and in direct proportion to, its loading on the same single, statistical construct g_f that emerges from factor analyses of scores from a variety of intelligence tests? And, if this is the case, how are we functionally to reify g_f? For example, can we assume, with authors such as Eysenck (1986), Jensen (1980, 1985, 1987) and others, that g_f is, in fact, functionally identical to information-processing rate, or with Carpenter, Just, and Shell (1990) that it is functionally identical to "working memory capacity?"

The data plotted in Fig. 1.1 do not answer this question, but illustrate what may be a necessary approach and eliminate at least one of these possibilities. CCF scores predict initial "executive", or "controlled" performance on the one

back test, but only at the relatively modest level of r= .437 to r=.481. Information-processing speed, measured in a straight-forward four-choice CRT task, makes much better predictions, averaging r =.655 to r=.741 over initial trials and r=.686 across all trials. When CCF scores and four-choice CRT scores are jointly entered as predictors of initial trial one back scores in a regression equation, their joint prediction of .84 is substantially greater than the independent prediction from either alone. Both predict independent variance.

It may seem banal to draw attention to the fact that a straightforward CRT task, that is very similar to the one back task, predicts performance on it rather better than does performance on the Cattell pencil and paper intelligence test. However, this illustrates two points. The first is that made by Ackerman (1989). Unlike intelligence test scores, speed scores predict performance (in a speeded task) throughout practice. This argues that the contributions of speed and CCF scores to predictions of initial "executive" performance of the one-back CRT task are not only statistically, but qualitatively, different. It follows that predictions from g_f and from speed scores are distinct, and that g_f cannot be reified in terms of information-processing speed. The second point has repeatedly been pressed above. "Executive" tasks are clearly various, and may make demands on any or all of a variety of different cognitive skills. It seems natural to expect that performance on any given executive task is likely to be best predicted by performance on other tasks, whose demands closely overlap with it, than by tasks that make very different demands. It is certainly an interesting speculation that there may be a single global property of the central nervous system which underlies performance on all "executive" tasks, and that the levels of this unitary property can be quantified and compared between individuals in terms of a single statistical construct, g_f. To test this it would be necessary to carry out very laborious investigations to determine whether or not scores on a very wide variety of "executive" tasks indeed share more variance on a single, common factor, distinct from other factors that express commonality of other, mutually independent and empirically specifiable, demands such as memory load or information-processing speed. Such data are not yet available.

It seems useful to return to the distinction between task performance indices, which we can measure in laboratory experiments, and the hypothetical "performance characteristics" of the systems that we are concerned to investigate. Levels of any given task performance index, such as CRT, will be determined in very complex ways by a number of different properties of the system from which it is measured. In neural network models these may include the number of layers of units in the system, the degree of connectivity between them, the average delay in mutual excitation between units, the overall level of random noise in the system, the average level of unit activation thresholds, and higher order statistical properties such as "system temperature." The levels of the indices that we measure are weighted resultants of complex interactions between these and

other properties. Collectively, the levels of system operating characteristics define boundaries of a multidimensional space which represents the limits of the system in all possible tasks. Isolated task performance indices that we measure in single experiments only define single points on these boundaries. In a very general sense the boundaries of the system operating space will expand and contract with improvements and declines in the overall levels of its operating characteristics. In this sense we can envisage "global" properties of the system that will determine the limits of performance across a variety of different tasks. This will mean that, within and between individuals, we might expect (relatively weak) correlations between attainable levels of performance on different tasks. The existence of these correlations between task performance indices does not, necessarily, define functional relationships between processes involved in these tasks, nor can any approximate general index of the volume of the system operating space be taken as a direct reflection of any of the functional processes within the system whose "global" efficiency and interactions determine that volume.

REFERENCES

Ackerman, P.L. (1989). Individual differences and skill acquisition. In P.L.Ackerman, R.J. Sternberg, & R.Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 165–217). New York: Freeman.

Anderson, J.R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Anderson, S.W., Damasio, H., Jones, R.D., & Tranel, D. (1991). Wisconsin card sorting performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13*, 909–922.

Baddeley, A.D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A.D., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology, 38A*, 603–618.

Baddeley, A.D., Maylor, E.A., & Godden, D. (in press). *British Journal of Psychology.*

Baddeley, A.D. & Wilson, B.A. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain & Cognition, 7*, 212–230.

Bourke, P.A., Duncan, J., & Nimmo-Smith, I. (1993). A general factor involved in dual task performance decrement. Cited in J.Duncan (1995), Attention, intelligence and the frontal lobes. In M.S.Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–731). Cambridge, MA: MIT Press.

Brandimonte, M., Einstein, G.O., & McDaniel, M. (Eds.) (1996). *Prospective memory: Theory and applications*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Brody, H. (1994). Structural changes in the aging brain. In J.R.M.Copeland, M.T.Abou-Saleh, & D.Blazer (Eds.), *Principles and practice of geriatric psychiatry* (pp. 30–33). Chichester: Wiley.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 234–248.

Case, R. (1992). The role of the frontal lobes in the regulation of cognitive development. *Brain & Cognition, 20,* 51–73.

Case, R. & Sandieson, R., (1992). Testing for the presence of a central quantitative structure: Use of the transfer paradigm. In R.Case (Ed.), *The mind's staircase* (pp. 117–132). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cattell, R.B. & Cattell, A.K.S. (1960). *The Individual or Group Culture Fair Intelligence Test.* Champaign, IL: IPAT.

Cerella, J. (1985). Information processing rate in the elderly. *Psychological Bulletin, 98,* 67–83.

Crossman, E.R.F.W. (1958). A theory of the acquisition of speed-skill. *Ergonomics, 2,* 153–166.

Dehaene, S. & Changeux, J-P. (1989). A simple model of prefrontal cortex function in delayed response tasks. *Journal of Cognitive Neuroscience, 1,* 244–26.

Dehaene, S. & Changeux, J-P. (1991). The Wisconsin card sorting test: Theoretical analysis and modelling within a neural network. *Cerebral Cortex, 1,* 62–79.

Duncan, J. (1986). Consistent and varied training in the theory of automatic and controlled processing. *Cognition, 23,* 279–284.

Duncan, J. (1995). Attention, intelligence and the frontal lobes. In M.S.Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–731). Cambridge, MA: MIT Press.

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

Duncan, J., Williams, P., Nimmo-Smith, M.L., & Brown, I. (1993). In D.Meyer & S. Kornblum (Eds.), *Attention and Performance* (Vol. 14), Cambridge, MA: MIT Press.

Einstein, G.O. & McDaniel, M.A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology, Learning, Memory and Cognition, 16,* 717–726.

Eysenck, H.J. (1986). The theory of intelligence and the psychophysiology of cognition. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Flood, D.G. & Coleman, P.D. (1988). Neurone numbers and sizes in aging brain: Comparisons of human, monkey and rodent data. *Neurobiology of Aging, 9,* 453–463.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fuster, J. (1989). *The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe* (2nd ed.). New York: Raven Press.

Goldman-Rakic, P.S. (1987). Circuitry of primate prefrontal cortex and regulation of behaviour by representational knowledge. In F.Plum & V.Mountcastle (Eds), *Handbook of physiology* (Vol. 5; pp. 373–417). Bethesda, MD: American Physiological Society.

Hasher, L., Stoltzfus, E.R., Zacks, R.T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 163–169.

Heim, A.W. (1970). *The AH 4 Group Test of General Intelligence.* Windsor: NFER-Nelson.

Horn, J.L. (1982). The theory of fluid and crystallised intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F.I.M.Craik & S.Trehub (Eds.), *Aging and cognitive processes.* Boston: Plenum.

Horn, J.L. & Cattell, R.B. (1967). Age differences in fluid and crystallised intelligence. *Acta Psychologica, 26,* 107–129.

Jason, G.W. (1985). Manual sequence learning after focal cortical lesions. *Neuropsychologia, 23,* 483–496.

Jensen, A.R. (1980). Chronometric analysis of mental ability. *Journal of Social and Biological Structures, 3,* 181–224.

Jensen, A.R. (1985), The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioural and Brain Sciences, 8,* 193–219.

Jensen, A.R. (1987). Individual differences in the Hick paradigm. In P.A.Vernon (Ed.), *Speed of information processing and intelligence.* Norwood, NJ: Ablex.

Just, M.A. & Carpenter, P.A. (1992). A capacity theory for comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kimberg, D.Y. & Farrah, M.J. (1993). A unified account of impairments following frontal lobe damage: The role of working memory in complex organised behaviour. *Journal of Experimental Psychology: General, 122,* 411–428.

Kimura, D. (1977). Acquisition of a motor skill after left-hemisphere damage. *Brain, 100,* 527– 542.

Kyllonen, P.C. & Krystal, R.C. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence, 14,* 389–433.

Levine, D.S. & Prueitt, P.S. (1989). Modelling some effects of frontal lobe damage: Novelty and perseveration. *Neural Networks, 2,* 103–116.

Luria, A.R. & Tsvetskova, L.D. (1964). The programming of constructive ability in local brain injuries. *Neuropsychologia, 2,* 95–108.

Maylor, E.A. (1990). Age and prospective memory. *Quarterly Journal of Experimental Psychology, 42A,* 471–493.

McLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Milner, B. (1982). Some cognitive effects of frontal lobe lesions in man. *Philosophical Transactions of the Royal Society of London, B, 298,* 211–226.

Nettelbeck, T., Rabbitt, P.M.A., Wilson, C., & Batt, R. (1996). Uncoupling learning from initial recall: The relationship between speed and memory deficits in old age. *British Journal of Psychology, 87,* 593–608.

Newall, A. & Rosenblum, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R.Anderson (ed.), *Cognitive skills and their acquisition* (pp. 45–63). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Parashos, I.A. & Coffey, C.E. (1994). Anatomy of the aging brain. In J.M.Copeland, M.T. Abou-Saleh, & D.Blazer (Eds.), *Principles and practice of geriatric psychiatry* (pp. 36–50). Chichester: Wiley.

Rabbitt, P.M.A. (1959). Effects of independent variations in stimulus and response probability. *Nature,* April 12, p. 1212.

Rabbitt, P.M.A. (1965). An age decrement in the ability to ignore irrelevant information. *Journal of Gerontology, 20,* 233–238.

Rabbitt, P.M.A. (1968). Three kinds of error in signalling responses in a serial choice task. *Quarterly Journal of Experimental Psychology, 20,* 179–188.

Rabbitt, P.M.A. (1993). Does it all go together when it goes? *Quarterly Journal of Experimental Psychology, 46A,* 385–434.

Rabbitt, P.M.A. (1996). Intelligence is not just Mental Speed. *Journal of Biosocial Science, 28,* 425–449.

Rabbitt, P.M.A. & Yang, Qian (1996). What are the functional bases of individual differences in memory ability? In D.Herrmann, C.McEvoy, C.Hertzog, P.Hertel, &

M.K.Johnson (Eds.), *Basic and applied memory research: Theory in context* (Vol. 1; pp. 127–159). Mahwah, NJ: Lawrence Erlbaum Associates.

Reitan, R.M. & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4,* 161–198.

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

Salthouse, T.A. (1985). *A cognitive theory of aging.* Berlin: Springer-Verlag.

Salthouse, T.A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Salthouse, T.A. (1996). Predictors of cognitive aging: What can be learned from correlational research? Paper presented at the Cognitive Aging Conference, Atlanta, GA., April.

Schneider, W. (1993). Varieties of working memory as seen in biology and in connectionist/ control architectures. *Memory & Cognition, 21,* 184–192.

Schneider, W. (1994). The neurobiology of attention and automaticity. *Current Opinion in Neurobiology, 4,* 177–182.

Schneider, W. & Detweiler, M. (1988). The role of practice in dual task performance: Towards workload modelling in a connectionist/control architecture. *Human Factors, 30,* 539–566.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

Shallice, T. & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobe in cognitive estimation *Cortex, 14,* 294–303.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic information processing: II. Perception, learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.

Spearman, C. (1927). *The abilities of man.* London: Methuen.

Sternberg, R.J. & Powell, J.S. (1982). Theories of intelligence. In R.J.Sternberg (Ed.), *Handbook of human intelligence,* Cambridge: Cambridge University Press.

Walsh, K.W. (1978). *Neuropsychology: A clinical approach.* New York: Churchill Livingstone.

Weiskrantz, L. (1992). Introduction: Disassociated issues. In A.D.Milner & M.D.Rugg (Eds.), *The neuropsychology of consciousness* (pp. 1–10). London: Academic Press.

West, R.L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292.

Wilkins, A.J. & Shallice, T. (1987). Frontal lesions and sustained attention. *Neuropsychohgia, 25,* 359–365.

CHAPTER TWO
Cognitive Models of Ageing and Frontal Lobe Deficits

Christine Lowe and Patrick Rabbitt *Age and Cognitive Performance Research Centre, University of Manchester, U.K.*

INTRODUCTION

The main tasks for cognitive gerontology are to describe how mental abilities change with age and to relate these observations to the ageing of the brain and central nervous system. There has been much progress in identifying age-related changes in memory and in relating these to neurophysiology (Moscovitch & Winocur, 1992). In contrast, excellent descriptive studies of other cognitive abilities, such as problem solving, selective attention, and the ability to make rapid decisions, have not yet led to useful functional models of changes in cognitive processes or of the relationships of these changes to the ageing of the brain. This is partly because speculation has been dominated by the idea that age-related changes in cognitive skills are all alike, entrained by degradation of a single performance parameter, information-processing speed.

It is indisputable that as people grow old they become slower at performing tasks of any kind (Birren, 1956, 1974; Birren & Williams, 1980; Cerella, 1985, Salthouse, 1985, 1992, 1993). It is also clear that individuals who achieve higher scores on tests of general "fluid" intelligence also perform most laboratory tasks more rapidly and accurately (Cerella, Di Cara, Williams, & Bowles, 1986; Eysenck, 1986; Jensen, 1980, 1985, 1987; Vernon, 1983, 1985; Vernon & Jensen, 1984). All functional models for task performance must, of necessity, be based on and tested in terms of the only two performance indices that we can measure and compare in laboratory tasks: speed and accuracy. Thus the hypothesis that all individual differences in efficiency in all cognitive skills that are associated with age, or with intelligence, are solely determined by differences in one of the only two things that we can measure does not take us very far. It also confounds levels of description.

Measurements of changes in task performance with age have obvious practical importance but are theoretically unilluminating unless they can be interpreted in

terms of functional models. No functional models, whether they are based on the "box and arrow" information flow systems popular in the 1950s and 1960s (Broadbent, 1958) or on more recent production system models (Anderson, 1993) or connectionist networks (Dehaene & Changeux, 1989) or other hypothetical system architectures, attempt to account for the efficiency of all cognitive processes solely in terms of a single system performance parameter, information-processing speed. Most models accurately predict variations in the times taken to perform different tasks, but do this in terms of relationships between a variety of other hypothetical system performance parameters which, although they can be precisely quantified, cannot be sensibly expressed in units of time or of information-processing rate. For example, models framed in terms of production systems might account for differences in decision times in terms of changes in numbers, weightings and integrity of links; connectionist models might do so in terms of degrees of connectivity of units, of numbers of layers of units intervening between input and output or of levels of unit activation thresholds (see Rabbitt, 1996a, b). For the same reasons it is unhelpful to confuse differences in response speed measured in behavioural tasks with particular physiological characteristics of the central nervous system (CNS), such as rate of spread of activation in a neural network (Salthouse 1985) or synaptic delay, or "random neural noise" (Eysenck, 1986). It is possible that investigators have confounded levels of description in terms of measurable task performance indices, hypothetical functional system performance characteristics, and hypothetical neurophysiology because they have been misled by the apparently impressive results from two methodologies that have been developed to analyse decision times measured in laboratory tasks.

Cerella (1985) brilliantly analysed 140 published studies which had compared choice reaction times (CRTs) of younger and older adults across three or more levels of difficulty in each of a variety of different tasks. He used a graphical technique of data analysis introduced by Brinley (1965), plotting mean RTs for groups of older against corresponding mean RTs for groups of younger adults for each level of all tasks. He found that data points for all levels of difficulty were well fitted by the same straight lines which had very similar slopes for tasks of all kinds. That is, whatever the qualitative nature of the demands that tasks make and so, presumably, by whatever functional processes they are carried out, mean CRTs for older adults can be accurately estimated by multiplying mean decision times for young adults by the same, simple constant. This finding of similar proportional increases with age in mean decision times from all tasks has often been replicated. Cerella interpreted this finding in terms of a model for CRT "decomposition" developed by Sternberg (1969), who also discussed its use to interpret differences in decision times resulting from acute ingestion of alcohol and from individual differences in pathology (Sternberg, 1975). The basic assumption of "decomposition theory" is that the observed times taken for all

decisions that involve the perception of a signal and the production of a response to it represent the simple sums of the durations of linear series of successive and independent subprocesses. On these assumptions, by devising tasks that vary demands on one hypothetical subprocess, but not on any others, and by comparing RTs across these tasks, we can test whether conditions such as old age, or high intelligence, affect the efficiency of this particular subprocess more than others. In this framework of description Cerella's (1985) meta-analysis included data from a variety of tasks which, in terms of Sternberg's (1975) logic, make different levels of demands on the different independent, successive processes intervening between the perception of signals and the completion of responses to them. Cerella (1985) concluded that age slows all subprocesses, whatever their qualitative nature, to the same proportional extent.

Cerella's analysis has been extremely influential. However, it has the disappointing methodological corollary that we cannot hope to learn much about changes in functional processes in old age by comparing the response latencies of older and younger people on any task because, whatever the qualitative nature of the decisions that they make, the elderly will be slowed by the same proportional extent. Many investigators also drew the more radical, and logically quite unrelated, conclusion that "global slowing of information processing" across all functional processes not only provides a description of empirical observations but is also a *theoretically sufficient functional explanation* of all age-related changes in all mental abilities (Salthouse, 1985, 1991).

A second influential methodology which has been used to support the "global slowing" hypothesis is to compare older and younger adults on candidate tasks, such as memory tests, which may be scored in terms of accuracy rather than speed, and also to give them a variety of other tasks in which their decision times are measured. Hierarchical regression analyses are then made to discover whether any age-associated variance in memory test performance remains after individual differences in speed on the other tasks has been taken into consideration. Hertzog (1989), Salthouse (1985), Salthouse and Babcock (1991) and Schaie (1989) all report that, in a number of different cognitive tasks, age-related variance is significantly reduced or entirely eliminated after individual differences in speed have been partialled out. For example Salthouse (1993) reports that, after individual differences in speed have been taken into consideration, age accounts for less than 1% or 2% of variance in performance on many complex tasks. Commenting on similar results, Salthouse and Babcock (1991) conclude that age differences are not determined by the number or the relative complexity of the different functional operations that are necessary to perform a working memory task but only by the "simple speed with which even the most elementary operations are successfully executed."

These results must be interpreted cautiously. One obvious problem is that of "task purity" (Weiskrantz, 1992). Though performance on tasks as diverse as

"inspection time," "choice reaction time," "visual search" and "digit-letter substitution coding" can all, alike, be assessed in terms of "milliseconds per response", this obviously does not mean that they are all equally "pure" measures of information-processing speed and that they do not make demands on any other performance characteristic of the cognitive system. The details of published studies make it clear that age-related variance in memory tasks is only marginally reduced by partialling out variance in very simple tasks, such as four-choice CRT, and markedly declines only when regression equations incorporate data from much more complicated tasks which make demands on speed of learning or of memory retrieval of complex signal/response codes (Rabbitt & Yang, 1996).

Single-factor models tend to attribute age-related cognitive change to diffuse neurophysiological changes that affect the entire central nervous system (CNS) rather than to brain changes that occur earlier, and proceed faster in some "modular" subsystems than in others. They offer no way to accommodate the accumulating biological and neurophysiological evidence that, while ageing does bring about generalised loss of cortical tissue, this process is more marked in the frontal and temporal lobes than elsewhere in the brain (Petit, 1982; Stuss & Benson, 1986). For example Mittenberg, Seidenberg, O'Leary, & DiGiulio (1989) found that loss of neural tissue is most marked in the frontal lobe with 17% average loss between 20 and 80 years in comparison to less than 1 % in both occipital and temporal lobes. Haug et al. (1983) also reported similar disproportionate tissue loss in the brain regions that have strongest connections to the frontal lobes. Shaw et al. (1984) demonstrated selective frontal reduction in cerebral blood flow. The growing number of such reports has increased the demand for tests of the possibility that age-related cognitive changes reflect differentially greater impairments of functions supported by the frontal and temporal lobes of the brain (Whelihan & Lesher, 1985).

Until recently, models for "modular" ageing of the cognitive system have mainly discussed temporal lobe changes because damage to the temporal lobes and associated structures and pathways often results in the amnesic syndrome, and this encourages speculation that evident losses of memory efficiency in old age may be linked to selectively greater temporal lobe attrition (Mayes, 1988; Mayes & Daum, this book; Parkin & Walter, 1992; Robbins et al., 1994). It has also been suggested that relatively earlier and faster "frontal lobe ageing" may result in correspondingly marked changes in "central executive" functions in later life (Della Sala, Logie, & Spinnler, 1992). Parallels between behavioural changes associated with frontal lesions and with normal ageing have long been discussed at an anecdotal level; Veroff (1980) pointed out that "there is a striking similarity between the functional deficits seen in frontal lobe syndrome and those seen in normal ageing." He drew parallels between the characteristic symptoms of frontal lobe damage and reports of the increase, in old age, of distractibility,

rigidity and slowing of thought processes (see Miller, 1977) with accompanying loss of the ability to think abstractly (see Thaler, 1959), to think flexibly (see Botwinick, 1977) and to process complex novel material (see Cattell, 1943). Mittenberg et al. (1989) examined the performance of older people on a battery of neuropsychological tests using principal components analyses and concluded that frontal deficits were the primary component of the age-related cognitive impairments.

Objective attempts to examine performance by elderly participants both on batteries of "frontal" or "executive" tests and on single tasks have mainly been guided by the speculation that old age impairs the ability to suppress irrelevant information, or to respond to some parts of a complex display while inhibiting responses to others (e.g. Hartman & Hasher, 1991; Hasher, Stolfus, Zacks, & Rypma, 1991; West, 1996). Hartley and Teal (1994) suggest that older adults perform relatively poorly on tests sensitive to frontal lobe damage even after variance associated with declines in speed and memory ability has been partialled out. However evidence from studies using similar test batteries has been mixed. Robbins et al. (1994; and this book) report principal components analyses of data from a battery of neuropsychological tests, and on measures of information-processing speed and of intelligence obtained from large groups of normal elderly. They find some separation of "frontal" measures, from each other and from "temporal" measures, and some association between these measures and participants' ages but, as Robbins remarks, "the results from these analyses do not strongly support the hypothesis that functions sensitive to frontal lobe damage are especially sensitive to ageing effects."

When interpreting such findings at least four methodological issues are crucial. The first is that, as we have seen, when tasks are scored in terms of decision times or estimates of information-processing rate, it is necessary directly to test, and discount the strong version of the hypothesis that authors such as Salthouse (1985, 1992) have derived from Cerella's (1985) analysis: whatever the qualitative task demands may be, age slows times for all decisions in exactly the same proportion to their absolute durations, and so their presumed difficulty. Thus, if "frontal" tasks are relatively difficult and slow, the absolute amounts of age-related slowing observed in these tasks will be more marked than in other easier and faster tasks that do not involve the frontal lobes. The second methodological issue is that, in order to be sure that a putatively "frontal" or "executive" task picks up age-related changes that are not simply consequences of "global slowing," it is necessary to examine variance associated with age after differences in speed have been taken into consideration. A third methodological difficulty, trenchantly pointed out by Burgess (this book), has been that the test/retest reliability of "frontal" or "executive" tasks is often unacceptably low because tasks can only be valid tests of executive function when they are novel, and "a task can only be novel once." This means that age changes in frontal

function can only be detected on unpractised tasks. A final methodological problem raised by Rabbitt (this book) is that terms such as "inhibition," which have been used as convenient labels for hypothetically distinct aspects of "executive" function, have uncertain construct validity. It has been convenient to borrow such terms from everyday language as labels to describe task demands, but this gives us no assurance that they in any way reflect the underlying functional processes by means of which these demands are met. Computer simulations of several "frontal" tasks have shown that identical production system architectures can be used to meet very disparate task demands. In particular, Kimberg and Farrah (1993) have formally shown that identical production system architectures can be used to program computer simulations of tasks that have been supposed to make entirely different and characteristic demands on the human cognitive system such as the planning and execution of sequences of responses, the "conflict" condition of the "Stroop test," the "concept shifting" condition in the Wisconsin card-sorting task, and failures of "contextual memory". This illustrates the possibility that, although words such as "inhibition" or "concept shifting" are useful descriptive terms for different types of *task demands,* this does not necessarily imply any difference in underlying processes. Interestingly, there is as yet no good empirical evidence that individuals who perform particularly poorly on one test of "inhibition" or "concept shifting" also perform poorly on other tasks that make identical, or logically similar, demands. Indeed, as Burgess (this book) points out, patients with frontal damage who perform poorly on some frontal tasks may nevertheless perform quite normally on many others. It is important to note that constructs such as "inhibition" have not been validated experimentally, and that to do this requires formal tests whether individuals who fail to inhibit responses to irrelevant or intrusive stimuli in one task also have greater than average difficulty when other, different, tasks make this particular demand.

To explore how these methodological difficulties might be overcome it was necessary to compare large groups of more and less elderly people on a battery that includes some tasks which have been considered to be tests of frontal or executive function and others that, while they are equally difficult and slower to perform, are not considered to involve "frontal" or "executive" processes.

EXPERIMENT 1

This was made possible by work carried out to evaluate a battery of cognitive tests to detect possible declines in cognitive function following anaesthesia in elderly surgical patients. For this purpose it was necessary to test a large and diverse elderly population, to include both "executive" and "non-executive" tasks, and to give volunteers all tasks twice in order to determine test/retest reliability.

Description of Test Battery

Tasks have been described in detail by Rabbitt (1996a).

Task 1: Four-Choice Reaction Time (4-choice RT). Choice reaction time (CRT) has been considered to be an exceptionally "pure" index of information-processing speed (e.g. Jensen, 1987; Salthouse, 1985). Volunteers responded as fast as possible using the appropriate one of a row of four keys on the keyboard of a laptop computer to 2cm black circles that appeared, one at a time, in random order, within four illuminated square "boxes" drawn in a horizontal row across the centre of the computer monitor screen. They completed runs of 52 responses.

Task 2: "One-back" 4-choice RT. This was identical to task 1 except that volunteers did not respond to the signal currently on the display but to its immediate predecessor. "Executive" features of this task are the continuous updating of working memory necessary to respond to previous signals and the suppression of responses to the current signal in order to respond to its predecessor.

Tasks 3 through 7: Memory Search. Recent reviews have suggested that working memory efficiency is particularly sensitive to cognitive changes in old age (Salthouse, 1992). Accordingly, we included a visual search task with an incrementing working memory component. Volunteers scanned printed sheets of 120 capital letters printed in random order in 12-point Times Roman in 12 regular horizontal lines across an A4 page in order to detect and cross out 15 target letters which occurred at random positions among 108 background letters. Scanning time for single pages was timed by a stopwatch. In successive conditions targets were fixed sets of 1, 2, 3, 4, or 5 different symbols, and background items were drawn from the remaining 21 letters of the Roman alphabet. Comparisons between these conditions allowed information-processing speed to be examined as a function of increasing working memory load.

Task 8: Substitution Coding. Many studies have shown that scores on substitution coding tasks, as used in the WAIS subtests and standardised for elderly populations, show markedly high correlations with scores on intelligence tests (Rabbitt, 1993) and also with chronological age (Salthouse, 1985). Recent work shows that substitution coding tasks do not merely measure individual differences in coding speed and accuracy but also individual differences in working memory efficiency (Piccinin, Hofer, & Rabbitt, submitted, 1995; Rabbitt & Han, unpublished manuscript). The volunteers' task was to "encode" as many digits as possible within a fixed period of 2m by looking up the appropriate codes and entering them in the corresponding cells. The performance index was the derived average time taken to encode each item.

Tasks 9, 10, and 11: The "Stroop" Test. Hasher, Stolfuss, Zacks, and Rypma (1991) have argued that, in addition to losses in efficiency of memory and information-processing speed, cognitive ageing is accompanied by a progressive and specific loss of the ability to inhibit unwanted responses or to ignore

irrelevant or competing signals. The Stroop test is the prototypical paradigm for examining individual differences in susceptibility to interference. In the baseline condition, Stroop 1, volunteers were timed at reading aloud, as fast and as accurately as possible, lists of 40 colour names printed in black type. In a second baseline condition, Stroop 2, they were timed at naming aloud 40 coloured patches. In Stroop 3, the "interference" condition, they were timed at naming aloud the colours in which the names of 40 other, conflicting, colours were printed. In all conditions errors were identified by checking responses against printed checklists and reading times were measured with a stopwatch. Performance indices were mean reading time per item and numbers of errors in each condition.

Tasks 12, 13, 14, 15, and 16: Concept Shifting. Maylor, Godden, Ward, Robertson and Baddeley (in press) report that older people have particular difficulty with speeded tasks in which they must alternately discriminate between members of different classes of stimuli, such as the colours and shapes of symbols. A category shift task (CS) was, therefore, included in the battery. In all conditions participants were given A4 sheets on which 16 5mm diameter circles were spaced at equal intervals around a 6cm diameter circular perimeter. In the first condition, CSA, 12 of the circles were inset in random order with the numbers 1 to 12 and volunteers crossed these out, in ascending order (i.e. 1, 2, 3, 4, 5, etc.). In the second condition, CSB, the circles were randomly inset with the 12 letters, A through L, which were crossed out in ascending order (i.e. A, B, C, D, etc.). In the fourth, "switching" condition, CSC, circles were randomly inset with both letters and digits and volunteers crossed out the six letters A through F and the digits 1 through 6 in an alternating ascending series of digits and letters (i.e. 1, A, 2, B, 3, C, 4, D, 5…etc.). In addition, to provide a baseline measure of speed on this task, volunteers were also asked to cross out blank circles in predetermined order as rapidly as possible. They were given two trials, CS1 and CS2. For each subtest experimenters noted errors as they occurred and timed performance with a stopwatch. Performance indices were errors and average times taken to cross out all target items.

Task 17: The Baddeley Semantic Reasoning Task. Volunteers were given printed pages of propositions which they had to consider, and validate as true or false, by crossing out T or F after each. Equal numbers of propositions are "true" and "false" and are framed in the active, active negative, passive and passive negative modes as illustrated here: "a follows b: ab (T/F) (answer T); b does not follow a: ab (T/F) (answer F); a is followed by b: ab (answer T)" and "a is not followed by b: ab (answer F)." Scores were numbers of errors and numbers of propositions correctly completed in two minutes.

Measure of General Intellectual Ability. To provide normative data stratified by level of ability, to validate tests against some well-standardised criterion other than age, and to examine age-related variance which might remain after variance

associated with differences in general ability had been taken into account, all volunteers were also scored on a well-standardised measure of general ability: all four parts of the Cattell and Cattell (1960) "culture fair" intelligence test (CCF).

All tests were taken by 56 men and 67 women aged from 60 to 83 years (mean age=68.08, *SD*=5.46), individually in a quiet room during two 90-min sessions one week apart. They were each paid an honorarium of £4 for each session.

Results

The battery included tasks of very different levels of difficulty and, as a result, mean times for individual decisions varied from 256ms to 1800ms. This allowed a reasonable test of Cerella's (1985) suggestion that age slows decisions for all tasks by the same proportion. Four of the tasks, one-back 4-CRT, Stroop, concept shifting, and memory search, may intuitively be considered as making demands on "executive" functions, or on "working memory" or on both, and so to be especially sensitive to the cognitive changes that occur in old age. This let us test the hypothesis that the effects of age on these tasks would be disproportionately greater than on others.

Since all tasks were scored in terms of decision speed, and since all volunteers were also scored on the Cattell and Cattell (1960) "culture fair" test, it was also possible to discover whether age-related variance in scores on frontal tasks remains after variance in speed on other tasks had been taken into consideration.

Volunteers completed the battery twice, with an intervening interval of two weeks plus or minus five days. This allowed us to test whether any age effects observed when "frontal" tasks were novel on trial 1 were reduced on trial 2. Finally, by comparing volunteers' decision times across different "frontal" and "executive" tasks, it was possible to test whether individuals who were exceptionally slow or inaccurate on one class of "frontal" or "executive" task also performed relatively poorly on others.: in particular, whether failures of inhibition in the conflict condition of the Stroop test (Stroop 3) could be shown to have operational construct validity inasmuch as they predicted failures of concept shifting or unusual difficulty in suppressing responses to currently displayed signals in the one back 4-CRT task.

1. Does Age Affect Decision Times on "Frontal" or "Executive" Tasks Proportionately more than on Others? Rabbitt (1996a) pointed out that the methodology used by Cerella and others offers only very weak tests of the hypothesis that age increases decision times on all tasks by the same proportions because least squares fits for "Brinley functions" lack the statistical power necessary to make confident comparisons between groups. The present data once again clearly illustrate this. Mean decision times for each task were obtained for volunteers aged from 61 through 70 years, and for those aged from 71 to 83 years. This allowed calculation, for each task, of what might be called the

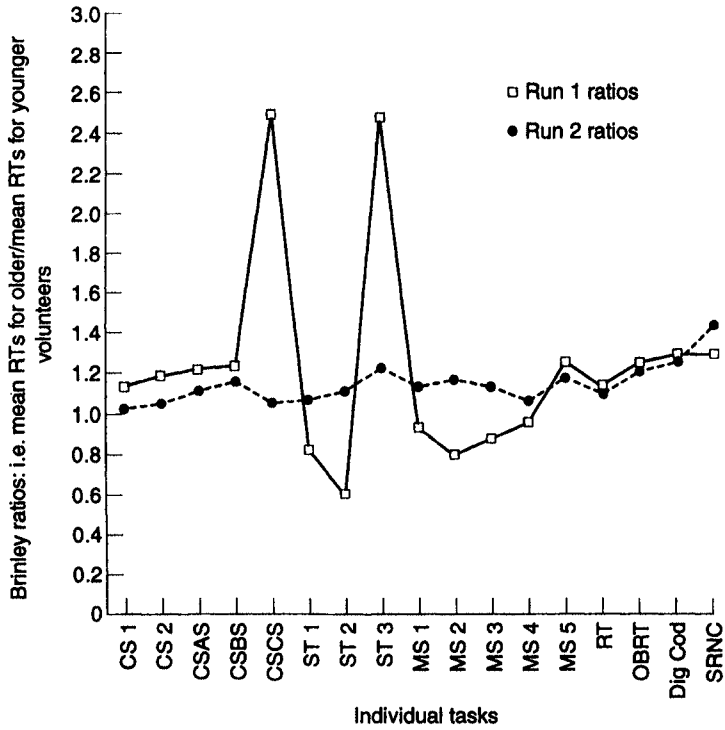


FIG. 2.1. Brinley ratios of mean reaction times for older and younger volunteers across all tasks, for first and second trials.

"Brinley ratio" of decision times of the older to decision times of the younger individuals. Brinley ratios for data from the first trials on each task varied from 0. 614 (i.e. mean decision times for older were 38.6% less than those for younger volunteers) to 2.496 (i.e. decision times for older were 149.6% higher than those for younger volunteers). Figure 2.1 shows the Brinley ratios of mean RTs for older/mean RTs for younger volunteers across all tasks for both their first and their second runs.

During Session 1 Brinley ratios for most tasks were very similar and, just as in Cerella's (1985) meta-analysis, fell between the limits of 1.13 and 1.3. However, for all conditions of the visual search tasks (MS) ratios fell below 1.0, indicating that the older volunteers, paradoxically, scanned faster than the young. Examination of errors showed that in run 1, but not in run 2, older individuals maintained search speed at the cost of missing many targets. This had the effect that, as the number of targets increased from 1 to 5, while younger individuals searched more slowly without loss of accuracy older individuals maintained the same search speed but traded this for accuracy and made increasing numbers of

errors. Exceptionally high ratios of 2.49 were obtained for the critical "conflict" condition of the Stroop task (Stroop 3) and for the critical "switching" condition of the concept shifting task (CS3). These values do not seem to reflect the fact that older and younger volunteers made different choices in trading off speed against accuracy, since both age groups made the same numbers of errors in these conditions. However, in spite of these large discrepancies between Brinley ratios for individual tasks, a "Brinley function" plotted across all data points for mean old RT vs. mean young RT was excellently fitted (R^2=.97) by a straight line of slope 1.24, just as in Cerella's (1985) classic analysis.

This shows that impressive least squares fits may divert attention from the fact that the proportionate slowing due to age is very different in different tasks. In this experiment the Stroop 3 and the CS3 tasks were especially sensitive to age-related changes. This replicates findings by Rabbitt (1996a) for the same tasks, when means for groups with high scores were plotted against means for groups with low scores on the Cattell and Cattell (1960) culture fair intelligence test. In this previous analysis, as in the present exercise, although the "Brinley function" for high vs.low Cattell scores was excellently fitted by a straight line, ratios for individual tasks differed from each other by as much as 30% or 40% of baseline speed. Rabbitt (1996a) found that, when Brinley ratios for a first and second run were compared, they correlated significantly at $r > .7$. In other words, Brinley ratios for some tasks are consistently higher than for others. Rabbitt concluded that not all tasks are equally sensitive to individual differences in general ability and that the appearance of simple scalar effects in Brinley plots can be very misleading. Figure 2.1 shows that the unusually high Brinley ratios for Stroop 3 and for CS 3 disappear after even a single session of practice. Closer inspection also showed that, during session 2, older volunteers reduced their errors on all conditions of the MS task to comparable levels to younger individuals and that, as a result, their scanning times increased with the number of different targets for which they had to search and became slower than those of the young in all MS conditions so that Brinley ratios for all conditions were now above 1. Because practice brought about such marked qualitative changes in performance on MS tasks and on Stroop 3 and CS3, we felt justified in omitting these tasks when we compared the consistency of Brinley ratios between runs 1 and 2. The rank order of Brinley ratios across all other tasks correlates at R=.698, p=0.021. In short, contrary to impressions from analyses such as Cerella (1985), old age slows performance on some tasks consistently more than on others.

Besides illustrating the limitations of the Brinley function methodology these data suggest that old age may, indeed, affect performance on two prototypical "frontal" tasks more than on other "non-frontal" tasks. It also directly illustrates the point made by Burgess and others in this book that "frontal" tasks may often work only once. Even with a single practice run and a gap of 1 to 2 weeks

between first and second tests, practice abolishes the age differential in the Stroop and in the concept shifting tasks.

Since there appears to be no reliable way of testing whether the Brinley ratios for the conflict condition of the Stroop test (Stroop 3) and the switching condition of the concept shifting task (CS3) are significantly greater than the Brinley ratios for other tasks, the question arises whether these tasks actually do pick up cognitive changes in old age other than those that reflect general slowing. One approach is to use the second methodology, developed by Hertzog (1989), Salthouse (1985, 1990, 1991, 1993), and Schaie (1989), and used by Maylor et al. (in press) to test whether any age-related variance in task performance remains after variance associated with individual differences in processing speed has been partialled out. Duncan (1995) and Duncan, Burgess, and Emslie (1995) report that scores on the Cattell "culture fair" intelligence test account for all individual differences in performance on some "executive" tasks. Maylor et al. (in press) took this into account and examined age effects after partialling out both variance in indices of speed and in culture fair test scores in the conflict condition of the Stroop test and in an executive switching task. They found that significant age-related variance still remained after this had been done. They concluded that age changes in "executive" function are not entirely explained either by global slowing of information processing or by age-related changes in general fluid intelligence.

We repeated this analysis with data from all tasks and both runs. For the Stroop 3 and CS3 tasks, the most appropriate indices of information-processing speed seemed to be times for their easy "non-executive" conditions (Stroop 1 and 2 and CS 1 and 2). On the first trial, speed on the two easy conditions of the CS task and culture fair scores each accounted for significant proportions of variance between individuals on the more difficult "switching" version of the task, i.e. CS3. After this variance had been taken into consideration, age also predicted an additional small (6%) but significant proportion of individual variance in performance. Similarly for the conflict condition of the Stroop (Stroop 3), age still accounted for a modest (9%) but significant proportion of variance even after significant contributions made by individual differences in speed, measured in Stroop 1 and 2 and by culture fair test scores, had been partialled out. On the second trial, speed and culture fair scores still remained significant predictors of CS3 scores, but age made no additional contribution. The same was true for Stroop 3:

It seems that when the concept shifting and Stroop tasks were first encountered they both picked up age-associated variance in performance that is independent of, and additional to, differences associated with concomitant declines in information-processing speed, and in general intellectual ability as assessed by Cattell scores. However, on the second trial age-related changes in

performance were completely accounted for by slowing of information-processing speed and concomitant decline in fluid intelligence.

Equivalent analyses were carried out for all other tests in the battery. In all cases, both on the initial and on the second trials, age made no further contribution after variance between individuals associated with differences in information-processing speed and in culture fair test scores had been partialled out. Note that decision times in the one-back CRT task and in the substitution coding task were much longer than in the conflict condition of the Stroop or in the switching condition of the concept shifting task. On this criterion, Stroop 3 and CS3 were not the most "difficult" tests in the battery. We conclude that the age sensitivity of the Stroop and concept shifting tasks relates to the particular functional processes they involve rather than simply to their intrinsic difficulty, or the time that they take.

It seems that both Stroop 3 and concept shifting 3 pick up age-related changes in cognitive performance, but the question remains whether or not this is because they both involve the same functional processes that are affected by old age. Evidence for this would be that individuals who were markedly slowed on Stroop 3 were also markedly slower on CS3. Because some individuals are faster than others on all tasks, a simple comparison between rank orders of decision times is uninformative. What is necessary is a comparison between the amount of slowing, with respect to baseline performance, in the Stroop 3 and the CS3 tasks. However, although age-related slowing is greater for some tasks than for others, the scaling effect noted by Cerella (1985) also operates. Individuals' decision times are slowed in proportion to their durations, so that those who were slower on baseline conditions of the Stroop and CS task, Stroop 1 and 2 and CS 1 and 2, would also be proportionally slowed on the critical conditions of Stroop 3 and concept shifting 3. Individuals who were slower on all tasks would, therefore, necessarily also show a greater absolute difference between baseline and critical conditions on the Stroop and concept shifting tasks. Because of this the hypothesis of consistent individual differences in slowing in critical Stroop and CS conditions must be tested by comparing proportionate, rather than absolute, slowing relative to baseline. Percentage slowing from baseline was computed for each individual for Stroop 3 and CS 3. The rank order correlation between percentage slowing in Stroop 3 and CS3 was not significant (r=0.14). There is no evidence that individuals who are markedly slowed on one of these tasks are also markedly slowed on the other, and there is no reason to suppose that selective slowing on the conflict condition of the Stroop and the switching condition of the concept shifting task are mediated by age-related changes in the same functional processes.

EXPERIMENT 2

This study was carried out as a further test of Maylor et al.'s observation (in press) that older individuals are especially inconvenienced by tasks that demand rapid switching of attention between different sets of stimuli. A group of 90 volunteers, aged from 60 to 82 years (mean=71, *SD*=5.5) were tested on six conditions of an "attention switching" task modelled on that used by Maylor et al. Each person was given a printed booklet which contained the six test conditions and was told that they would be given one minute to search for every occurrence of each of two items, which would be described anew for each condition. Participants were explicitly asked to order their search so that they always scanned from left to right and line by line.

On each sheet were printed rows of regular outline geometrical shapes (triangles, squares, circles, etc.) and each shape had a number (from 1–9) printed within it. The subtasks represented two conditions: "simple," dual-item/single-element search and "difficult," dual-item/dual-element search. In the simple conditions volunteers crossed out *either* every incidence of the number 2 and of the number 7, *or* every incidence of a circle and a square.

The difficult versions of this task included four counterbalanced subconditions in which volunteers had to search for both a particular digit and a particular shape: (1) crossing out every instance of the digit 7 and every square; (2) crossing out every instance of the digit 2 and every circle; (3) crossing out every instance of the digit 7 and every circle; and (4) crossing out every instance of the digit 2 and every square. There was a brief interval of approximately one minute between each search condition.

Results

An initial multiple regression analysis examined age and speed (latency to locate a target) on the simple/single-element condition as predictor variables for performance on the difficult/dual-element conditions. (Speed scores for both the simple conditions and the four difficult conditions were highly correlated, so RTs for first and third conditions were used in the analysis.) Both simple latency and age were significant predictors ($F(3,89)=24.246$, $p<.0001$) and the model accounted for 36% of the variance in task performance. A further regression analysis included performance scores on a measure of fluid intelligence, the Heim (1970) AH 4 test, as a covariate. When variance associated with AH 4 test scores was entered into the model, variance associated with age was not wholly accounted for and remained a significant predictor of task performance ($p<.03$).

The results replicate those reported by Maylor et al. (in press) and also those reported for Experiment 1 in showing that the entire age-related decline in performance at a particular "executive" task cannot be entirely accounted for by

concomitant declines in information-processing speed, or even in fluid intelligence (Spearman's g_f).

A further analysis was made to discover whether individuals who were markedly slowed on one of the "difficult" switching conditions of this task were also markedly slowed on others, that is, to discover whether "switching" had demonstrable validity as a construct identifying a particular kind of functional operation necessary to carry out a particular kind of task demand that might be selectively disabled in some individuals but not in others. To do this, proportional slowing on the switching conditions relative to the non-switching conditions was calculated, as described for Experiment 1, and compared across individuals. The rank order correlation was not significant ($r<0.1$) so, once again, the results provide no evidence that "switching" of attention between different categories of symbols involves a functional mechanism that is selectively damaged in some individuals, causing them to be disabled on all tasks of this kind.

Discussion

These two studies illustrate some methodological issues that complicate attempts to test how age-related changes in cognitive performance are related to ageing of the brain. One possibility is that brain changes are diffuse and "global" and so affect all mental abilities to the same proportional extent. An associated idea, which is by no means a logically necessary entailment, is that satisfactory functional models of cognitive differences associated with individual differences in age or in general intelligence can be based on the premise that a single performance characteristic of the human cognitive system, such as "information-processing speed", underlies, and entails, changes in all cognitive skills. This latter proposal has been based on two lines of evidence: The first is that, because greater age (and lower intelligence) slow times for all kinds of decisions to the same proportional extent, there is no evidence that they affect some functional processes more than others (Cerella, 1985). The second line of evidence is that, on all tasks, no age-related differences in performance are observed after variances associated with individual differences in information-processing speed have been taken into consideration.

The conclusion that old age brings about the same proportional slowing of performance in all decisions, whatever their nature, has been based on the graphical methodology of "Brinley plots", as used by Cerella (1985). Experiment 1 replicates an earlier demonstration by Rabbitt (1996a) that this methodology is unsatisfactory because excellent least squares fits to linear "Brinley functions" may be obtained even when, in both absolute and proportional terms, age-related slowing is markedly and consistently greater on some tasks than on others. This means that measurements of differences in decision times can be used to test whether age affects performance on some tasks more than on others. Since speed and errors are the main performance indices

that we can measure in laboratory experiments, this is methodologically liberating because it opens the way for decision times on tasks to be used to test whether old age affects the efficiency of some decisions, and so of the functional processes underlying these decisions, more than others. The findings are also theoretically interesting because disproportionately great age-related slowing (unusually high "Brinley ratios") was observed only on tasks that have been considered as tests of "frontal" function, but not on others. This is consistent with speculations that because age-related neurophysiological changes are more marked, and occur earlier in the frontal lobes than elsewhere in the brain, "frontal" tasks are exceptionally sensitive to age-related changes in cognitive performance.

Another idea that has inhibited investigations of "local" as distinct from "global" age-related changes has been that, on any given task, age-related differences disappear once differences in information-processing rate have been taken into consideration. Both Experiments 1 and 2 confirm that this is not always the case. On the critical conditions of the "Stroop test," a "concept shifting" task, and a version of the Maylor et al. (unpublished manuscript) attention shifting task the effects of age remained significant even after variance associated with individual differences in speed on baseline conditions had been partialled out. Once again these results are consistent with the idea that old age brings about relatively early and marked changes in "frontally mediated" tasks.

In a recent experiment, Duncan (1995) found that differences in "frontal" tasks are well predicted by differences in g_f, as assessed by scores on the Cattell and Cattell "culture fair" intelligence test. He made the entirely reasonable suggestion that measures of g_f may be better indices of integrity of frontal function than many of the other tests that we have. This may be true, but the results of Maylor et al. (in press) and those described above show that, though it may indeed be a good measure of frontal function, the Cattell culture fair test certainly does not detect *all* individual variance in performance between individuals on the Stroop test, on the concept switching task, or on other tasks that require rapid switching of attention between different categories of stimuli. Duncan's (1985) suggestion that g_f, the single factor that appears to account entirely for the common variance between individuals' scores in intelligence tests, may account for individual differences in frontal function is, of course, interesting. In principle, it is very likely that efficient and sensitive test batteries for frontal function could be devised. In this context, Duncan's (1995) proposal makes the reasonable point that the best such batteries that are currently available are known as "intelligence tests" and that the Cattell and Cattell (1960) culture fair test is one of the most sensitive of these. We might argue that the culture fair and other intelligence tests work well as tests of frontal function because many of the great variety of problems that they incorporate make specific task demands on functional processes that take place in the frontal lobes, such as

"inhibition," "planning," "attention switching," or "goal maintenance." In this framework of description, scores on the Cattell and other intelligence tests sensitively reflect individual differences in frontal function because they are, in effect, aggregates of scores on a variety of different tests of "frontal" function. If we accept this hypothesis, we may suppose that it might be possible to discover different subsets of problems that, while collectively serving as effective intelligence tests because they load heavily on measures of g_f, nevertheless also assess competence on distinct, specifiable, functions such as "inhibition," "switching," etc.

Duncan (1995) does not seem to take this view, but seems rather to accept Spearman's (1927) suggestion that scores on intelligence tests cannot sensibly be statistically "decomposed" or "fractionated" to yield estimates of levels of different subvarieties of "intelligence" and so are best described in terms of a single common factor of general fluid ability, g_f. In other words further "decomposition" of frontal function is not sensible, and $g_{f(fluid)}$ is equivalent to $g_{f(frontal)}$.

This point of view would become less plausible if there were evidence that concepts such as "inhibition," or "attention switching" have validity as independent functional constructs, that is, if it were found that individuals who are poor at one version of a task involving inhibition or switching were also poor at other tasks that make the same demands, but not necessarily on other tasks—in other words, if we could show the "double disassociations" that, until now, have provided the methodological underpinning of cognitive neuropsychology. The evidence would be still more compelling if individuals who were disabled on all versions of the Stroop test were not, necessarily, poor at all versions of switching tasks, and vice versa. A still more compelling finding would be that predictions of difficulty on one subclass of tests to difficulty on another remained significant even after individual differences in scores on intelligence tests, such as the Cattell culture fair, had been taken into consideration. This would mean that there were specific deficits in function that are associated with frontal impairment, but which could not be completely accounted for in terms of individual differences in scores on a good measure of Spearman's (1927) single, general factor, g_f.

Both Experiment 1 and Experiment 2 tested these possibilities, but with inconclusive results. In Experiment 1 both the Stroop 3 and CS3 tasks were shown to be age-sensitive, and age-related differences in performance on these tasks could not entirely be accounted for by individual differences in speed or in culture fair test scores. However, difficulties with the conflict condition of the Stroop test (Stroop 3) did not predict difficulties with the switching condition of the concept shifting task (CS3). This leaves open the possibility that "inhibition," as assessed by Stroop 3, and "attention switching," as assessed by CS3, are both valid, but functionally distinct, constructs. In Experiment 2 a particular class of

attention switching tasks was also found to pick up age-related variance in performance even after concomitant variance in information-processing speed and in culture fair scores had been partialled out. However, the amount of difficulty that individuals experienced in the critical condition of any one of these switching tasks did not predict the amount of difficulty that they experienced on any other. There was no evidence that "attention switching" is a functional construct that can be validated across a range of different tasks. Other series of experiments carried out in our laboratory also fail to find that differences in levels of competence at coping with interference in some versions of the Stroop test predict levels of competence in others. To this extent our attempts to show that frontal function can be "fractionated" in terms of classes of independent functional operations have failed.

The main contribution of these studies is, therefore, to overcome practical objections to methodologies that have been thought to be unsuitable, or ineffective, for investigating age differences in performance and to show that, if these methodologies are sensibly used, age-related cognitive changes are significantly greater in some tasks than in others. The tasks that we have found to be "age-sensitive" are also ones that have often been used to detect impairments in frontal function. We have shown that age-related changes in performance on these "frontal" tasks are not removed when individual differences in scores on the culture fair intelligence test, and so in g_f, are taken into consideration. Thus while the culture fair intelligence test may indeed, as Duncan (1995) suggests, be an effective, it is not a perfect index of differences in performance on frontal tasks. We are, of course, accustomed to living with the idea that in terms of measurements of cognitive performance it is likely that no indices can be perfect.

As a secondary contribution these results also show that, once methodological difficulties have been understood and taken into consideration, it is possible to use systematic studies of performance of large groups of individuals on a variety of different "frontal" tasks to discover whether "frontal" performance can be "fractionated" in terms of subgroups of specifiable cognitive functions, which can be ostensively defined in terms of particular classes of task demands. Ideal demonstrations would show both that these functions had good construct validity in the sense that performance on a particular task's demands was consistent, while performance on different task demands was relatively independent across a variety of different situations. Thus these experiments show us ways to raise, but do not yet resolve, the question whether individual differences in competence at "frontal" tasks and changes in this competence in old age can be well described in terms of a plurality of different indices or, as Duncan (1995) and Duncan, Burgess, and Emslie (1995) would suggest, only in terms of a single, "aggregated" index, such as g_f.

REFERENCES

Anderson, J.R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Birren, J.E. (1956). The significance of age changes in speed of perception and psychomotor skills. In J.E.Anderson (Ed.), *Psychological aspects of aging.* Washington DC: American Psychological Association.

Birren, J.E. (1974). Psychophysiology and speed of response. *American Psychologist, 29,* 808– 815.

Birren, J.E. & Williams, M.V. (1980). Cognitive issues: Speed of behaviour. In L.W.Poon (Ed.), *Aging in the 1980's: Psychological issues.* Washington DC: American Psychological Association.

Botwinick, J. (1977). Intellectual abilities. In J.E.Binnen & K.W.Schaie (Eds.), *Handbook of psychology of ageing* (pp. 580–605). New York: Van Nostrand Reinhold.

Brinley, J.F. (1965). Cognitive sets, speed and accuracy in the elderly. In A.T.Welford & J.E. Birren (Eds.), *Behaviour aging and the nervous system.* New York: Springer-Verlag.

Broadbent, D.E. (1958). *Perception and communication.* Oxford: Pergamon Press.

Cattell, R.B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 3,* 153–193.

Cattell, R.B. & Cattell, A.K.S. (1960). *The individual or group Culture Fair Intelligence Test.* Champaign, Illinois: IPAT.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98,* 67–83.

Cerella, J., Di Cara, R.E., Williams, D., & Bowles, N. (1986). Relations between information processing and intelligence in elderly adults. *Intelligence, 10,* 75–91.

Dehaene, S. & Changeux, J-P. (1989). A simple model of prefrontal cortex function in delayed response tasks. *Journal of Cognitive Neuroscience, 1,* 244–26.

Della Sala, S., Logie, R.H., & Spinnler, H. (1992). Is primary memory deficit of Alzheimer patients due to a Central Executive impairment? *Journal of Neurolinguistics, 7,* 325–346.

Duncan, J. (1995). Attention, intelligence and the frontal lobes. In M.S.Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge: MIT Press.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia,* 35(3), 261–268.

Eysenck, H.J. (1986). The theory of intelligence and the psychophysiology of cognition. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 3.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Hartley, A.A. & Teal, J.V. (1994). It's what's up front that counts; speed, memory and attention as predictors of age related variance on neuropsychological tests sensitive to frontal lobe damage. Unpublished conference abstract.

Hartman, M.S. & Hasher, L.S. (1991). Aging and suppression of previously relevant information, *Psychology and Aging, 6,* 587–594.

Hasher, L.S., Stolfuss, E.R., Zacks, R.T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17(1),* 163–169.

Haug, H., Barmwater, U.J., Eggers, R., Fischer, D., Kohl, S., & Sasi, N.L. (1983). Anatomical change is aging brain: Morphometric analysis of human prosecephalon. In S.CervosNavarro & H.I.Sarkander (Eds.), *Neuropharmacology* (vol. 21, pp. 1–12). New York: Raven Press.

Heim, A.W. (1970). *The AH 4 Group Test of General Intelligence.* Windsor: NFER-Nelson.

Hertzog, C. (1989). Influences of cognitive slowing on age differences in intelligence. *Developmental Psychology, 25(4)*, 636–651.

Jensen, A.R. (1980). Chronometric analysis of mental ability. *Journal of Social and Biological Structures, 3,* 181–224.

Jensen, A.R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioural and Brain Sciences, 8,* 193–213.

Jensen, A.R. (1987). Individual differences in the Hick paradigm. In P.A.Vernon (Ed.), *Speed of information processing and intelligence.* Norwood, NJ: Ablex.

Kimberg, D.Y. & Farrah, M.J. (1993). A unified account of impairments following frontal lobe damage: The role of working memory in complex organised behaviour. *Journal of Experimental Psychology: General, 122,* 411–428.

Mayes, A.R. (1988). *Human organic memory disorders.* Cambridge: Cambridge University Press.

Maylor, E.A., Godden, D.R., Ward, T., Robertson, I.H. & Baddeley, A.D. (in press). Aging, selective attention and the influence of modality. *British Journal of Psychology.*

Miller, E. (1977). *Abnormal ageing: The psychology of senile and presenile dementia.* London: Wiley.

Mittenberg, W., Seidenberg, M., O'Leary, D.S. & DiGiulio, D. (1989). Changes in cerebral functioning associated with normal ageing. *Journal of Clinical and Experimental Neuropsychology, 11(6),* 918–932.

Moscovitch, M. & Winocur, G. (1992). The neuropsychology of memory and aging. In F.I.M. & T.Salthouse (Eds.), *The handbook of ageing and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Parkin, A.J. & Walter, B.M. (1992). Recollective experience, normal aging and frontal dysfunction. *Psychology and Aging, 7*, 290–298.

Petit, T.L. (1982). Neuroanatomical and clinical neuropsychological change in ageing and dementia. In F.I.M.Craik & S.Trehub (Eds.), *Ageing and cognitive processes.* New York: Plenum Press.

Rabbit, P.M.A. (1993). Does it all go together when it goes? The nineteenth Bartlett Memorial Lecture. *Quarterly Journal of Psychology, 46A,* 385–434.

Rabbitt, P.M.A. (1996a). Do individual differences in speed reflect "global" or "local" differences in mental abilities? *Intelligence, 22,* 69–88.

Rabbitt, P.M.A. (1996b). Intelligence is not just mental speed. *Journal of Biosocial Science, 28,* 425–449.

Rabbitt, P. & Yang, Q. (1996). What are the functional bases of individual differences in memory ability? In D.Herrmann, C.MacEvoy, C.Hertzog, P.Hertel, & M.K.Johnson (Eds.), *Basic and applied memory research, Vol 1.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

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

Salthouse, T.A. (1985). *A theory of cognitive aging.* Amsterdam: North Holland.

Salthouse, T.A. (1992). What do adult age differences in the Digit Symbol Substitution Test reflect? *Journal of Gerontology, 47(3),* 121–128.

Salthouse, T.A. (1993). Speed mediation of adult age differences in cognition. *Developmental Psychology, 29(4),* 722–738.

Salthouse, T.A. & Babcock., R.L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27(5),* 763–776.

Schaie, K.W. (1989). Perceptual speed in adulthood: Cross-sectional and longitudinal studies. *Psychology and Aging, 4(4),* 443–453.

Shaw, T.G., Morter, K.F., Sterling Meyer, S., Rogers, R.C., Hardenberg, J. & Cuitaia, M.M. (1984). Cerebral blood flow changes in benign aging and cardiovascular disease. *Neurology, 34,* 855–862.

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist, 57,* 421–457.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 17,* 1–32.

Stuss, D.T. & Benson, D.F. (1986). *The frontal lobes.* New York: Raven.

Thaler, M. (1959). Relationships among Wechsler, Weigl, Rorschach, EEG findings and abstract behaviour in normal aged subjects. *Journal of Gerontology, 11,* 404–409.

Vernon, P.A. (1983). Speed of information processing and intelligence. *Intelligence, 7,* 53–70.

Vernon, P.A. (1985). Individual differences in general cognitive ability. In L.C.Hartledge & C.F.Telzner (Eds.), *The neuropsychology of individual differences: A developmental perspective,* New York: Plenum.

Vernon, P.A. & Jensen, A.R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5,* 411–423.

Veroff, A.E. (1980). The neuropsychology of aging. *Psychological Research, 41,* 259–268.

Weiskrantz, L.S. (1992). Introduction: Disassociated issues. In A.D.Milner & M.D.Rugg, (Eds.), *The neuropsychology of consciousness* (pp. 1–10). London: Academic Press.

West, R.L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292.

Whelihan, W. & Lesher, E. (1985). Neuropsychological changes in frontal functions with ageing. *Developmental Neuropsychology, 1,* 371–380.

CHAPTER THREE
Testing Central Executive Functioning with a Pencil-and-paper Test

Alan Baddeley *Department of Psychology, University of Bristol, U.K.*

Sergio Della Sala and Colin Gray *Department of Psychology, University of Aberdeen, U.K.*

Costanza Papagno and Hans Spinnler* *Third Neurology Department, University of Milan, San Paolo Hospital, Italy*

INTRODUCTION

Executive Functioning and the Frontal Lobes

For well over a century, it has been known that the frontal lobes are involved in the ordering of the motor movements required in the effective performance of any action with a decision element (Bianchi, 1895; Feuchtwanger, 1923; Jastrowitz, 1888; Kleist, 1934; Luria, 1973). More recently, however, it has become clear that the role of the frontal lobes includes the control of cognitive processes. In the past 20 years or so, a considerable body of evidence has converged upon the view that the frontal regions of the brain, rather than themselves performing specific cognitive operations such as memorising, learning or reasoning, are concerned instead with the *deployment* of the capacity to carry out such processes, which take place elsewhere in the brain. That is to say, the frontal lobes are thought to have a supervisory or "executive" function; in fact, many neuropsychologists use the adjectives "frontal" and "executive" interchangeably, a practice which ignores the problems inherent in mixing functional and anatomical terms (Tranel, Anderson, & Benton, 1994).

The assumption that the terms "frontal" and "executive" are synonymous is also reflected in clinical assessment. A number of neuropsychological tests have been devised (or adapted) specially for the purpose of tapping the supposed "executive dysfunction" believed to be one of the sequelae of frontal lobe damage. In the absence of independent proof of this hypothesis, interpretation of the scores from such tests would seem to have an element of circularity, but their advocates point to a number of studies showing that scores tend to be lower on

* The authors' names are given in alphabetical order.

"frontal" tests after injury to the frontal regions (for a review of such work, see Stuss, Eskes, & Foster, 1994). Neuropsychological tests, however, do not perform well as detectors of anatomical change (for a review, see Reitan & Wolfson, 1994).

Although the location of executive functioning in a specific region of the brain may seem satisfactorily to anchor a hypothetical construct in a bedrock of anatomical reality, there is good reason to make a clear distinction between anatomical and functional concepts (Baddeley, 1996; Baddeley & Wilson, 1988). From the theoretical point of view, however, the concept of executive functioning is very attractive, because it seems to make sense of a wide range of otherwise puzzling phenomena. Yet, despite the theoretical flexibility of the executive hypothesis—indeed perhaps *because of* that flexibility—there has always been concern that, without clear theoretical specification and empirical demonstration of its components and the manner in which they interact, "executive functioning" could only too easily become a convenient *deus ex machina* with which "awkward" results could be explained away *a posteriori* (see Della Sala & Logie, 1993).

This latent unease about executive functioning has recently found expression in the observation that the empirical basis for the status of putative tests of frontal function is by no means as strong as many had supposed. As Tranel and his associates have expressed it: "Modern neuropsychology has met only qualified success in establishing procedures that can be unequivocally linked to executive functions…" (Tranel, Anderson, & Benton, 1994, p. 144).

A Correlational, Factor-analytic Approach to Executive Functioning

One approach to the investigation of executive functioning is to study a substantial number of patients with frontal lobe damage by presenting them with a battery of tasks that could reasonably be regarded as requiring "executive" functioning. This strategy has been followed by Duncan, Johnson, Swales, and Freer (in preparation) and by Della Sala, Gray, Spinnler, and Trivelli (in preparation). This kind of research has yielded somewhat contradictory results. On the one hand, Duncan et al., in a group of head-injured patients, found only low correlations among various "frontal" tests; indeed each of these tests correlated just as highly with supposedly "non-frontal" tests as they did with one another. This finding, however, contrasts with the results of the study by Della Sala et al.with a group of patients who had sustained only a single frontal lobe lesion, in which the correlations among the tests in the frontal group were not only positive but also often very substantial, the R-matrix exhibiting the "positive manifold" pattern regarded by most theorists of intelligence as demonstrating the existence of a general factor of intelligence, common to any and every test of ability (e.g. Spearman, 1927). Moreover, the "frontal" tests

justified their inclusion in a separate group by correlating more highly with one another than with some supposedly "non-frontal" tests. The results of the study by Della Sala and his associates, however, also highlight the need to distinguish between *anatomical* localisation of damage and *functional* deficit. The two often fail to correspond, suggesting that the link between anatomy and cognition is tenuous.

Despite the commonality between the frontal tests that was demonstrated in the study by Della Sala et al., there are several major problems with the correlational approach. The Wisconsin Card Sorting Test, for example, is widely regarded as the "frontal" or "executive" test *par excellence;* but this esteem is based largely upon Milner's (1963, 1964) studies of epileptic patients with frontal lesions, a narrow basis for such a generalisation (Reitan & Wolfson, 1994). Certainly, the Wisconsin test is often failed by those who have sustained frontal lobe damage. It is also, however, failed by patients with lesions in other areas (e.g. Anderson, Damasio, Jones & Tranel, 1991; Teuber, 1964); and it is sometimes adequately performed by patients with massive frontal lesions (e.g. Brazzelli, Colombo, Della Sala, & Spinnler, 1994; Eslinger & Damasio, 1985). Moreover, the results of some experiments by Dunbar and Sussman (1995) indicate that performance on the Wisconsin Card Sorting Test may be more dependent upon the phonological aspects of working memory than it is upon central executive functioning.

The problem with the correlational approach may be that each "frontal" test is, at best, an exceedingly impure measure of executive functioning. Indeed, executive functioning may also be involved in the so-called "non-frontal" tests. We should note that while there is reason to question the "frontal" (i.e. executive) status of some tests, it is also incumbent on those who speak of other tests as being "non-frontal" (i.e. non-executive) to justify their own claim. Spinnler (1991) has pointed out that it would be a most difficult enterprise to design an experiment sensitive only to attentional processes, and not to more posterior functions such as language and perception. It may be that the distinction between "frontal" and "non-frontal" tests is a matter of degree, rather than a true dichotomy. It is noteworthy that in the study by Della Sala et al. (in preparation), a strong general factor emerged even when the "frontal" and "non-frontal" tests were included together in the same R-matrix: from the loadings of all the tests (except digit span) on a single factor, good approximations to the correlations in the R-matrix could be produced, although the approximations were not as good as when the factor analysis was carried out on the tests in the putative frontal group alone.

Which factor analysis gives the "correct" result? The answer is neither and both. Of all the theorists of intelligence, Godfrey Thomson was most aware of the dangers arising from reification of the factors emerging from factor analytic research: Thomson stressed that the factors produced by factor analysis are

statistical realities but *psychological* fictions: "Factors are fluid descriptive mathematical coefficients, changing both with the tests used and with the sample of persons…" (Thomson, 1939, p. 299).

Perhaps, having established that there is a pattern of correlations consistent with the interpretation that there is a "general factor" operating, albeit to varying extents, in any and every test of cognitive functioning, the factor-analytic strategy can take us no further. To unpack the general factor, other approaches must be sought.

Testing Cognitive Models of Executive Functioning

The great problem with correlational and factor-analytic research is that, since correlation does not imply causation, this strategy is ill suited to the testing of precise hypotheses about cognitive processes. Substantive hypothesis testing is best approached by following an experimental strategy in which the putative causal variables can be manipulated by the experimenter rather than merely sampled as they occur in participants.

In recent years, cognitive psychologists have been studying executive functioning by carrying out experimental tests of cognitive models. There is good reason to hope that such models may help us to obtain a clearer picture of executive functioning than the shadowy factors yielded by the correlational approach.

There are actually several cognitive models to which we might turn to try to shed light upon executive functioning, especially in the neuropsychological context. There are, for example, Duncan's model of goal neglect (Duncan, Burgess, & Emslie, 1995) and Shallice and Burgess's supervisory attentional system (1991). In this chapter, however, we shall draw upon the *working memory* model, proposed by Baddeley and his associates (Baddeley, 1986; Baddeley & Hitch, 1974).

Working memory can be defined as a system for the temporary maintenance and manipulation of information. The model comprises an attentional control system, the central executive (CE), which coordinates the operation of two slave systems, the phonological loop, which is assumed to be a mechanism for the retention of speech-based material, and the visuospatial scratchpad, which constructs and preserves visual images. By far the most complex of these components of working memory is the central executive, which is thought to act as a general attentional resource and to be involved in reasoning, decision making, calculation, comprehension and long-term as well as short-term retention (Baddeley, 1986, 1990). In the working memory model the two slave systems, the phonological loop and the visuospatial scratchpad, have been extensively studied and each has its own distinctive empirical characteristics. Since, however, the central executive has proved to be much more difficult to

study than the slave systems, it has been somewhat neglected. There is, nevertheless, a growing body of literature which has cast at least some light on executive functioning (see Della Sala & Logie, 1993).

Dual-task Methodology

In the working memory model, the central executive coordinates the two slave systems. A fruitful approach to the study of the central executive may therefore be to employ a paradigm requiring the simultaneous operation of the phonological loop and the visuospatial scratchpad. One way of doing this is by using dual-task methodology. In the present context, the dual-task paradigm is advantageous in two respects: first, the experimenter can exercise fine control, both on the nature of the tasks and their individual levels of difficulty; second, in contrast with traditional "frontal" tests such as the Wisconsin, the use of exactly the same tasks in both the single and dual conditions should permit inferences about attentional load that are unconfounded by effects that might arise from a change of task.

Experimental Neuropsychological Evidence for Executive Impairment

If the central executive is indeed coordinating the two slave systems, then impairment of the executive should express itself, especially under dual-task conditions, as an inability to coordinate the phonological loop and the scratchpad.

This possibility was explored by Baddeley, Logie, Bressi, Della Sala, and Spinnler (1986), in a study in which performance of patients with dementia of the Alzheimer type (AD patients) was compared with that of matched healthy controls. Participants were required to perform simultaneously a computerised tracking task and one of several concurrent tasks. The most clearly interpretable of the latter was auditory digit span, since the two tasks had been carefully equated for difficulty in the two groups. The AD patients, in contrast with the controls, showed a marked decrement in their performance under dual-task conditions.

While the results of the study by Baddeley et al. (1986) are consistent with the interpretation that AD patients find it especially difficult to coordinate the tasks, suggesting a problem with the CE, there is also the possibility that they were simply overtaxed by information overload in the dual-task con dition. Accordingly, Baddeley and coworkers (Baddeley, Bressi, Della Sala, Logie & Spinnler 1991), in a follow-up study with the same group, retested the AD patients on the dual-task paradigm. The rationale was that if the deficit in the ability to allocate cognitive resources is indeed a central feature of AD, dual-task performance should show a more pronounced decrement as the disease progressed than should the two individual tests when performed alone. This

proved to be the case, confirming the hypothesis that, in Alzheimer's disease, there is a characteristic impairment of the central executive component of working memory. In another experiment, designed to rule out the confounding variable of task difficulty, Baddeley et al. (1991) showed that there was no tendency for more difficult single tasks to show greater sensitivity to the progression of the disease. The lack of an effect of single-task difficulty over time, in contrast with a marked decrement in dual-task performance, lends powerful support to the hypothesis of a CE impairment in AD.

The Development of a Pencil-and-paper Version of the Dual-task Paradigm

Pilot Work Based on Fitts's Law: The Dot Placement Task. The computerised dual-task paradigm used in the studies reviewed so far presents problems for day-to-day clinical use, since it requires expensive equipment, making it difficult to ensure standardisation. It was decided, therefore, to attempt to develop a simple pencil-and-paper version of the task for use in the clinical setting.

It must be admitted at the outset that, in trying to devise a portable test, we encountered a number of unexpected problems. The rationale of the first procedure we tried was based upon Fitts's law (see Fitts & Peterson, 1964), which states that the time taken to strike a target varies positively with the distance of the target and inversely with its width. The participant, paced by a metronome, attempted over a period of two minutes to tap a dot with a pencil inside each of a two-row array of identical circles drawn on a sheet of paper. Subsequently, the diameter of the circles was progressively reduced. The participant's auditory digit span (with oral response) was also ascertained, and the two tasks were combined in a dual-task paradigm.

Even under single-task conditions, all participants (whether patients or healthy controls) found it extremely difficult to follow the rhythm of the metronome; indeed, most stopped even trying to do so after a short time. (This was especially true of the AD patients.) Moreover, the AD patients had difficulty in following instructions on the order of target circles, and would typically stop at the head of each column and ask the experimenter what they had to do next. Often, having placed a dot in a circle, a patient would continue to place further dots in the same circle, rather than pro ceeding to the next. Another frequent problem was that when the examiner began to present the digits, the patient would stop tracking altogether.

The span procedure also presented problems. Having established a participant's span in an untimed test, we found that span length was too great for a situation in which there was a time limit. Another difficulty was that, in the dual-task situation, the patients frequently began writing numbers (instead of tapping dots) inside the circles. Because of the inordinate numbers of errors

produced with traditional span tests under timed conditions, we began to use shorter sequences, eventually settling on the greatest length at which all three sequences had been correctly reproduced. Performance was then expressed in terms of the number of sequences of this length that were correctly recalled over the timed interval, expressed as a proportion of the total number of presented sequences.

Pilot Studies with Mazes. Suspecting that the necessity to lift the pencil when proceeding from one circle to the next placed too great a load on the decision-making capacity of the patient, we began to experiment with mazes of decreasing path width, through which the participant could proceed by connecting dots without lifting the pencil. The maze tests, however, also produced too many errors. Moreover, when the dual-task condition was introduced, even the simplest of the mazes proved to be too difficult. The tracking task was therefore changed to one of traversing increasing complex mazes with paths of constant width. Meanwhile, the controls found the modified mazes too easy, and their scores showed an obvious ceiling effect. With the modified mazes, care had to be taken to avoid a learning effect by varying the pattern of corners from the single to the dual-task condition.

In the end, the mazes had to be abandoned altogether. The main difficulty was that of scoring the performance of a patient whose pencil had wandered off the path. Moreover, there remained the problem of the patient pausing and asking for further instructions. In the tracking task that was eventually decided upon and used in all subsequent researeh, the participant is presented with a sheet of paper upon which is drawn a pattern of squares, which are to be traversed with a continuous pencil line.

REVIEW OF STUDIES USING THE PENCIL-AND-PAPER VERSION OF THE DUAL TASK

The Paradigm

The paradigm used to gather the data in the present study drew heavily upon the experience gained from the earlier pilot work. Basically, the memory and tracking components were variants of the tasks reported in the previous section, suitably modified to achieve the required levels of difficulty and to avoid the problems encountered in the pilot study.

The Memory Span Task

Establishing the Baseline for Digit Span. The participant's digit span was first tested by using a standard procedure. A list of digits was read aloud by the examiner at the rate of one digit per second and the participant was asked to

repeat them in their order of presentation. Further lists were presented, each group of three being one item longer than those in the preceding group. The participant's digit span was taken to be the maximum length at which all three lists were reproduced without error. While this criterion may be unconventionally conservative, the pilot work had indicated that it was best for present purposes.

The Single-task (Memory) Condition. After the participants' span had been determined in the manner described, they were presented continuously, over a period of two minutes, with lists of digits at span length, the items of which had to be spoken back in order of presentation. The number of lists presented to a participant naturally varied, depending partly upon the participant's span and partly upon the rapidity of the verbal response. The performance measure, therefore, was the *proportion* of lists correctly recalled.

The Dual-task (Memory) Condition. In the final phase, participants were once again presented, over a period of two minutes, with fresh lists of digits at span length, while at the same time performing a tracking task (see below). As with the single-task condition, the memory measure was the proportion of the presented lists that were correctly recalled.

The Tracking Task

The Single-task Condition. Participants were required to use a felt pen to cross out boxes (of size 1 cm square) which had been linked to form a path laid out on an A4-size sheet of white paper. They were first given a number of practice trials with a short, 10-box path, to accustom them to the procedure, and to ensure that they really had understood what was required of them (see Fig. 3.1).

In the testing session proper, participants were required, over a two-minute interval, to follow paths through sheets with 80 boxes (see Fig. 3.2). The participant was asked to start at one end of the chain and place a cross in each successive box as quickly as possible. If all the boxes on the sheet had been traversed before the time limit of two minutes had elapsed, a second sheet of boxes was presented. The score was the number of boxes successfully marked by the participant.

The Dual-task Condition. Once again, the participants performed the tracking task, while at the same time verbally reproducing lists of digits.

Some Studies with the Pencil-and-paper Version of the Dual-task Paradigm

This pencil-and-paper task has been successfully used by Della Sala and his associates (Della Sala, Baddeley, Papagno, & Spinnler, 1995) to replicate the findings of the study by Baddeley et al. (1986) with 12 AD patients and 12 matched controls. The results showed not only strong main effects of the group

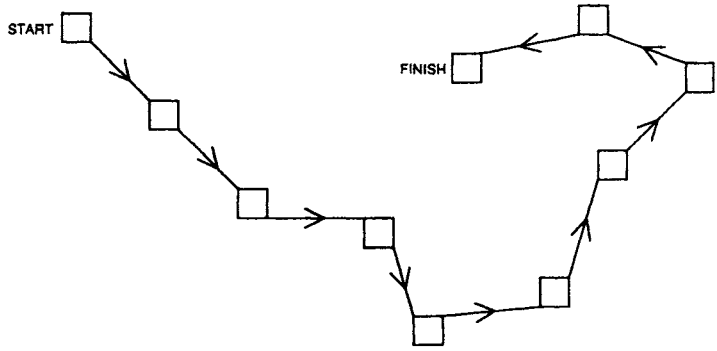


FIG. 3.1. The short, practice form of the tracking task.

and condition factors but also a striking interaction: in comparison with the controls, the AD patients showed a much larger decrement in performance under dual-task conditions. The pencil-and-paper task, therefore, when used in comparative group studies, is sensitive to one of the characteristic effects of AD, while at the same time being relatively insensitive to the normal effects of ageing.

Importantly, the new test also appears to be more sensitive than "frontal" tests to behavioural changes arising from frontal lobe damage. Baddeley, Della Sala, Papagno, and Spinnler (1997) compared dual-task performance in dysexecutive and non-dysexecutive patients, each with a focal frontal lesion, and found that the dysexecutive patients showed a larger decrement in memory performance under dual-task conditions than did the non-dysexecutive patients (there was no interaction in the tracking data). Two classic "frontal" tests, word fluency and the Wisconsin Card Sorting Test, failed to differentiate between the two groups of patients. Alderman (1996) found that some severely brain-damaged patients who failed to respond to a normally successful rehabilitation scheme based upon a token economy performed relatively more poorly on the dual task than patients who responded to the system in the usual way. Moreover, standard "frontal" tests, such as the Wisconsin Card Sorting Test and the verbal fluency test, showed little association with progress (or lack of it) in the token economy.

The pencil-and-paper dual-task technique has already begun to be used more widely. Greene, Hodges, and Baddeley (1995) have also used the test to study AD patients. They observed that not only did dual-task performance show a marked decrement in the AD patients in comparison with healthy controls, but also that performance on this pencil-and-paper task correlated significantly with scores on another, quite different, dual task procedure borrowed from Robertson, Ward, and Ridgeway's (1994) Test of Everyday Attention. The pencil-and-paper dual task has also been used to demonstrate a mild decrement in a sample of patients affected by Parkinson's disease (Dalrymple-Alford, Kalders, Jones, & Watson, 1994). Cowey and Green (1996) have used the test to compare frontally lesioned with hippocampally damaged patients. Interestingly, only the frontal

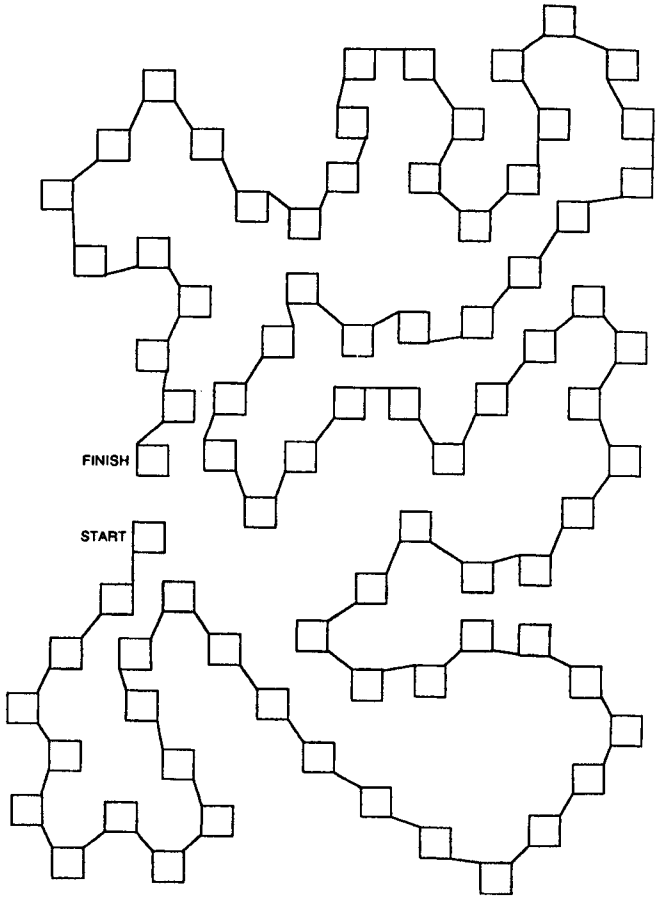


FIG. 3.2. The full version of the tracking task, used in the experiment proper.

group showed a strong dual-task decrement, supporting the hypothesis that the frontal regions of the brain are concerned with coordination of different activities.

The paper-and-pencil version of the dual task has proved to be very promising in several ways. It is easy and quick to administer, very portable and readily usable in a clinical context. There remain, nevertheless, a number of problems which must be overcome if the measuring instrument is to achieve its full potential.

In the dual-task situation, participants may decide either to perform both tasks to the best of their ability, or to concentrate on one at the expense of the other. An examination of dual-task data from several different studies suggests that participants vary in the way they distribute their attention between the two tasks. For example, in the study by Greene et al. (1995), it was found that patients seem

to have concentrated upon the memory task at the expense of tracking: that is, the two component tasks did not show the same pattern of change from the single —to the dual-task situation. In other work, however, patients appeared to concentrate upon the tracking task at the expense of memory (Della Sala et al., 1995). These inconsistencies among different studies may arise from varying aetiologies and rapidity of fatigue onset within sessions.

To consider performance only on each component task separately, therefore, may be to miss important aspects of dual-task performance, and there is clearly a need for a single performance index of dual-task performance which takes into consideration *both* component tasks, so that if participant A, while equalling participant B on, say, tracking, has a much lower score on memory, A will be assigned a lower overall performance score than will B.

The need for a single index is further underlined by the psychometric considerations that apply when a test is to be used as part of the clinical assessment of *an individual patient,* as opposed to the making of a comparison between groups. In psychological (and neuropsychological) research in general, it is important to distinguish between what is true of large groups or samples and what can be inferred from an individual score, which may be quite atypical of the group.

In interpreting an individual's score on a clinical test, attention must be paid to the *reliability* of the measuring instrument, which is the extent to which the individual, if tested again subsequently, would achieve a similar score, with a quantile value similar to that of the mark achieved on the first occasion of testing.

For a test to be really useful in a clinical setting, the clinician needs to be able to say whether an individual's score on a test is such that there is reason to suspect the presence of functional impairment. What is required is a cut-off point, or rule of thumb, which can indicate the need for closer investigation.

One might, for example, obtain a large sample of scores on a neuropsychological test and, after the fashion of those who assess a firm's claim to produce machinery with certain specifications, determine the lower tolerance limit, that is, the value below which we can be, say, 95% confident that no more than 5% of the healthy population lies. But the fact that an individual has achieved a score lower than those of 95% of the population would have considerably less impact were it to be known that an individual scoring below the fifth percentile on one occasion might well attain the 80th percentile on another. This would be the case if scores were assigned on a purely random basis, a procedure comparable in its effects to the use of an unreliable test.

A NORMATIVE STUDY

Aims

The purpose of this study was to test the performance of a substantial number of healthy participants on the pencil-and-paper version of the dual-task paradigm, with a view to investigating the statistical behaviour of a composite measure of dual-task performance designed to reflect the quality of performance on *both* tasks.

We also intended to study the association of dual-task performance with such demographic variables as age, level of education, and gender. While it is well known that some aspects of memory decline with age, the trajectory of executive functioning over the life span is much less well known.

We also planned to test some of the participants in the study on a second occasion to permit a preliminary investigation of the test—retest reliability of the overall measure of dual-task performance, with a view to arriving at a region of performance which should alert the clinician to the possibility of brain damage.

The Sample

One hundred and eight healthy individuals participated in the present study on a voluntary basis. They were selected so that they would show adequate variation on demographic characteristics (such as age, gender, and level of education) that might affect performance. Table 3.1 shows that the participants spanned almost the entire adult age range, the youngest being 20 and the oldest 99 years of age.

Searching for a Measure of Dual-task Performance

Dual-task performance is intended to be an indicator not of working memory nor of motor dexterity per se but of the patient's ability to deploy the available resources of memory capacity. Accordingly, we need a measurewhich is relatively independent of such variables as short-term memory, speed of box-marking, and so on. It was for this reason that the memory task was constructed on the basis of each individual's own digit span, and in the data set, the number of lists correctly recalled was expressed as a percentage of the number of lists presented, the lists being at span length for each individual. While we can expect digit span to be correlated with the proportion of lists recalled under either the single or dual task condition, the *difference* between the proportions correct, which should reflect the demand upon executive functioning, may not correlate appreciably with digit span.

TABLE 3.1 Characteristics of the Participant Sample

Age distribution							
Age	20–30	31–40	41–50	51–60	61–70	71–80	over 80
Frequency	12	12	12	24	24	24	12

Gender distribution		*Educational level*	
male	54	low (< = 8yrs)	54
female	54	high (> 8yrs)	54

A Two-component Measure of Dual-Task Performance

Let n_s and n_d be, respectively, the number of lists correctly recalled under the single and dual conditions; correspondingly, the total numbers of lists presented under those two conditions are N_s and N_d, respectively. The proportions of lists recalled under the single and dual task conditions, p_s and P_d are then given by $p_s = n_s/N_s$ and $p_d = n_d/N_d$ respectively. The proportional loss of memory performance under dual-task conditions (p_m) is given by $p_t = (t_s - t_d)/t_s$. For example, if, under single-task conditions, a patient were to be presented (over the fixed interval) with 12 lists of digits, of which 8 lists were recalled correctly, n_s=8, N_s=12 $p_s = 8/12 = .67$. If, under dual task conditions, 13 lists were presented to the same individual over the same interval, of which 7 were correctly recalled, n_d=7, N_d=13 and $p_d = 7/13 = .54$. The proportional loss of memory under dual-task conditions is $p_m = (p_s - p_d) = .67 - .54 = .13$.

Let t_s and t_d be the tracking performance (the number of squares marked) under single and dual conditions, respectively. The proportion p_t of single-task tracking performance lost in the dual-task situation is given by

$$p_t = (t_s - t_d)/t_s$$

Suppose that our patient marked 179 squares under single-task conditions and 164 in the dual task. Then t_s=179, t_d=164 and $p_t = (179 - 164)/179 = .08$

It is desirable to have a measure which positively reflects the patient's level of performance, i.e. a low score is "bad" and a high score is "good".

Our measure (μ) expresses an individual's dual-task performance as a percentage of single-task performance, the contributions from the two tasks being equally weighted. The measure μ is defined as:

$$\mu = \left[1 - \frac{(p_m + p_t)}{2}\right] \times 100$$

In the present case, p_m=.13, p_t=.08 and so

$$\mu = 100[1 - (.13 + .08)/2] = 90\%$$

Under dual-task conditions, therefore, and taking both tasks into consideration, our individual performed at 90% of the single-task level.

Table 3.2 (a, b, c, and d) shows that the mean value of the distribution of μ is 91.99% (the median is 91.41%, suggesting that μ is symmetrically distributed).

TABLE 3.2a Distribution of the Index of Dual-task Performance

Mean (N = 108)	*SD*	*Median*	*Lower quartile*	*Upper quartile*	*Minimum, maximum*
91.99	11.16	91.41	85.20	100.33	53.50, 120.45

TABLE 3.2b Means for the Seven Age Bands (Standard deviations are given in brackets)

Age	*up to 30*	*31–40*	*41–50*	*51–60*	*61–70*	*71–80*	*over 80*
Freq.	12	12	12	24	24	24	12
Mean	92.11 (8.57)	94.30 (11.24)	97.69 (9.24)	94.10 (6.12)	90.92 (10.44)	91.27 (10.19)	85.32 (18.13)

Table 3.2c Means for Gender (Standard deviations are given in brackets)

Gender	*Male*	*Female*	*t*	*p*
Mean	90.85 (10.85)	93.12 (11.45)	1.06	.2928

Table 3.2d Means for Educational Level (Standard deviations are given in brackets)

Educ. Level	*High*	*Low*	*t*	*p*
Mean	90.70 (10.76)	93.28 (11.50)	1.20	.2325

The standard deviation is 11.16%. So the participants were performing, on average, at about 8% below single-task level; but the minimum level was about 54% and the maximum about 120%. A number of participants, that is, actually performed *better* under the dual-task condition than they did under the single-task condition. (This may have been a practice effect.) With a score of 90%, our present participant lies only 2% (less than 1/3 of a standard deviation) below the mean, and is obviously well within the normal range.

Figure 3.3 is a histogram of the distribution of μ, with the corresponding normal distribution (N[92, 124.5]) drawn in for comparison purposes. The distribution of μ appears to be approximately normal, and it passes the stringent Kolmogorov-Smirnov (Massey, 1951), Lilliefors (1967) and Shapiro-Wilk (1965) tests for normality.

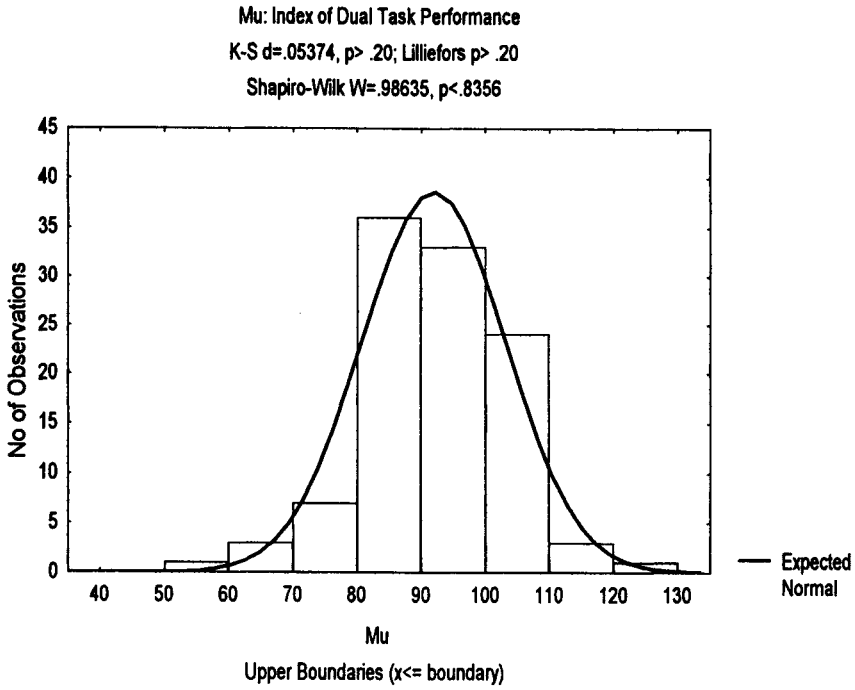


FIG. 3.3. Histogram of the distribution of the index of dual task performance (μ).

Correlations of the Index with Age and Digit Span

Since the measure μ was constructed so as not to reflect the individual's digit span, we should not expect a substantial correlation between μ and digit span. Indeed, the correlation (.17) is small and statistically insignificant. There is also a small and insignificant correlation of μ with age.

If there were different distributions for men and women, a bimodal distribution of μ would be expected when men and women are combined in a single sample. In fact, the mean values of μ in the male and female participants do not differ significantly. The means for the male and female participants are 90.85% and 93.12%, respectively: $t(106)=1.06$, $p=.2928$. Similarly, there is no significant difference in the mean values of μ for the "high" and "low" educational groups, which are 90.70% and 93.28%, respectively: $t(106)=1.20$, $p=.2320$.

The Reliability of the Dual-task Performance Index

Of the 108 participants who took part in this investigation, the retest scores of 33 were available for a preliminary investigation of test—retest reliability. The

Pearson correlation between the values of μ achieved by the 33 participants on the first and second occasions of testing is 0.44 (p=.01).

The disappointingly low reliability of μ derives partly from the noisiness of the memory component of the index: the reliability of the memory component is only .36 (p=.037), whereas that of the tracking component is .76 (p<.00005). Another factor, however, is the small negative correlation (−.0582) between the memory and tracking components of the dual task.

The reliability might be improved by obtaining a larger sample of the participant's behaviour and testing over, say, four-minute, rather than two-minute intervals. On the other hand, it is essential to avoid overtaxing or distressing patients, and longer intervals may not be realistic.

The finding that the reliability of the composite index μ is low has considerable implications for the interpretation of what otherwise might seem a conspiciously poor score. Given that the mean and standard deviation of the distribution of μ are 92% and 11.2% respectively, the 5% tolerance limit is 67. 21%, a value with a deviation of −24.8% from the mean of 92%. If the index μ had a high reliability, (0.95 or greater), it would be reasonable to regard any score below 67.21% as abnormally low and to consider further investigation. But the index μ at least when based upon the dual task in its present form, does not have a high reliability. This means that confidence zones should be constructed, within which the participant's true score is likely to lie; and with a reliability of, say, 0.44, such zones must cover a wide range of values, making the interpretation of an individual score difficult.

Suppose, for example, that a patient's score on μ is 42%, that is, a deviation of −50% (from the mean of 92%). Since this is well below the lower tolerance limit of 67.2%, it might seem that the patient's performance level is alarmingly low. That judgment, however, would fail to take into consideration the unreliability of the measurement.

The standard error of the measurement *(SEM)* is given by the formula

$$SEM = SD \times \mathrm{SQRT}(1 - \text{reliability}),$$

and in this case has the value 8.38%. The 95% confidence interval for a participant's true (deviation) score is obtained by first estimating the true score with t', where

$$t' = \text{reliability} \times (\text{deviation score}).$$

In the present example, the patient's deviation score is −50%, so that t'= −22% (i.e. a raw percentage score of 70%). It will be noticed that the estimate of the participant's true deviation score is considerably less "deviant" than the original observed score; in fact, the estimate is now greater than the lower tolerance limit (67.2%). This is "regression to the mean"; and the lower the reliability, that is to say the more error there is in the measurement, the more regression to the mean there will be.

The estimate of the true deviation score is placed at the centre of the 95% confidence interval $t' - 2SEM, t' + 2SEM$ which in this case is [−38.76%, −5.24%] or, transforming to raw percentages, [53.24%, 86.76%], which is a very wide range (34%) of possible values. The patient's true score may well be in the normal range.

Returning to our original participant (whose score was 90%), the 95% confidence interval for the true score is [74.4%, 107.9%], a range of values straddling the mean of 92% and confirming that this person is well within the normal range. In contrast, the confidence interval for the participant whose score was 42%, i.e. [53.24%, 86.76%], not only fails to straddle the mean of 92% but also includes the 5% tolerance limit (67.2%). It might therefore be felt that further investigation is advisable.

Had the reliability of μ been higher, say 0.8, the interpretation of a low score would have been much easier, because the confidence interval for the patient's true score would have been considerably tighter. Returning to the previous example, a patient with a μ of 42% (i.e. a deviation score of −50%) would have an estimated true score of −40% (a raw percentage score of 52%), which (although there has been some regression to the mean) is still below the 5% tolerance limit (67.2%). The 95% confidence interval for the patient's true deviation score is approximately [−50%, −30%], or (in raw percentages) [42%, 62%], a narrower interval spanning only 20%. Since the lower tolerance limit is outside and above this interval, we should have reason to suspect that the patient's true performance is indeed abnormally low.

CONCLUSION

The correlational, factor-analytic approach to the study of executive functioning has only limited power to illuminate the manner in which the central executive works. In general, because correlation does not imply causation, correlational research is unsuited to the testing of precise hypotheses about mechanisms. In particular, the factors yielded by factor analysis, being "fluid mathematical coefficients" (Thomson, 1939, p. 240) which reflect such considerations as the composition of the test battery and the nature of the patients tested, are better thought of as descriptions, rather than explanations.

In this chapter, we have considered one attempt to explore executive functioning in terms of one model in cognitive psychology, namely, the working memory model of Baddeley, Hitch, and their associates. The model predicts that where there is reason to suspect that executive functioning may be impaired, as with patients who have sustained frontal lesions, participants should find the dual-task situation especially difficult. This hypothesis has already received confirmation from a number of studies comparing frontal or Alzheimer patients with healthy individuals on dual-task performance. As predicted, the

performance of such patients showed an especially marked decrement under the dual-task condition.

A pencil-and-paper version of the original computerised dual-task paradigm has already been successfully used to replicate some of the comparative research described above.

The finding of a robust difference between groups of participants, however, does not imply that there is necessarily an easy interpretation of the score of an individual. The statistical properties of a single, composite measure of dual-task performance were investigated, and it was found that its reliability was rather low. In practical terms, this means that a low score carries less weight than if the measure were highly reliable. Confidence intervals for participants' true scores are rather too wide to be very useful in any but the most extreme cases.

From a variety of viewpoints reviewed in this chapter, there is good reason to expect that the dual-task paradigm will prove to be an effective way forward in future exploration of the precise nature of executive functioning. The suggested measure (μ) seems to be a reasonable index of dual-task performance. From the point of view of clinical practice, however, the dual-task technique requires further development to achieve the level of reliability required to formulate a decision rule that would be applicable in individual cases.

ACKNOWLEDGEMENT

We would like to thank Ms Sarah Stevenson for collecting the data.

REFERENCES

Alderman, N. (1996). Central executive deficit and response to operant conditioning methods. *Neuropsychological Rehabilitation, 27,* 479–504.

Anderson, S.W., Damasio, H., Jones, R.D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical Experimental Neuropsychology, 13,* 909–922.

Baddeley, A.D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A.D. (1990). *Human memory: Theory and practice.* Hove and London: Lawrence Erlbaum Associates.

Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A,* 5–28.

Baddeley, A.D., Bressi, S., Della Sala, S., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease. *Brain, 114,* 2521–2542.

Baddeley, A.D., Della Sala, S., Papagno, C, & Spinnler, H. (1997). Dual-task performance in dysexecutive and nondysexecutive patients with a frontal lesion. *Neuropsychology, 11,* 187– 194.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.Bower (Ed.), *The psychology of learning and motivation* (Vol. VIII pp. 47–90). New York: Academic Press.

Baddeley, A.D., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology, 38A,* 603–618.

Baddeley, A.D., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7,* 212–230.

Bianchi, L. (1895). The functions of the frontal lobes. *Brain, 18,* 497–522.

Brazzelli, M., Colombo, N., Della Sala, S., & Spinnler, H. (1994). Spared and impaired cognitive abilities after bilateral frontal damage. *Cortex, 30,* 27–51.

Cowey, C.M., & Green, S. (1996). The hippocampus: A "Working Memory" structure? The effect of hippocampal sclerosis on working memory. *Memory, 4,* 19–30.

Dalrymple-Alford, J.C., Kalders, A.S., Jones, R.D., & Watson, R.W. (1994). A central executive deficit in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 57,* 360–367.

Della Sala, S., Baddeley, A.D., Papagno, C. & Spinnler, H. (1995). Dual Task Paradigm: A means to examine the central executive. In J.Grafman, K.J.Holyoak, & F.Boller (Eds.), Structure and functions of the human prefrontal cortex: Special issue of *Annals of the New York Academy of Sciences, 769,* 161–172.

Della Sala, S., & Logie, R.H. (1993). When working memory does not work: The role of working memory in neuropsychology. In F.Boller & J.Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 8, pp. 1–62). Amsterdam: Elsevier.

Dunbar, K., & Sussman, D. (1995). Toward a cognitive account of frontal lobe function: Simulating frontal lobe deficits in normal subjects. In J.Grafman, K.J.Holyoak, & F. Boller (Eds.), Structure and functions of the human prefrontal cortex: Special issue of *Annals of the New York Academy of Sciences, 769,* 289–304.

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

Eslinger, P.J., & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation. *Neurology, 35,* 421–429.

Feuchtwanger, E. (1923). *Die Funktionen des Stirnhirns.* Berlin: Springer.

Fitts, P.M. & Peterson, J.R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67,* 103–112.

Greene, J.D.W., Hodges, J.R., & Baddeley, A.D. (1995). Autobiographical memory and executive function in early dementia of Alzheimer type. *Neuropsychologia, 33,* 1647–1670.

Jastrowitz, M. (1888). Beiträge zur Localisation im Grosshirn und über deren praktische Verwerthung. *Deutsche Medizinische Wochenschrift, 14,* 81–83, 108–112, 125–128, 151–153, 172–175, 188–192,209–211.

Kleist, K. (1934). *Gehirnpathologie.* Leipzig: Barth.

Lilliefors, H.W. (1967). On the Komogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 64,* 399–402.

Luria, A.R. (1973). The frontal lobes and the regulation of behavior. In K.L.Pribram & A.R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 3–26). New York: Academic Press.

Massey, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association, 46,* 68–78.

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

Milner, B. (1964). Some effects of frontal lobectomy in man. In J.M.Warren & K.Akert (Eds.), *The frontal granular cortex and behavior* (pp. 313–334). New York: McGraw-Hill.

Reitan, R.M., & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4,* 161–198.

Robertson, I.H., Ward, T., & Ridgeway, V. (1994). *The test of everyday attention.* Flempton: Thames Valley Test Company.

Shallice, T., & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shapiro, S., & Wilk, M. (1965). An analysis of variance test for normality. *Biometrika, 52,* 591–611.

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Spinnler, H. (1991). The role of attentional disorders in the cognitive deficits of dementia. In F. Boller & J.Grafman (Eds.), *Handbook of neuropsychology* (Vol. 5, pp. 79–122). Amsterdam: Elsevier.

Stuss, D.T., Eskes, G.A., & Foster, J.K. (1994). Experimental neuropsychological studies of frontal lobe functions. In F.Boller, H.Spinnler & J.A.Hendler (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 149–185). Amsterdam: Elsevier.

Teuber, H.-L. (1964). Discussion. In J.M.Warren & K.Akert (Eds.), *The frontal granular cortex and behavior* (pp. 332–333). New York: McGraw-Hill.

Thomson, G.H. (1939). *The factorial analysis of human ability.* London: University of London Press.

Tranel, D., Anderson, S.W., & Benton, A. (1994). Development of the concept of "executive function" and its relationship to the frontal lobes. In F.Boller, H.Spinnler, & J.A.Hendler (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 125–148). Amsterdam: Elsevier.

CHAPTER FOUR
Theory and Methodology in Executive Function Research

Paul W.Burgess *Department of Psychology, University College London, U.K.*

INTRODUCTION

The term "executive functions" is a relatively new one to neuropsychology (see, for example, Baddeley & Wilson, 1986, 1988; Lezak, 1983; Stuss & Benson, 1986). It is used to describe a range of poorly defined processes which are putatively involved in activities such as "problem solving" (Levin, Goldstein, Williams, & Eisenberg, 1991), "planning" (Shallice, 1982), "initiation" of activity (Burgess & Shallice, 1996c; Shimamura, Janowsky, & Squire, 1991), "cognitive estimation" (Shallice & Evans, 1978) and "prospective memory" (e.g. Cockburn, 1995; Shimamura, 1989). Historically these processes have always been linked to the frontal lobes (e.g. Benton, 1991) and damage to them has resulted in the range of symptoms previously known as the "frontal lobe syndrome" (see Stuss & Benson, 1986, for review).

It is probably fair to say that until recently the study of executive function has been a Cinderella area of (cognitive) neuropsychology. For instance, one of the standard textbooks in the area (Ellis & Young, 1988) has no mention at all of the area. Another (McCarthy & Warrington, 1990) has only 22 pages devoted to what they cautiously call "problem solving," when memory as an equally broad topic commands approximately three times as much space, and the neuropsychology of language commands approximately seven times the space! This state of affairs exists despite the generally accepted view that executive functions are those processes which "enable a person to engage successfully in independent, purposive, self-serving behaviour" (Lezak, 1995, p. 42). Given this characterisation, it is hardly surprising that those texts which concern themselves with more general aspects of the practice of neuropsychology (e.g. Kolb & Whishaw, 1990; Lezak, 1995) have more to say on the topic of executive/frontal lobe function, and the subject is one of primary concern to those working in rehabilitation settings, since the range of behavioural and cognitive disturbances

which these so-called "dysexecutive" (Baddeley & Wilson, 1988) patients often show suggest poor outcome and difficult management (e.g. Alderman & Burgess, 1994; Alderman, Fry, & Youngson, 1995; Burgess & Wood, 1990; Thomsen, 1987; Von Cramon & Von Cramon, 1994). For these reasons it would seem that the recent increase in research in the area has been to some extent driven by a practical need to understand executive function, rather than for pure theoretical reasons alone.

A further aspect of the study of executive functions has been the prevalence of group study methodology. Most investigators in the area of executive function have produced group studies at one time or another (e.g. Kopelman, 1991; Milner, 1963, 1982; Petrides, 1985; Schacter, 1987; Shallice, 1982; Shallice & Evans, 1978; Janowsky, Shimamura, & Squire, 1989; Stuss and his colleagues, 1982, 1991; and *all* chapters in this volume, to mention but a very few). Moreover, most of these large group studies have sought specifically to link certain types of behavioural manifestation with damage to certain brain structures.

In large part these two aspects of executive function research (group studies and brain structure links) do not sit easily with the creed of "ultra-cognitive" neurpsychology, where results from group studies are regarded as unsuited to theory development, and questions regarding the neurological basis of behaviour are irrelevant (Shallice, 1988, p. 203). For researchers within this tradition, the primary and hard fought-for development over the last 10 or 20 years has been rejection of exactly these methodologies and concerns (e.g. Caramazza, 1986; Caramazza & Badecker, 1991; Ellis, 1987; McCloskey, 1993). Thus those who might be described as "ultra-cognitive neuropsychologists" have little option but to disregard much of executive function research. Perhaps this is the reason for the complete omission of the subject in Ellis and Young's otherwise excellent textbook.

However, the possibility I want to consider here is that for both practical and theoretical reasons the exclusive study of either groups or single cases is unlikely to be sufficient a methodology in executive function research. The strategy most likely to succeed is one where both approaches are admitted. In particular I want to argue that within the area of executive functions, the key finding of theoretical interest to the ultra-cognitive neuropsychologist— the double dissociation—can often be misleading. The primary reasons, which are all interrelated, are as follows:

1. Executive tasks are less pure measures than non-executive ones.
2. The relative lack of correspondence between behaviour and putative cognitive processes in executive functions may lead to apparent dissociations between behaviours without this necessarily reflecting fractionation of underlying processes.

3. Serial dependency in sequences of complex behaviour may lead to misleading dissociations, on the one hand, and preclude the discovery of functional independence, on the other.
4. The nature of executive control is that it serves to adapt behavioural routines. This means that no two behavioural sequences which stress these functions to an identical degree will be exactly the same. Consequently, in the absence of necessary associations between observed behaviours, simple dissociations may be uninstructive.

These points will be discussed in detail, followed by a brief examination of the implications of measuring cognitive processing in novel situations using single-case design. I must stress that this chapter is *not* an attack on the utility of single-case methods. Far from it. It is intended as an appeal for principled discussion about the unique methodological demands made by the nature of executive system functioning.

THE NATURE OF EXECUTIVE FUNCTION

Current theoretical interpretations of the neuropsychological data from dysexecutive patients (i.e. patients who show problems with executive control) owe much to the distinction between automatic and controlled behaviour in traditional cognitive psychology (see, for example, Atkinson & Shiffrin, 1968; Schneider & Shiffrin, 1977); current consensus regards the executive system as a process or set of processes whose primary purpose is to facilitate adaptation to novel situations (see, for example, Burgess & Cooper, 1996; Karnath, Wallesch, & Zimmerman, 1991; Shallice, 1988, p. 345; Sirigu et al, 1995). This system works by "modulation and control of more fundamental or routine cognitive skills" (McCarthy & Warrington, 1990, p. 343; see also for related accounts Damasio, 1985; Duncan, 1986; Fuster, 1980; Luria, 1973; Shallice, 1982; Teuber, 1972). These routine cognitive skills are generally regarded as those which have been overlearned by practice or repetition, and thus can include anything from motor, reading, or language skills to semantic memory (Burgess & Shallice, 1996b). However most real-life situations require at least some adaptation of these skills (since no two situations are exactly alike), and Norman and Shallice (1980, 1986; see also Shallice & Burgess, 1991b) outline five types of situation where routine, automatic activation of behaviour would not be sufficient for optimal performance:

1. Those that involve planning or decision making.
2. Those involving error correction or troubleshooting.
3. Situations where responses are not well-learned or contain novel sequences of actions.
4. Dangerous or technically difficult situations.

5. Situations which require the overcoming of a strong habitual response or resisting temptation.

That these characterisations were necessary reveals a key aspect of current views on the role of executive processes in cognition: Neuropsychologists would hardly feel it necessary to define the circumstances under which speech production processes are in use; or, taking an example which is more theoretically oriented, situations which might stress the graphemic buffer (e.g. Ellis, 1982). In these examples, there is presumed to be a high degree of correspondence between the putative process and a specific behaviour. In other words, the cognitive system underlying the behaviour is supposed to be dedicated to it, or at least to a range of highly similar behaviours. Thus there are thought to be dedicated cognitive processes underlying reading, spelling, object recognition, and so forth. Neuropsychological (double dissociation) evidence of specific impairments has led to the current view of these cognitive systems as relatively "informationally encapsulated", and some have argued that they are examples of modularity in functional architecture beyond only the input level that Fodor (1983) envisaged, or the output level as Marshall (1984) suggests, (see Shallice, 1988, pp. 269–273; Shallice, 1984). Thus if Fodor's "central processes" now exist, they are probably synonymous with what neuropsychologists call executive processes. These are processes which guide behaviour in many different situations and, critically, work either by action upon the output from routine processing resources, or by direct input to them. Fodor was gloomy about the prospects for the study of central processes, maintaining that if they had the properties which he ascribed to them (e.g. non-modular, and not informationally encapsulated) they would be "bad candidates for scientific study" (Fodor, 1983, p. 127). With an unintended neuropsychological pun, he maintained that "…enthusiasm for a frontal assault on central processes…seems to have considerably abated" (p. 126). The latter statement hardly applies to neuropsychology (Fodor's point was made in reference to AI work). But what about the former? Are executive processes non-modular and, if they are, does this mean that they are bad candidates for study?

EXECUTIVE FUNCTIONS AND THE PROCESS-BEHAVIOUR[1] DISTINCTION

As an example of the assumed process—behaviour correspondence in the cognitive neuropsychology of routine processing, consider for instance Bruce and Young's (1986) model of face recognition. This model characterises well the neuropsychological (and cognitive psychological) findings from prosopagnosic patients; it does so largely by defining stages in face processing and recognition (e.g. "expression analysis") and assuming that, at least to a large degree, different cognitive processes support these stages. Thus a double dissociation between performance on tests of facial expression and identity (e.g. Etkoff,

1984) suggests that recognition of faces and expression analysis share little in terms of processing resources (see also McNeil & Warrington, 1993). The model (like others in the area) is based on studies involving close putative process—behaviour correspondence: the evidence for preserved "facial expression processing" comes from tests where, for instance, the participant is asked to judge facial emotion from photographs. The finding of preserved ability to recognise faces (and therefore intact "face recognition units," in Bruce and Young's terms) typically might come from asking the participant to recognise people from pictures of their faces. Here the processes thought to be tapped when the participant performs the behaviour of, for instance, responding to the question "tell me who this person is" are assumed to be highly dedicated to that situation: one does not suppose that they are tapped to nearly the same degree if the participant is shown a word and asked to define it. Thus we have a process (or a set of them) which is very closely dedicated to supporting behaviour in one narrow type of situation. This is fundamental to the notions of "informational encapsulation," and "routine processing" (specialisation of cognitive systems through repetition).

By contrast, executive processes theoretically manifest themselves in a range of quite different situations, the only unifying feature of which might be the involvement of that process. This is the rationale behind, say, Duncan's (Duncan, Burgess, & Emslie, 1995; Duncan, in preparation) contention that executive processes are synonymous with Spearman's *g*. If we accept this theoretical standpoint, a number of methodological and practical considerations follow.

Methodological Consequences of Low Process—Behaviour Correspondence

Increased "Measurement Error." Let us assume that any neuropsychological measure is subject to two primary sources of measurement error: the first is random error arising from factors which change from person to person and from one occasion to another (e.g. fluctuations in arousal or fatigue, external noise, interuptions, and so forth); the second is non-random involvement of cognitive

[1] Just in case there is any doubt, definitions of "process" and "behaviour" as used in this chapter are as follows: Cognitive processes are: (1) Event independent. They are not necessarily tied to any one behaviour and may be used in many. (2) Are not related, performance-wise, to the output of previous processes. (3) Are only assessed through measurement of a behavioural operation. They not directly observable. (4) The possibility of infinite reduction is a pragmatic and theoretical danger. (5) The possibility of parallel processing exists. Behaviours are: (1) Directly observable, event dependent. (2) Behaviours relevant to one goal are performed serially. (3) Even the simplest behaviour may require the activation of many processes. (4) Behaviours have an endpoint which is observer-defined.

processes which are not intended to be measured. This second source might occur, for instance, with the involvement of reading systems in performing Warrington's Recognition Memory Test for words, where the intention of the test is to measure mnestic functions, but where variations in reading ability may affect performance (Burgess & Shallice, 1994) in subtle and poorly understood ways. The first source of error is what is traditionally referred to in psychometrics as measurement error, but from the point of view of theorising in neuropsychology the second source is probably more relevant because it is non-random, and thus more troublesome for statistical interpretation.

This second type of measurement error in many areas of neuropsychological enquiry (e.g. semantic memory deficits) is generally ignored. It is assumed that a patient's ability, for instance, to give information about animals identified from pictorial representation is so highly indicative of some internal state that the possibility of measurement error can be safely disregarded. However, measurement of a control process, when that process only receives input from some slave process and outputs through another, will always be contaminated by effects attributable to these slave processes, quite apart from the error attributable to other sources such as process—behaviour translation, pure observation error and so forth.

Empirical support for this viewpoint is provided by data taken from the study in this volume by Wilson, Alderman, Burgess, Emslie and Evans. If executive tasks are contaminated by the influence of slave processes to a greater extent than slave processes are contaminated by each other, then performance on executive tasks should generally show greater correlations with indicators of slave processes than can be found between measures which are good indicators of supposedly "informationally encapsulated" resources, or at least between resources which are psychologically supposed to be quite distinct.

As part of the Wilson et al. study, certain traditional executive tests were administered to a group (N=94) of mixed-aetiology neurological patients, alongside a number of putatively non-executive tasks of memory, language, and perception.

As an example of an "executive task", let us consider performance on the Wisconsin Card-Sorting Test or WCST (number of categories achieved on the Nelson version). Two "non-executive" tasks one might choose which theoretically should share few direct processing resources with the WCST, and with each other, might be Warrington's (1984) recognition memory test (RMT) for words, and the Position Discrimination subtest from the VOSP (Warrington & James, 1991). The RMT for words requires the participant to make a judgment about 50 words presented one-by-one. They are, without delay, then asked to identify the target words when presented with a foil. The test is generally considered to measure memory ability for verbal material. Performance on this task should, *prima facie,* share little in the way of processing resources with the

Position Discrimination subtest of the VOSP. Here the participants are presented with two squares which each have a dot marked in them. They are asked to identify the square which has the dot in the centre, rather than just offset from it. This task has no obvious verbal or memory component and is generally regarded as a test of visuospatial or perceptual ability. The assumption that verbal memory and perceptual abilities share little in common should therefore be enough to predict strongly that there will be little relationship between these tasks in an unselected neurological population. And indeed this appears to be the case: the correlation between these tasks was found to be .141, which is not significantly different from zero. Moreover, neither test was significantly correlated with participants' age (RMT words, −.15; Position Discrimination, −.13).

But what about their relationship with the WCST? Performance on the WCST, unlike the other two tests, was significantly related to age (r=−.28, p<.01), to position discrimination (r=.36, p<.002) and was very strongly related to performance on RMT words (a surprisingly high correlation of .59, p<.001, which may support contentions that the WCST loads upon mnestic systems (see, for examples Corcoran & Upton, 1993; Kimberg & Farah, 1993).[2]

This kind of finding is not restricted to the WCST. It also holds, albeit to a less remarkable degree, for performance on another task with a strong executive component, Trail-making B (Reitan, 1971). In this test, participants are required to draw links between marked circles on a page in an alternating sequence (letter, number, letter…etc.). Here, although the correlation with RMT words was quite low (.18), the correlation with Position Discrimination was highly significant (r=.50, p<.005). Since it would be hard indeed to make the case that the Position Discrimination Test is greatly executive in nature, one is forced to conclude that the correlation between the two tests may well arise from some other (possible visuo-spatial) component they share.

These results do indeed support the view that many (if not most) executive tasks tap a range of processes incidental to their main purpose, and that this measurement error (probably more accurately referred to as "task impurity"; cf. Weiskrantz, 1992)—if one is using the tasks as measures of executive processing—is likely to be greater than in non-executive tasks.

"Task Impurity" and Patient Selection. The high degree of task impurity of executive measures has profound implications for the choice of participants for

[2] In fact Kimberg and Farah (1993) characterise failure on the WCST as being due to a weakening of "associations among the elements of working memory" (p. 422) caused by frontal lobe damage. Corcoran and Upton (1993) agree that the WCST makes demands upon working memory systems, but use this as an explanation for their finding that patients with hippocampal lesions performed more poorly than ones with frontal lesions. As regards the present argument, I know of no assertion that the RMT for words is a prototypical *working* memory measure, so will provisionally assume that the cause of the RMT—WCST association reported here lies elsewhere.

study, and explains in part the prevalence of localisation studies in this area compared with other areas of neuropsychology (e.g. reading disorders). In situations with high process—behaviour correspondence, it is a relatively straightforward matter to exclude the possibility of peripheral contributory factors to a person's performance: in the case of written sums, perhaps visual problems, dyslexia or whatever. However, situations which tap executive processes are usually more complex, thus increasing the difficulty of excluding possible contributions to performance by peripheral problems.

For example, a number of investigators have reported executive deficits in patients who perform extremely well on conventional neuropsychological examination (e.g. Eslinger & Damasio, 1985; Goldstein, Bernard, Fenwick, Burgess, & McNeil, 1993; Shallice & Burgess, 1991a). These patients only showed their deficits in complex situations such as real-life shopping tasks. Quite clearly, behaviour of this complexity must stress practically all cognitive systems in addition to executive ones. In order to decide that the problem such a patient demonstrates is selectively executive in nature, one needs to know the state of practically all other putatively non-executive systems. For instance, if a patient were to forget an item on their shopping list, is this the sort of failure a non-dysexecutive amnesic might make, or is it more a failure to realise a delayed intention? This is the nub of the low process—behaviour correspondence situation: similar observed behaviours can have quite different causes. In practice, the only patients who are likely to conform to the criteria of no peripheral impairment in the context of clear executive problems are those who have relatively isolated frontal lobe lesions. Therefore executive function researchers have tended to select patients on localisation grounds. The assumption is that when frontal lobe lesioned patients fail a particular task, they do so because of damage to the executive system; when posteriorly lesioned cases fail the task, they do so for "non-executive" reasons. A number of studies have directly examined the differences in behaviour between frontal and non-frontal patients (e.g. Shallice, 1996a, c; Incisa Della Rocchetta & Milner, 1993; Miller, 1984), but some assume that the link between localisation and behaviour is so strong that lesion control groups are unnecessary (e.g. Goel & Grafman, 1995; Janowsky, Shimamura, & Squire, 1989). Such selection methods are of course anathema to the ultra-cognitive neuropsychologists, and where they are used to circumnavigate the complicated issue of defining an "executive task" (which will be discussed below) as opposed to linking brain structures with a type of behaviour, it is hard not to be sympathetic to their viewpoint. It should be possible to discuss acquired executive disorders without reference to brain structures. Historically, however, the two have been so often associated that the terms "dysexecutive syndrome" and "frontal lobe syndrome" are currently synonymous. This is unlikely to remain the case (see e.g. Stuss, 1992).

Task Impurity and Convergent Evidence

A further potential problem exists for one particular type of design for executive function investigation where patients' performances in a range of situations are compared. Given the low process—behaviour correspondence in executive processing situations, the background demands (e.g. peripheral contributory factors) will differ greatly from task to task. Even quite minor changes in task demands may have considerable effect upon performance. This means that if one uses psychological rather than lesion location criteria for selection (i.e. choosing to study patients who fail a particular test regardless of lesion type or location) it is possible that many patients would be studied who are not failing the task owing to disruption of the executive component of the task, but because of damage to peripheral systems. Moreover there would be concern about the generalisability of the findings. In addition, "convergent evidence" will be difficult to achieve, with results apparently highly unstable across tasks (and possibly within tasks; see O'Carroll, Egan, & MacKenzie, 1994), owing to differing influences of peripheral contributory factors.

"EXECUTIVE FUNCTIONS" IS A THEORETICAL DEFINITION

The foregoing argument makes it clear that there is currently a fundamental problem with executive function research. If one wishes to study, say, calculation deficits, one might screen potentially suitable patients by giving them a calculation task, or a range of calculation tasks. This is because one is interested in studying people who cannot calculate. However, there is no equivalent for "executive functions," because the term is not an operational definition, but a theoretical one. There is no prototypical screening measure. This makes selection a difficult matter—whom should one study? A patient is dyscalculic if he or she cannot perform calculation tasks normally. A patient is dysexecutive if he or she cannot perform…what exactly? Many dysexecutive patients can perform any number of executive tests normally (see Shallice & Burgess, 1991a). Thus "screening" for dysexecutive problems in effect becomes full-blown investigation requiring many hours of testing and consultation with people who know the patient well.

In practice, as already discussed, executive function researchers have tended to get around this practical problem by selecting patients according to lesion. Since the link between executive dysfunction and frontal lobe lesions is well established, selection by this method sidesteps the issue of patient selection on psychological grounds. However, the subject of executive control will not be equivalent in theoretical terms to, say, the study of amnesia, until there are agreed empirical characteristics of executive disorder which will allow selection on psychological grounds alone.

This process has perhaps begun in detailed analyses of certain types of executive disorder: for example, confabulation (Burgess & Shallice, 1996b; Dalla Barba, 1993; Kopelman, 1987), perseveration (Goldberg, 1986; Sandson and Albert, 1984), and utilisation behaviour (Lhermitte, 1983; Shallice, Burgess, Schon & Baxter, 1989). Psychometric evaluation of such symptoms has, however, proved awkward. A classic dense amnesic, for instance, will always fail a certain range of psychometric (memory) tests. This is not the case for patients who show individual dysexecutive symptoms, for whom all manner of psychometric profiles are possible (Burgess & Shallice, 1994; Burgess, Baxter, Rose, & Alderman, 1996; Kopelman, 1995); and those cases who show a severe generalised dysexecutive syndrome do not show relative preservation of a wide range of functions as in classic amnesia (as of course one would expect with severe damage to system-wide control processes). Is it therefore possible to approach investigation of the executive system in a similar fashion to the study of other areas of neuropsychology?

THE PROBLEM OF COGNITIVE CONGRUENCE

One of the most well-replicated findings in psychometrics is that, in the undamaged brain, performance on virtually any cognitive task correlates positively with performance on any other. This is of course fundamental to the notion of *g* or general intelligence. Explanations for this finding range from the view that all tasks of any complexity share some specific information-processing resource (e.g. Spearman, 1927), through the view that *g* is a product of the sampling of overlapping but independent structural "bonds" (Thomson, 1939), to the contention that the finding is merely a product of the necessary biological development of the brain, and therefore has little to say at a psychological level. No direct view will be taken at present on the root cause of this finding, but the statistical fact itself will be referred to as "cognitive congruence."

Cognitive congruence is of course one of the primary reasons for the utility of neuropsychological study. However, it can cause considerable problems for those involved in the study of executive functions. Consider, for instance, the table of correlations given in Table 4.1a, which shows the interrelationships between performance on four executive tasks, plus age and NART performance, taken from the controls in the study of Wilson et al. (this volume). The results show a number of significant relationships between the executive measures. However, three of the four measures are also significantly correlated with irregular word reading (NART) performance. Since most neuropsychologists will have seen patients in their laboratory who show reading deficits in the context of, say, normal trail-making performance, or the converse pattern, we can assume that the significant correlations between NART and three of the executive measures reflect the cognitive congruence of the normal brain.

One appropriate consideration, therefore, is whether the relationship between the four executive tasks in controls is similarly merely a consequence of cognitive congruence, or whether the relationship between them reflects shared processing beyond *g*. Table 4.1b shows the table of correlations in controls once the influence for which NART performance is an indicator is removed by partial correlation for each variable. As the results show clearly, the previously statistically significant relationships between three of the measures (Cognitive Estimates, Trail-making B and Verbal Fluency) are completely removed by this manipulation. One significant relationship (that between WCST and Trail-Making), however, remains. One interpretation of these results is that NART is a good indicator of abilities across a range of neuropsychological domains in controls (including executive processing resources, cf.cognitive congruence), and thus statistically adjusting for the level of the *g* factor removes the relationships from the data. However, it would seem that for the relationship between WCST and Trail-making this argument does not hold. Thus one might conclude that the relationship between the WCST and Trail-making tests, whilst not particularly strong, is nevertheless greater than is predicted by the general congruence of cognitive abilities. However, it is not possible to be confident about the source of the relationships between the other tasks. The pattern of correlations here could result either from shared processing resources, or from mere biological happenstance.

This would seem to be an ideal question to address by studying brain-damaged patients. While the *g* hypothesis is not inconsistent with a continuing degree of cognitive congruence in such a group (the fluid intelligence, or hold-don't hold view; for a review see Rabbitt, 1993; Zimmerman & Woo-Sam, 1973), even the most ardent exponent of this position could hardly deny that neurological patients can show relatively isolated deficits in certain cognitive domains, such as perception, language, etc. If the executive tasks used here differ in the demands they make upon such (purely incidental) systems, which may be damaged randomly in a mixed-aetiology neurological group, one might expect less relationship between the measures than in the controls on these grounds alone.

Consider, however, Table 4.1c. The correlations between the executive measures are generally rather stronger than found in the controls. Indeed, the correlation between WCST and Verbal Fluency in the patient group is significantly higher than that found in the control group (r=−.14 in 74 controls, r=−.47 in 78 patients, p<.05). This is quite surprising, and at first sight would seem to support the *g* view (e.g. perhaps depletion in *g* by brain damage leads to greater variance in patients and stronger correlations), or at least a non-modular one (see e.g. Goldberg, 1995), given the relationships between background and executive measures outlined in the section above (on increased measurement error). This is the sort of result which might justify Fodor's gloom.

TABLE 4.1a Correlations between Executive Measures, Age and NART from the Controls from Wilson et al.

	Age	*NART*	*WCST*	*CEsts*	*FAS*
NART	.31**				
WCST[1]	–.29**	.18			
CEsts[2]	–.46***	–.63***	.24		
FAS[3]	.27*	.64***	–.14	–.40***	
TRAILS[4]	.06	–.51***	–.35**	.32**	–.33**

NOTE: *Ns* vary from 74 to 216.

TABLE 4.1b Correlations Between Controls After Removing Effects Attributable to NART Performance

	Age	*NART*	*WCST*	*CEsts*	*FAS*
NART	.31**				
WCST	.39***	.00			
CEsts	–.23*	.00	–.16		
FAS	–.07	.00	–.08	.01	
TRAILS	.33**	.00	.31**	.00	–.03

TABLE 4.1c Correlations between Executive Measures, Age and NART from Patients in Wilson et al.

	Age	*NART*	*WCST*	*CEsts*	*FAS*
NART	.28				
WCST	.19	–.22*			
CEsts	–.24	–.53***	.58***		
FAS	–.16	.48**	–.47***	–.59***	
TRAILS	–.01	–.46**	.38**	.56***	–.58***

NOTE: *Ns* range from 43 to 92.

Fortunately, the situation is not quite so grim. Consider Table 4.1d, which shows the correlations between tasks once NART performance is partialled out. Whilst the correlations between the executive measures are reduced somewhat, for 5/6 comparisons they are nevertheless significantly different from zero, supporting the rationale behind the clinical use of the NART. However, the key test is whether the relationships remain after partialling for a measure which is intended to be sensitive to brain damage rather than one (the NART) which is specifically designed to be insensitive. If they do not, the relationships may merely reflect acquired alteration in some non-specific "fluid intelligence" factor rather than a theoretically orthogonal executive one. In fact, partialling for performance on such a test (the Position Discrimination subtest of the VOSP) before correlation of the four executive measures still leaves four relationships

TABLE 4.1d Correlations from Patients After Removing Effects of NART Performance

	Age	*NART*	*WCST*	*CEsts*	*FAS*
NART	.28**				
WCST	.31**	.00			
CEsts	–.06	.00	.50***		
FAS	–.28*	.00	–.29*	–.44**	
TRAILS	.13	.00	.14	.44**	–.47**

*p<.05; **p<.01; *** p<.001.
[1] Modified WCST (Nelson, 1976), number of categories achieved.
[2] Cognitive Estimates Test (Shallice & Evans, 1978), error score. This test requires the participant to make estimates to questions such as "what is the weight of a full pint bottle of milk?".
[3] Verbal fluency for the letters F, A, and S (60s each; e.g. Miller, 1984). The participants are asked to say aloud as many words as they can that begin with a given letter within one minute. The score is the total of all words produced to three letters.
[4] Trail-making B section (e.g. Reitan, 1977; described in text).

which are significant at .05 or better (Cognitive Estimates and FAS fluency: WCST and Cognitive Estimates; Trail-making and Cognitive Estimates; Trail-Making and FAS). Moreover, three significant relationships exist even if one partials for WAIS FSIQ, which must be considered to have an executive component (Burgess & Shallice, 1996a; Cognitive Estimates and WCST; FAS and WCST; Trail-making and FAS). Thus it would seem that performances on the executive measures seem to share a greater relationship with each other than can be explained merely by cognitive congruence or than can be accounted for by the average level of performance across a range of non-specific tasks (e.g. the WAIS).

Not all commentators support this finding. For instance, Duncan et al. (in press) reports correlations between executive and non-executive tasks which are "at least as high" as those between executive tasks alone. Others have reported quite low correlations between executive tasks (e.g. Crockett, Bilsker, Hurwitz, & Kozak, 1986; Kopelman, 1991; Shoqeirat, Mayes, MacDonald, Meudell, & Pickering, 1990), although some have shown quite different patterns (e.g. Owen, Sahakian, Hodges, Summers, & Robbins, 1995; Perret, 1974; Wilson, Alderman, Burgess, Emslie, & Evans, 1996). The instability of these relationships appears puzzling at first. They become less so if one makes two assumptions. The first is that executive functions are potentially fractionable, and the second is that, following brain damage, some degree of cognitive congruence might still exist even when the absolute level of overall ability changes. Thus all patients will not show selective deficits in executive processes, and the relationships between psychometric tests in those that do not (and even between executive tasks not tapping the selective deficit in those that do) will reflect cognitive congruence,

tempered by the presence of selective non-executive cognitive deficits plus the usual sources of measurement error.

This view requires a notion of how cognitive congruence might exist in an essentially modular brain with acquired deficits. One simple analogy of the view taken here would be where all patients in a hospital, regardless of illness, are required to run 100 metres. One is interested in inferring from those with specific leg injuries the role of this organ in running. But the only available measure is the time taken, and one is not allowed to observe the race directly, nor know the pathology of the patients. Clearly as a group, the patients will perform more poorly than a control group of nurses, but how is one to decide which of the patients is relevant to the question of interest? After all, the patient with a fever will not be running slowly for the same reason as the patient with a broken leg. One piece of information which might help would be if each patient were also to swim 100 metres using front crawl. This style requires little leg-power, and so those with leg injuries would be generally less disadvantaged in this activity compared with the running event. Thus some calculation based on the degree of deficit relative to controls in the two events (and degree of deficit within patients) might now enable a decision about whom to study further.

Considered as a whole, however, the patients would tend to show "physical congruence" in performance across the two quite different events due to premorbid fitness, current drug regimen, general level of fatigue and so forth. Additionally, there is the knock-on effect of the particular illness for other physical systems. None of these factors need be directly relevant to the point in question. In this case, it is only those who show deficits relative to background level of physical performance (of which swimming ability might be one indicator) who are truly of interest. Applied to the current context, the view is that level of current "fluid intelligence" may be a useful background against which selective deficit may be defined, but the relationships between executive measures in patients without selective deficit must be approached with caution. Associations in patients with selective deficits are, however, potentially highly informative.

FRACTIONATION OF THE EXECUTIVE SYSTEM?

Is it possible, then, to find individuals amongst the group considered above who show selective deficits despite the group as a whole demonstrating congruence? Consider Table 4.2, which gives details of cases who show patterns of discrepancy where performance on at least one test is above the mean of a group of age-and NART-matched controls (*Ns* vary from 129 to 142) from the Wilson et al. study (this volume), and performance on at least one of the others is below the fifth percentile. Together the three patients would seem to show that performance on the WCST can dissociate from Cognitive Estimates, Fluency or Trail-making, and that Trail-making and Cognitive Estimates may dissociate. However, it could be argued that these dissociations arise because of specific

TABLE 4.2 Patients' Performance Discrepancies on Traditional Tests of Executive Function (%ile of Control Performance)

	Age	*NART*[1]	*FSIQ*[2]	*CEst*	*Fluency*	*WCST*	*Trials*
Case 1	76	116	106	84	8	12	< 0.1
Case 2	58	98	97	73	5	69	2
Case 3	35	92	89	66	62	< 0.1	66

[1] National Adult Reading Test WAIS FSIQ equivalent (Nelson & O'Connell, 1978).
[2] WAIS—R Full-Scale IQ.

impairments in peripheral processes. For instance, one might argue that the dissociation in patient 1 between Trail-making and Cognitive Estimates may be a function of motor slowing with age, combined with an improvement in performance on Cognitive Estimates with age (due to greater life experience, e.g. semantic knowledge). Thus this patient's pattern of performance may say little about the nature of executive function per se.

In order to avoid such complications Burgess and Shallice (1996c) sought to explore the possibility of dissociation in executive processes by using a task where the background demands of the task were kept similar. In the first part of the Hayling Sentence Completion Test ("initiation" condition) the patient is presented with a series of sentences which have the last word omitted and asked to provide the word which completes the sentence. In each case what this last word might be is strongly cued by the sentence frame. In the second part of the test ("inhibition" condition) the patient is required to produce a word which *does not* fit at the end of the sentence. A response which reasonably completes the sentence receives an error score of 3, a word semantically related to a word in the sentence receives an error score of 1, and an unrelated word a score of 0.

Patients who satisfied a number of criteria, including having a lesion (predominantly primary tumours) involving no more than two lobes, were tested. Patients whose lesions involved the frontal lobes (anterior group) and those with purely posterior lesions were closely matched on age (means 45.1, 43.0), NART IQ (Nelson & O'Connell, 1978; (109.3, 111.5) and FSIQ (102.6, 106.1) Twenty controls matched on age (49.7) and NART IQ (112.0) were also tested.

There were two main findings. First, the anterior group showed significantly longer response latencies on straightforward completion of the sentences than either the controls ($p<.002$) or the posterior group ($p<.05$), although the posteriors did not differ significantly from the controls. The second finding was that in the second part of the test (inhibition) patients with anterior lesions had a much higher average error score than either the posterior group or the controls. There was no interaction with hemisphere. In other words, the anterior group found it much more difficult not to produce the triggered response than did the other groups.

Most relevant to the present argument, however, was the finding of extremely low correlations between the patients' performances on the initiation and inhibition sections of the Hayling test. Overall the correlation was .19, which is not significantly different from zero. Moreover, partialling for age and FSIQ reduced this figure to .07. Individual case scores confirmed this potential independence. Calculating the significance of the individual difference scores based on the split-half reliabilities of the test showed that performance on the two sections of the test may doubly dissociate (patient 1: initiation first percentile, inhibition 76th percentile; patient 2: initiation 66th percentile, inhibition 0.1 percentile). Interestingly, the only patients who showed significant dissociations had anterior lesions (although there were more anteriorly lesioned cases under consideration).

There is, however, a complicating factor for this argument. In the inhibition condition, it was noted that participants often used a strategy to generate anomalous words. For instance, the most common was to look around the testing room and name objects within view. Examination of the patients' responses revealed that the anterior patients showed less evidence of having used such a strategy. Moreover strategy use and sentence completion (e.g. task failure) were highly, and inversely, related. It was argued that, in this task, response suppression and strategy use are mutually causally related: a patient will require some response suppression abilities in order to utilise a strategy. However inability to formulate a strategy will result in a greater load upon response suppression abilities.[3] If such reciprocal causation is a characteristic of the operation of the executive system, this may have considerable methodological implications: understanding such links generally requires large data sets, and collecting such sets from one individual may prove difficult for reasons which will be discussed later. However, reciprocal causation may only be a characteristic of the Hayling test. Is there any evidence that such causal relations may exist in other situations tapping executive systems?

[3] This study was a good lesson in not prejudging the demands of an executive task: The validity of terms such as "planning" or "response suppression" has not yet been established, nor have the demands of many executive tasks been examined in detail. It is therefore likely that recent findings (e.g. Delis et al., 1992; Della Malva et al., 1993; Goel & Grafman, 1995; Karnath, Wallesch, & Zimmerman, 1991), suggesting that our present characterisations of what executive tasks measure are quite wrong, will be extended. These are welcome developments beyond the present "labelling" phase in executive function research.

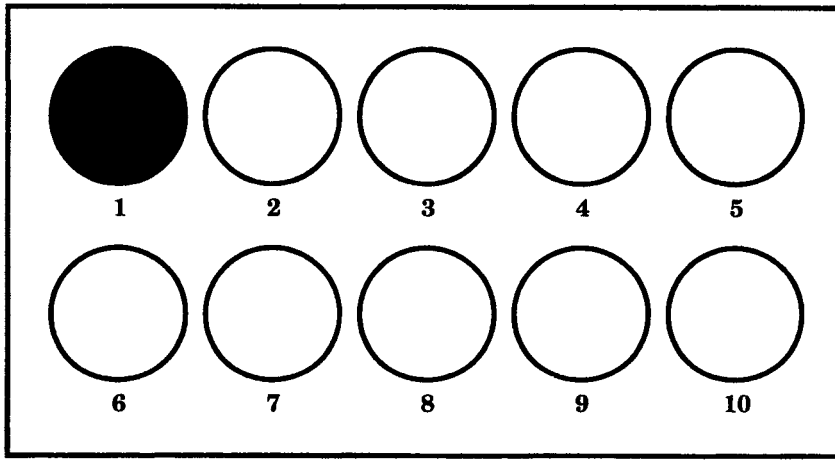


FIG. 4.1. Brixton test stimulus, position 1.

RECIPROCAL CAUSATION AND EXECUTIVE FUNCTIONS

In fact, there is evidence that reciprocal causal links are probably not restricted to the Hayling test, but may be more widely characteristic of situations which tap executive processes (see e.g. Della Malva, Stuss, D'Alton, & Willmer, 1993, and Roberts, Hager, & Heron, 1994, for similar findings within quite different paradigms). Indeed, Burgess and Shallice (1996a) found a specific link of this nature in the most characteristic situation where dysexecutive patients can have difficulty: set attainment. They administered a test which consisted of a series of plates similar to those shown in Fig. 4.1. On each plate, one of the 10 positions was filled, but the position of the filled circle changed with each trial. The changes in position were governed by a series of simple rules (e.g. +1 or −1,) which changed without warning, and the patients were charged with the task of predicting the filled position on each subsequent trial. Thus if the filled position had previously followed the sequence 1, 2, 3, the patient was required to say that the filled circle on the next plate would be at position 4 (+1 rule).

This task (called the Brixton Test) was administered to a group of patients with focal lesions which involved the frontal lobes, a group of posteriorly lesioned patients matched for age and current WAIS FSIQ, and a group of similarly matched healthy controls. In accordance with previous findings (e.g. Cicerone, Lazar, & Shapiro, 1983; Milner, 1963; Nelson, 1976), the patients with anterior lesions were poorer at set attainment (e.g. performance on the test overall) than the posterior group or the controls. However, when the nature of the anteriors' responses was analysed, they did not show a tendency to perseverate (thus adding to growing evidence (e.g. Corcoran & Upton, 1993;

Delis, Squire, Bihrle, & Massman, 1992) that perseveration is not necessarily a primary characteristic failure of dysexecutive patients). Instead, they showed an abnormally high incidence of bizarre responses (i.e. responses for which no apparent rationale could be discovered, and which normal individuals did not make). They also showed an exaggerated tendency to abandon a correct rule once it had been attained. Critically for the present argument, incidence of bizarre responding and overall set attainment performance were very highly correlated. It was argued that, in this situation, not being able to attain set might lead to pseudo-random guessing (since the positions of the stimuli would appear apparently random to someone who had no idea of the underlying rules). However, if the predominant deficit in this group was a propensity for wild guessing, this would interfere with their engagement in the task and make set attainment more difficult. Thus, at least in this situation, set attainment problems and tendency towards bizarre responding seem necessarily linked, even though they may be theoretically orthogonal.

It is equally possible, however, that they are not. Take the first example of the reciprocal link between response suppression and strategy use. What is response suppression for? The traditional view is that (if "response suppression" exists at all as an outward sign of a cognitive event) it serves to prevent an automatic response in situations where that response would not be optimal. Thus it serves to *enable* a behaviour to appear. In the case of the Hayling test it may serve to facilitate strategy use. Conversely, what is strategy use for? In many situations it must serve to lessen the "cognitive load" of a task. In the case of the Hayling test, the cognitive load is the response inhibition demands of the task. Even if all processes underlying performance of executive tasks are theoretically fractionable, the key to understanding behavioural adaptation (surely the purpose of executive processes) may nevertheless rest with understanding such reciprocal causation: Goethe's famous comment that dissecting a fly and studying its parts will not tell you how it flies could almost have been intended for the neuropsychology of executive function.

Dissociations Between Processes or Dissociations Between Operations?

There is a further complicating issue as regards the potential fractionation of the executive system: Processes may be theoretically fractionable, but operational constraints may prevent this discovery. Norman and Shallice's five situations requiring the use of executive control, described previously, take for granted a response history: the person is in state A and wants to get to state B. The executive processing allows the person to achieve this and can only be understood by examination of the translation from state A into state B. But these two states do not exist in some imaginary problem space where physical and temporal laws do not exist: the actions which enable state A to state B translation

have to be performed in a certain sequence. Thus if, for instance, the processes underlying planning, and those that enable a person to effect a self-generated plan are theoretically separable, they will not be empirically: one cannot fail or succeed in carrying out a plan one has not made.

Consider for instance the model presented in Fig. 4.2. This attempts to explain the relationship between the stages in performance of a non-routine behavioural sequence which requires the realisation of a delayed goal (or "prospective memory"), and is based upon the behavioural observations of Shallice and Burgess's three patients with Strategy Application Disorder (Shallice & Burgess, 1991b; see also Goldstein et al., 1993). According to this account, the realisation of a delayed intention operates through the setting of "markers", which are messages to treat some aspect of an anticipated future situation as non-routine. "Planning" in such a situation is a multistage operation, the end purpose of which is to set up these markers, which are hopefully then activated at the appropriate future time (see Burgess & Shallice, in press, for a fuller account of how this procedure might operate). Following marker activation, a series of reevaluations are performed, if necessary, before action initiation. Clearly, although the behaviours being performed at each stage might appear quite different, the processes underlying the behaviours might be shared to a high degree. But how is one to test this hypothesis? The most obvious test would be a finding of double dissociation between, say, marker creation and marker activation. However, where one behaviour is dependent upon another, a true double dissociation may never be achievable even if the supporting processes are orthogonal. Thus the results gleaned from the study of complex dependent behavioural sequences will always tend to militate against fractionation, at least (as we shall see) as far as the study of single cases is concerned.

THE STUDY OF COMPLEX BEHAVIOURAL SEQUENCES

If these temporal and physical constraints in the study of complex situations are not enough, Kimberg and Farah (1993) present another difficulty. They isolate as one of the main reasons why the nature of the executive system is still poorly understood in neuropsychology the fact that executive processing[4] "is called into play only when the activities of multiple components of the cognitive architecture must be coordinated" (p. 422). Thus if a methodology is used where a task is broken down into its component parts, no deficit will be discovered in dysexecutive patients.

This is an important point, but it may be possible to use methods which study the contributions of component parts while leaving the overall structure together. For instance, in a recent study (Burgess & Taylor, in preparation), a group of 50 brain-damaged patients and 31 healthy controls were given a simplified version of the Six Element Test (Shallice & Burgess, 1991a). This test requires

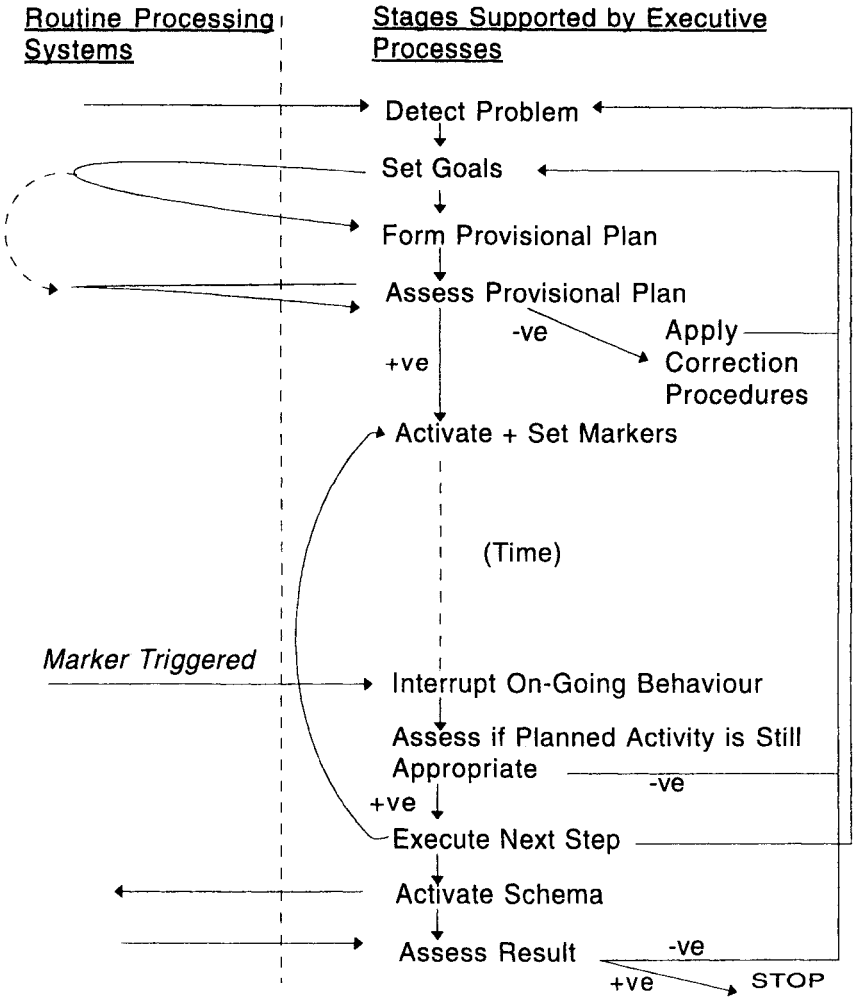


FIG. 4.2. Behavioural stages involved in coping with a novel situation requiring the realisation of a delayed intention according to Shallice and Burgess, 1991b. Vertical lines indicate the succession of stages, and horizontal lines indicate the flow of information between routine and non-routine processing (adapted from Burgess & Cooper, 1996).

participants to schedule the performance of three different types of tasks (dictation, arithmetic and written picture naming), each of which is split into two parts (A and B). They are told that they are required to attempt at least part of each subtask within a 10-minute time period, but that they are not permitted to attempt parts A and B of the same type consecutively.

Clearly failure on a task of this complexity might happen at a number of stages. For instance, the subject may fail to form a plan before attempting the

task, or fail to follow it subsequently. Moreover, failures might occur due to damage to peripheral systems; poor auditory comprehension abilities, for instance, might lead to poor rule learning (and consequently a poor plan). In order to investigate the relationships between the processing supporting these different operations involved in the SET, without having to "pull apart" the task itself (and thus fall foul of Kimberg and Farah's objection), the SET was administered to the participants in a novel fashion. In addition to the normal test administration procedure, the participants were assessed on their ability to learn and understand the rules, directly and indirectly, before commencing the test ("rule learning"); the complexity and detail of their plan of how they intended to attempt the test ("planning"); how closely they followed their plan subsequently in performing the test ("plan following"); how accurately, after performing the test, they were able to recount what they had done ("monitoring"); and their recall, cued and spontaneous, of the task rules after the test was completed ("retrospective memory"). Background neuropsychological testing confirmed that the unselected neurological group included patients who had relatively isolated impairments in peripheral systems (e.g. memory, perception, language), thus enabling an examination of how these systems interacted with executive ones in performance of the task overall. (Ultra-cognitive neuropsychologists might like to note that this group study relies on the *heterogeneity* of the group; not all group methods make the homogeneity assumption; see e.g. McCloskey, 1993.)

Using structural equation modelling (for an exposition of this method as applied to neurophychology, see Francis, 1988), a simple two-factor model was applied to the data, where rule learning, planning, monitoring score and retrospective recall of the task rules were hypothesised to be loaded upon factor one. A second factor, which is specifically executive in nature, was hypothesised to be linked to plan-following ability, and overall performance on the SET. The largest point biserial correlations between group and the other variables were with SET performance and plan following. Therefore group was allowed to be an indicator for this second factor. The overall model was a good fit to the data (2 (14)=17.94, p=.21; GFI=.99; AGFI=.99), and accounted for the data markedly better than a single-factor model. Thus this study supports the view that many apparently different operations in the performance of complex tasks are tightly bound up with each other. Even where theoretical processing independence may be demonstrated, mutual dependency (and reciprocity) in the actual behaviour

[4] In fact Kimberg and Farah (1993) maintain that there is no entity which could be termed an executive system. Instead they suggest that the weakening of associations among elements of working memory leads to deficits which "may have the appearance of an executive dysfunction but…are a consequence of damage to the available knowledge representation and not to executive control structures" (pp. 422–423).

hierarchy may make these distinctions empirically opaque as regards understanding the processing demands made by different task situations. For instance how can the formulation of a plan be doubly dissociated from the understanding of task rules? Those who cannot understand the task parameters will inevitably devise poor plans. Task comprehension and planning may well be theoretically separable functions, but how could one measure this empirically when one behaviour is so dependent on the results of previous cognitive processing? A mere single dissociation is the most that could be hoped for.

Mutual dependencies of this type may lead to failures at different stages in the behavioural "tree" (see Fig. 4.2) having differing strengths of "knock-on" effect, thus mimicking some *g* factor in a potentially fractionable system. However, it was possible to support the view that at least as regards the performance of the SET, processes which underlie ability to follow a plan are separate from, but linked to, those which support other behaviours. These processes appeared to be especially vulnerable to neurological damage. Overall, the data supported the notion of a separate executive processing system which underlies plan following (i.e. prospective memory) behaviour in complex situations (see also Cockburn, 1995; Einstein & McDaniel, 1990).

ANALYSIS OF BEHAVIOURAL SEQUENCES WITH FEW STRUCTURAL CONSTRAINTS

The example just outlined is one where, although the processes being stressed at any one point may cross boundaries between stages in the behavioural sequence (e.g. "planning" or "rule learning"), there is some temporally determined structure to the order of the different stages (e.g. one cannot report in retrospect—the "monitoring" score above—what one has not yet done). Many everyday strings of behaviour are of course subject to such constraints, and thus proceed, roughly, in a given order. These kinds of constraints are particularly useful to those who might want to use methods such as structural equation modelling because they provide limits on theorising about causal relationships.

Some behavioural sequences, however, enjoy more freedom in their manifestations. Indeed one possible definition of how much a given situation might stress executive processes would be the degree to which the behavioural sequence can be interpreted by the individual in pursuit of a particular goal (which is part of Shallice & Burgess's, 1991a, explanation for the failure of their patients only in relatively open-ended situations). In such situations behaviour may tend to follow an overall structural framework, but individual patterns of behaviour may differ markedly from one occasion to another.

One such situation appears to be where participants are recalling events which have happened to them. In a recent study (Burgess & Shallice, 1996b), the manner in which normal individuals undertake autobiographical recollection was examined in detail in order to determine whether confabulation in neurological

patients (one of the most striking forms of dysexecutive problem) might conceptually be accounted for as a breakdown in recollection control processes. If such control processes might be found at work in normals, an account might be developed regarding potential consequences of their failure, analogous to the manner in which action lapses (see e.g. Reason, 1993) in normals can be related to utilisation behaviour in patients (see Shallice, Burgess, Schon, & Baxter, 1989).

Eight normal participants were asked a series of 14 questions such as "describe the last time you had dealings with the police" or "when was the last time you cleaned your car?". The participants were instructed to describe their thought processes and the memories/images that came to mind as they answered. If possible each new idea was to be given a single-word label only. This "voicing aloud" was tape-recorded. They were asked to continue for up to one-and-a-half minutes for each question, although no effort was made to enforce this guideline. After they had finished an answer the tape-recording was replayed so that they could elaborate their brief labels and comment on what they had said, especially as regards its accuracy. The original answers and the commentary were then transcribed word-for-word for each participant. This procedure was employed so as to reduce as far as possible the interference that producing a protocol necessarily entails in the primary task, and so as to allow participants to check the accuracy of their initial recall process.

The transcripts were used to investigate the hypotheses put forward by a number of memory theorists (e.g. Conway, 1992; Morton, Hammersley, & Bekerian, 1985; Norman & Bobrow, 1979; Williams & Hollan, 1981) who maintain that recollection involves stages of forming a description for retrieval, followed by post-retrieval verification procedures. Pilot studies had shown that various sections of the verbal recollection protocols contained elements which corresponded to these processes of description and verification, as well as those which are aimed at resolving impasses or memory failures.

The verbal protocols were split into small sections each of which was then independently rated by two judges as belonging to one of 25 different categories of element type. There was satisfactory agreement between the judges. The 25 element types could be grouped into four different broad categories of types of process.

The first type were memories themselves; the second were memory editing processes (where the person is assessing a retrieved memory for accuracy/suitability or giving notice of recall failures or confusions); the third related to the formation of general descriptions or hypotheses about what is to be recalled; the last category consisted of pure strategic, problem-solving elements, termed "mediator elements."

By examining the frequency with which a given element was followed by another across all the participants' protocols, it was possible to calculate those combinations which occurred more frequently than chance. These data were then

used to form a model of the most common forms of recall structure. The results indicated that, for instance, elements linked to the verification procedures often directly preceded or followed the retrieval of a memory, and these memory elements in turn tended to appear with greatest frequency towards the end of recollection. "Description" elements, however, were less closely linked to memories (that is, there was often at least one other element between a description and a memory) and appeared with especially high frequency in the early stages of the participants' answers to the questions. The mediator elements were particularly interesting. These elements are evidence that recollection sometimes involves inferential reasoning as well as frank problem solving. For instance it was not uncommon for participants to answer the question "what was the weather like yesterday morning?" by trying to remember first what they were wearing, thus giving a hypothesis which they could "try out" on their memory store. On other occasions, they made direct calculations about, say, the timing of a particular episode in a more straightforward fashion. Other mediator elements (which tended to be furthest away from memories in the recall structure, occurring primarily at the start of autobiographical recollection of an episode) included meta-cognitive judgments of the difficulty of the retrieval task before recall had started ("the last time I cleaned my car…*Oh, this is going to be difficult…*let me think") or notice that a particular memory is not available, but is likely to be in the future. In summary, autobiographical recollection did not appear to consist of instant, easy access to a memory, which was then recalled in its entirety and without error. Instead there was a general tendency towards the use of strategic and problem-solving procedures at the beginning of recall (when the memory was not yet available), followed by the formation of precise specifications (descriptions) for recall which enabled the recall of candidate memories. Output was then subject to a range of checking and error correction procedures at each stage. The overall relationship between elements from these different stages is shown in diagram form in Fig. 4.3, with examples of element structures appearing in this figure given in Table 4.3.

That these control processes (descriptions, editing, and mediations) are necessary in recollection appears to be a direct consequence of the way event memories are represented. Recollection appears not to proceed by the activation of an invariant "record" of an event; the process is more one of deliberate reconstruction, as many authors have underlined (see Barsalou, 1988; Bartlett, 1932; Conway, 1990,1992; Norman & Bobrow, 1979; Reiser, Black, & Kalamarides, 1986; Williams & Hollan, 1981).

Any such reconstruction must of course enforce the same empirical constraint as was described earlier in the context of the Six Element Test: the results of processing at one stage are dependent upon the results of processing at an earlier stage. Overall, the complexity of the processing stages involved in recollection was emphasised by some interesting phenomena shown by participants in this

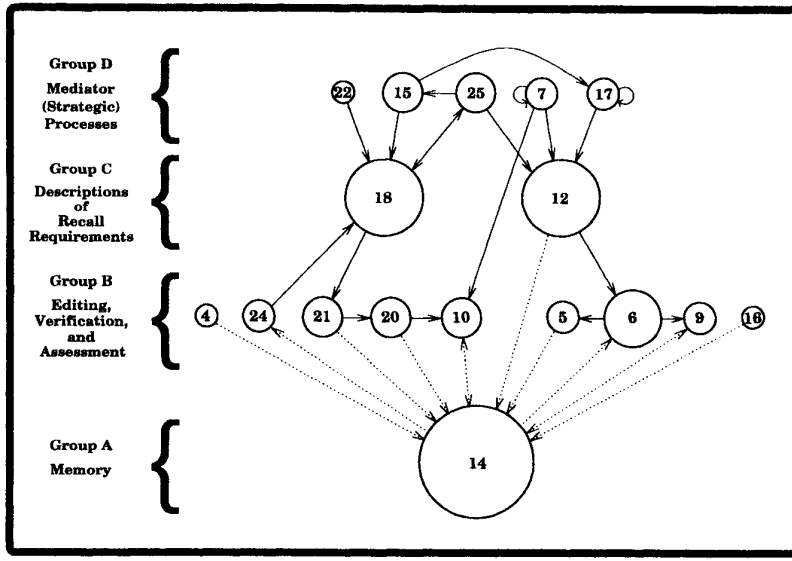


FIG. 4.3. Diagrammatic representation of protocol element sequential dependencies (se Table 4.3 for examples of individual elements; full description is given in Burgess & Shallice, 1996b. Reproduced by permission of Psychology Press).

study. They quite often made errors in recall, with relatively common mistakes being the incorrect insertion of personal semantic information in a memory, and the conflation of memory experiences (where two often similar events are mixed up). Moreover, at times participants were aware of not being able to remember certain details of an event which were retrieved later, or thought that they had recalled all the elements of an event when in fact they had not. Burgess and Shallice suggest that disorders such as confabulation can be explained as a failure in the memory control processes, leading to an increase in incidence of these "normal" mistakes, plus some which are rarely seen in intact individuals (a theoretical account of these findings is given in Burgess & Shallice, 1996b).

The "chaining" method of representing behavioural sequences as used here is pragmatically quite useful—if one were to collect enough data from a single individual this might suffice for some purposes. However it has its limitations: the model presented in Fig. 4.3 only represents single-order relations, for instance. Given one element, one is able to represent the most likely following element, but the structure (as outlined here) is much less true of conditional sequences further afield: such as, given element X is followed by element Y, what is the likelihood of element Z? Since the values between elements within such relational structures are unlikely to remain static throughout a response history such matters might come to be important. I know of no method currently in use in neuropsychology which can accommodate such dynamic changes in

TABLE 4.3 Element Strings from Actual Protocols

Element	*Code*[1]	*Protocol Segment*
Q: When did you last see a relative you don't live with?		
Yet Unavailable	25(D)	"... I can't actually think of a specific time—
Hypothesis	12(C)	I think the last person I must have seen would've probably been dad,
Correction Poss.	6(B)	but I could be wrong—
Correction	5(B)	no—it must've been mum
Memory	14(A)	at Jane's birthday."
Q: Describe an object you have recently given or received.		
Familiarity Imp.	10(B)	"[at the] ... beginning of this week
Memory	14(A)	I bought her some flowers
Verification	24(B)	I suppose that's, that's giving her a present
Memory	14(A)	Just on the way back from work."
Q: What was the weather like yesterday morning?		
Familiarity Imp.	10(B)	"But I think I have an impression
Memory (incorrect)	14E(A)	of the weather being stormy.
Correction Poss.	6(B)	But somehow that doesn't ring true."
Q: When was the last time you went to the coast?		
Recall Spec.	18(C)	"... aah, let me think away here ... um ...
Success Point	21(B)	yes, I think I've got it.
Memory	14(A)	Its Barclay's Bank down in Poole."

[1] This refers to the numerical code assigned to a given protocol element, as used illustratively in Fig. 4.3.

structural relations. It seems likely that our future understanding of the operation of executive control will be markedly constrained by the lack of such methods.

Returning to the issue of the limitations of studying one individual in executive function research, a comparison of participants' responses to the questions in the present study revealed considerable variation in individual styles. Thus although the general structure outlined in Fig. 4.3 held for each participant in the sense that when element X appeared it tended to be followed by element Y, the proportion of elements from each category varied markedly between participants. For instance, the proportion of overall elements that were memories for participant 1 was 35.3%. However the incidence was almost twice as many for participant 2 (65%). Similarly 25.4% of the first participant's spoken elements were of the "mediator" type (problem solving), while for the second only 4% of total elements were from this category. It is clear that while overall relational structures might exist in complex behavioural sequences, individual response styles may quite markedly alter incidences of certain types of behaviour. It would be quite wrong, for instance, to have concluded from the

study of participant 1 that about 35% of spoken elements in autobiographical recall tend to be memory elements. Clearly their share may be much greater. Such matters of proportional (as contrasted with absolute) incidence tend to be of particular significance in the study of executive functions. Furthermore, it is probable that in the study of behavioural sequences with low structural constraints the most informative aspects will lie in the interrelationships between behaviours. Data from any single individual could be misleading in this respect.

Single Case vs. Group Study Design

Many aspects of the nature of executive functions already discussed have considerable implications for the choice of whether to use single-case or group study. There is no doubting the theoretical power of the single case who shows a particularly striking form of loss of executive control. However, the nature of the executive system—if it is anything like the characterisation presented here—presents some unique methodological and theoretical problems which may often be best addressed by the study of more than one individual at a time. Outside the few patients who show striking dyscontrol syndromes such as confabulation, utilisation behaviour and marked perseveration, most patients with executive problems overlap with the normal distribution in their performance on the measures in question. The effect exerted by executive control, as suggested by Shallice (1988), is often quite subtle, and the variance attributable to individual differences in peripheral functions can swamp that attributable to executive dyscontrol. In the study of, for instance, language disorders, many patients may be found where the severity of the acquired problem is never (or extremely rarely) observed in the normal population. This is not often the case with executive disorders. It is only when aspects such as cognitive congruence and task impurity are taken into account that the subtle effects can be detected. Traditionally these problems have been solved by using group study design, or at least small groups of single cases with case-by-case control matching (e.g. Duncan et al., 1995; Shallice & Burgess, 1991a) although (given the appropriate patient) there is no theoretical reason why these matters could not be similarly addressed at the single-case level. However, there is the matter of understanding reciprocal causation in the executive system, and the discovery of processing demands which may support seemingly quite different behaviours. Single-case design of the form currently used in cognitive neuropsychology is unlikely to be a sufficient method to answer these more global questions about the executive system. However, there is no need to use group design in such a manner as to fall foul of two of the "ultra-cognitive" neuropsychologists' principal objections to such methods: the problem of averaging artefact and assumptions of homogeneity. The single point above all others that makes the study of executive functions different from other areas of neuropsychology is, for the reasons outlined above (e.g. task impurity, individual response styles, causal dependence

in complex behavioural sequences) dissociations need be no more instructive than associations. Often they may be less so. In addition, there are specific matters relating to the use of single cases if one makes the assumption that the executive system is most taxed in novel situations.

PROBLEMS WITH MEASUREMENT OF BEHAVIOUR IN NOVEL SITUATIONS

A. Variation in Novelty. The degree of novelty across tasks will vary with every patient. Thus in any single case it is extremely difficult to know how much a task is tapping the executive process of interest. Moreover, one's findings even across a series of single cases are likely to inform understanding of only that specific, single situation. While it may be interesting to know exactly why a particular patient fails, say, the WCST, there is the possibility that the results would only be true of that single patient, in that particular situation.

B. Measurement of Behaviour in Novel Situations is Like Shooting a Moving Target. The cognitive neuropsychologist who is studying calculation deficits can be reasonably sure that if on day 1 the patient cannot do addition sums, for instance, then barring some change in neurological state the patient will still be unable to do addition sums on day 2, or 3. This is not true for tests which measure processing in non-routine situations. By definition novelty decreases with repetition. Therefore the construct validity of a test whose purpose is to measure adaptation to novelty reduces with every testing. This is undoubtedly one of the key obstacles to our understanding of executive functions, and no doubt explains much of the "riddle of the frontal lobes," as Teuber (1964) called it: the purpose of non-routine processing is to make itself redundant. Thus one might learn from repeated testing of a single case how automaticity occurs in a given situation, but the purest method of studying individuals' behaviour in novel situations is to observe a large group on their first exposure (to one test, or to a range of them as in factor analysis) and then attempt to extract general principles from their behaviour (see e.g. Karnath, Wallesch, & Zimmerman, 1991).

C. The Variety of Response and Strategy Possibilities. If a prototypical situation in which executive processing is particularly required is where an impasse or decision point has been reached, or where an adaptation or suppression of routine schemata is indicated, this will typically be where a number of response options or strategies are available. On any particular occasion an individual can only produce one of these response options or use one of the possible range of strategies. On another occasion the same person might choose quite a different approach—a situation which should be quite familiar to most clinicians dealing with dysexecutive patients (see Shallice & Burgess, 1996, for further exposition on this point). Thus the performance of a single case on just one occasion is of limited value in understanding how impairment can affect

behaviour. Yet repeated testing of the same individual is of course confounded by the issue of novelty. Data from a large number of participants studied on their first trial might however approximate the range of potentially observable behavioural operations.

CONCLUSION

This chapter is not intended as a rejection of cognitive neuropsychogical case study method in the investigation of executive functions. Quite the contrary—it has been argued that some matters are best investigated in this way. However, it seems likely that the particular demands and restrictions of the subject area will require the adaptation of methods relatively new to neuropsychology. Some problems seem almost insurmountable given current methodological technology, and must await new developments.

The view of the executive system presented here is that control processes are potentially fractionable, even though the system is by its nature not "informationally encapsulated". The difficulty, however, lies in the interface between process and behaviour in complex sequences. I have adopted the widely accepted view that the executive system is particularly involved in the coordination of complex novel behavioural sequences. However, these sequences have to be carried out in certain orders owing to temporal and physical constraints which probably bear no relation to the functional organisation of the executive system. Since in any behavioural sequence, failure at any stage has knock-on effects for later behaviour, damage to processes supporting behavioural operations early in the sequence may lead to an impairment in overall performance which is in some sense disproportionate to the process damage. Other behavioural sequences undertaken by the individual might require the damaged process further down the behavioural sequence. In this case the apparent behavioural impairment (as evidenced by the end-point result) will be less severe.

In any one individual the degree of behavioural impairment will be the result of a complex interaction between the exact sequential demands of the situation, the particular processes which are impaired, and the novelty of that situation. In addition, there is the matter of the person's expectations and prior experience, quite independently of the degree of novelty. Moreover, the patient's response to an identical situation may never be identical: overall, it is argued that the function of the executive system is to facilitate adaptation by making itself redundant. If the account given here is even partially true it is little wonder that there is so much variety in the manifestation of executive system damage, and that these patients seem so variable in their behaviour. The study of executive functions makes special practical and theoretical demands, and the present challenge is to develop methods to deal with them.

ACKNOWLEDGEMENTS

I would like to thank Tim Shallice for his helpful criticisms of an earlier draft, and my colleagues Barbara Wilson, Nick Alderman, Jon Evans, and Hazel Emslie for letting me analyse data collected as part of the study they describe in this volume. Preparation of this chapter was supported by grant number 38964\Z \93\1.5 from the Wellcome Trust.

REFERENCES

Alderman, N. & Burgess, P.W. (1994). A comparison of treatment methods for behaviour disorder following herpes simplex encephalitis. *Neuropsychological Rehabilitation, 4,* 31–48.

Alderman, N., Fry, R.K., & Youngson, H.A. (1995). Improvement in self-monitoring skills, reduction of behaviour disturbance and the dysexecutive syndrome: Comparison of response cost and a new programme of self-monitoring training. *Neuropsychological Rehabilitation,* 5, 193–221.

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W.Spence (Ed.), *The psychology of learning and motivation.* New York: Academic Press.

Baddeley, A.D. & Wilson, B.A. (1986). Amnesia, autobiographical memory and confabulation. In D.C.Rubin (Ed.), *Autobiographical memory* (pp. 225–252). Cambridge: Cambridge University Press.

Baddeley, A.D. & Wilson, B.A. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7,* 212–230.

Barsalou, L.W. (1988). The content and organization of autobiographical memories. In U. Neisser & E.Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to memory* (pp. 193–243). Cambridge: Cambridge University Press.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Benton, A.L. (1991). The prefrontal region: Its early history. In H.S.Levin, H.M.Eisenberg, & A.L.Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 3–32). New York: Oxford University Press.

Bruce, V. & Young, A.W. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Burgess, P.W., Baxter, D., Rose, M., & Alderman, N. (1996). Delusional paramnesic misidentification. In P.W.Halligan & J.C.Marshall (Eds.), *Case studies in neuropsychiatry* (pp. 51–78). Hove, UK: Psychology Press.

Burgess, P.W. & Cooper, R. (1996). The control of thought and action. In D.Green (Ed.), *Cognitive science: An introduction* (pp. 340–367). London: Blackwell.

Burgess, P.W. & Shallice, T. (1994). Fractionnement du syndrome frontal. *Revue de Neuropsychologie, 4,* 345–370.

Burgess, P.W. & Shallice, T. (1996a). Bizarre responses, rule detection and frontal lobe lesions. *Cortex, 32,* 241–259.

Burgess, P.W. & Shallice, T. (1996b). Confabulation and the control of recollection. *Memory 4,* 359–411.

Burgess, P.W. & Shallice, T. (1996c). Response suppression, initiation and frontal lobe lesions. *Neuropsychologia 34,* 263–273.

Burgess, P.W. & Shallice, T. (in press). Autobiographical recollection and prospective remembering. In M.A.Conway and S.E.Gathercole (Eds.), *Cognitive models of memory.* Hove, UK: Psychology Press.

Burgess, P.W. & Wood, R.L. (1990). Neuropsychology of behaviour disorders following brain injury. In R.L.Wood (Ed.), *Neurobehavioural sequelae of traumatic brain injury* (pp. 110– 133). London: Taylor & Francis.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition, 5,* 41–66.

Caramazza, A. & Badecker, W. (1991). Clinical syndromes are not God's gift to cognitive neuropsychology: A reply to a rebuttal to an answer to a response to the case against syndrome-based research. *Brain and Cognition, 16,* 211–226.

Cicerone, K.D., Lazar, R.M., & Shapiro, W.R. (1983). Effects of frontal lobe lesions on hypothesis sampling during concept formation. *Neuropsychologia, 21,* 513–524.

Cockburn, J. (1995). Task interruption in prospective memory: A frontal lobe function? *Cortex, 31,* 87–97.

Conway, M.A. (1990). *Autobiographical memory: An introduction.* Buckingham, England: Open University Press.

Conway, M.A. (1992). A structural model of autobiographical memory. In M.A.Conway, D.C.Rubin, H.Spinnler, & W.A.Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp. 167–193). Dordrecht, The Netherlands: Kluwer.

Corcoran, R. & Upton, D. (1993). A role for the hippocampus in card sorting? *Cortex, 29,* 293–304.

Crockett, D., Bilsker, D., Hurwitz, T., & Kozak, J. (1986). Clinical utility of three measures of frontal lobe dysfunction in neuropsychiatric samples. *International Journal of Neuroscience, 30,* 241–248.

Dalla Barba, G. (1993). Confabulation: Knowledge and recollective experience. *Cognitive Neuropsychology, 10,* 1–20.

Damasio, A.R. (1985). The frontal lobes. In K.M.Heilman & E.Valenstein (Eds.), *Clinical neuropsychology* (2nd ed.). New York: Oxford University Press.

Delis, D.C., Squire, L.R., Bihrle, A., & Massman, P. (1992). Componential analysis of problem-solving ability: Performance of patients with frontal lobe damage and amnesic patients on a new sorting test. *Neuropsychohgia, 30,* 683–697.

Della Malva, C.L., Stuss, D.T., D'Alton, J., & Willmer, J. (1993). Capture errors and sequencing after frontal brain lesions. *Neuropsychologia, 31,* 363–372.

Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuropsychology, 3,* 271–290.

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

Duncan, J., Johnson, R., Swales, M., & Freer, C. (in press). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology.*

Einstein, G.O. & McDaniel, M.A. (1990). Normal ageing and prospective memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 717–726.

Ellis, A.W. (1982). Spelling and writing (and reading and speaking). In A.W.Ellis (Ed.), *Normality and pathology in cognitive functions.* London: Academic Press.

Ellis, A.W. (1987). Imitations of modularity, or, the modularity of mind: Doing cognitive neuropsychology without syndromes. In M.Coltheart, G.Sartori, & R.Job (Eds.), *The cognitive neuropsychology of language* (pp. 397–408). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Ellis, A.W. & Young, A.W. (1988). *Human cognitive neuropsychology.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Eslinger, P.J. & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal ablation: Patient EVR. *Neurology, 35,* 1731–1741.

Etkoff, N.K.L. (1984). Selective attention to facial identity and facial emotion. *Neuropsychologia, 22,* 281–295.

Fodor, J.A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Francis, D.J. (1988). An introduction to structural equation models. *Journal of Clinical and Experimental Neuropsychology, 10,* 623–639.

Fuster, J.M. (1980). *The prefrontal cortex.* New York: Raven Press.

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

Goldberg, E. (1986). Varieties of perseveration: A comparison of two taxonomies. *Journal of Clinical and Experimental Neuropsychology, 8,* 710–726.

Goldberg, E. (1995). Rise and fall of modular orthodoxy. *Journal of Clinical and Experimental Neuropsychology, 17,* 193–208.

Goldstein, L.H., Bernard, S., Fenwick, P., Burgess, P.W., & McNeil, J.E. (1993). Unilateral frontal lobectomy can produce strategy application disorder. *Journal of Neurology, Neurosurgery and Psychiatry, 56,* 271–276.

Incisa della Rocchetta, A., & Milner, B. (1993). Strategic search and retrieval inhibition: The role of the frontal lobes. *Neuropsychologia, 31,* 503–524.

Janowsky, J.S., Shimamura, A.P., & Squire, L.S. (1989). Source memory impairments in patients with frontal lobe lesions. *Neuropsychologia, 27,* 1043–1056.

Karnath, H.O., Wallesch, C.W., & Zimmerman, P. (1991). Mental planning and anticipatory processes with acute and chronic frontal lobe lesions: A comparison of maze performance in routine and non-routine situations. *Neuropsychologia, 29,* 271–290.

Kimberg, D.Y. & Farah, M.J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behaviour. *Journal of Experimental Psychology: General, 122,* 411–428.

Kolb, B. & Whishaw, I.Q. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: W.H.Freeman.

Kopelman, M.D. (1987). Two types of confabulation. *Journal of Neurology, Neurosurgery and Psychiatry, 50,* 1482–1487.

Kopelman, M.D. (1991). Frontal dysfunction and memory deficits in the alcoholic Korsakoff syndrome and Alzheimer-type dementia. *Brain, 114,* 117–137.

Kopelman, M.D. (1995). Delusional memory, confabulation and frontal lobe dysfunction: A case study in de Clerambault's syndrome. *Neurocase, 1,* 71–77.

Levin, H.S., Goldstein, F.C., Williams, D.H., & Eisenberg, H.M. (1991). The contribution of frontal lobe lesions to the neurobehavioral outcome of closed head injury. In H.S.Levin, H.M.Eisenberg, & A.L.Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 318–338). New York: Oxford University Press.

Lezak, M.D. (1983). *Neuropsychological assessment* (2nd ed.). Oxford: Oxford University Press.

Lezak, M.D. (1995). *Neuropsychological assessment* (3rd ed.). Oxford: Oxford University Press.

Lhermitte, F. (1983). "Utilization behaviour" and its relation to lesions of the frontal lobes. *Brain, 106,* 237–255.

Luria, A.R. (1973). *The working brain.* London: Penguin.

Marshall, J.C. (1984). Multiple perspectives on modularity. *Cognition, 17,* 209–242.

McCarthy, R.A. & Warrington, E.K. (1990). *Cognitive neuropsychology: A clinical introduction.* London: Academic Press.

McCloskey, M. (1993). Theory and evidence in cognitive neuropsychology: A "radical" response to Robertson, Knight, Rafal and Shimamura (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 718–734.

McNeil, J.E. & Warrington, E.K. (1993). Prosopagnosia: A face-specific disorder. *Quarterly Journal of Experimental Psychology, 46(A),* 1–10.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of pathology. *British Journal of Clinical Psychology, 23,* 359– 369.

Milner, B. (1963). Effects of different brain lesions on card-sorting. *Archives of Neurology, 9,* 90–100.

Milner, B. (1982). Some cognitive effects of frontal-lobe lesions in man. *Philosophical Transactions of the Royal Society of London B, 298,* 211–226.

Morton, J., Hammersley, R.H., & Bekerian, D.A. (1985). Headed records: A model for memory and its failures. *Cognition, 20,* 1–23.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313–324.

Nelson, H.E. & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the National Adult Reading Test. *Cortex, 14,* 234–244.

Norman, D.A. & Bobrow, D.G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology, 11,* 107–123.

Norman, D.A. & Shallice, T. (1980). Attention to action: Willed and automatic control of behaviour. Center for Human Information Processing (Technical Report No.99). (Reprinted in revised form in R.J.Davidson, G.E.Schwartz, & D.Shapiro [Eds.] [1986], *Consciousness and self regulation* [vol.4]. New York: Plenum Press.)

O'Carroll, R., Egan, V., & MacKenzie, D.M. (1994). Assessing cognitive estimation. *British Journal of Clinical Psychology, 33,* 193–197.

Owen, A.M., Sahakian, B.J., Hodges, J.R., Summers, B.A., & Robbins, T.W. (1995). Dopamine-dependent frontostriatal planning deficits in early Parkinson's disease. *Neuropsychology, 9,* 126–140.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual response in verbal categorical behaviour. *Neuropsychologia, 12,* 323–330.

Petrides, M. (1985). Deficits on conditional associative-learning tasks after frontal and temporal lobe lesions in man. *Neuropsychologia, 23,* 601–614.

Rabbitt, P. (1993). Does it all go together when it goes? *Quarterly Journal of Experimental Psychology, 46A,* 385–434.

Reason, J. (1993). Self-report questionnaires in cognitive psychology: Have they delivered the goods? In A.D.Baddeley & L.Weiskrantz (Eds.), *Attention: Selection, awareness and control. A tribute to Donald Broadbent* (pp. 406–423). Oxford: Clarendon Press.

Reiser, B.J., Black, J.B., & Kalamarides, P. (1986). Strategic memory search processes. In D.C.Rubin (Ed.), *Autobiographical memory* (pp. 100–121). London: Cambridge University Press.

Reitan, R.M. (1971). Trail Making Test results for normal and brain-damaged children. *Perceptual and Motor Skills, 33,* 575–581.

Roberts, R.J., Hager, L.D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393.

Sandson, J. & Albert, M.L. (1984). Varieties of perseveration. *Neuropsychologia, 22,* 715– 732.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology, 13,* 501–518.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B, 298,* 199–209.

Shallice, T. (1984). More functionally isolable subsystems but fewer "modules"? *Cognition, 17,* 243–252.

Shallice, T. (1988). *From neuropsychology to mental structure.* New York: Cambridge University Press.

Shallice, T. & Burgess, P.W. (1991a). Deficits in strategy application following frontal lobe lesions in man. *Brain, 114,* 727–741.

Shallice, T. & Burgess, P.W. (1991b). Higher order cognitive impairments and frontal lobe lesions in man. In H.S.Levin, H.M.Eisenberg, & A.L.Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125–138). New York: Oxford University Press.

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

Shallice, T. & Burgess, P.W. (1996). The frontal lobes and the temporal organisation of behaviour. *Philosophical Transactions of the Royal Society of London B, 351,* 1405–1412.

Shallice, T., Burgess, P.W., Schon, F., & Baxter, D. (1989). The origins of utilisation behaviour. *Brain, 112,* 1587–1598.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 14,* 294–303.

Shimamura, A.P. (1989). Disorders of memory: The cognitive science perspective. In F.Boller & J.Grafman (Eds.), *Handbook of neuropsychology* (vol. 3). Amsterdam: Elsevier.

Shimamura, A.P., Janowsky, J.S., & Squire, L.R. (1991). What is the role of frontal lobe damage in memory disorders? In H.S.Levin, H.M.Eisenberg, & A.L.Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 173–195). New York: Oxford University Press.

Shoqeirat, M.A., Mayes, A., MacDonald, J.C., Meudell, P., & Pickering, A. (1990). Performance on tests sensitive to frontal lobe lesions by patients with organic amnesia: Leng and Parkin revisited. *British Journal of Clinical Psychology, 29,* 401–408.

Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Agid, Y., & Dubois, B. (1995). Selective impairments in managerial knowledge following pre-frontal cortex damage. *Cortex, 31,* 301–316.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Stuss, D.T. (1991). Interference effects on memory functions in postleukotomy patients: An attentional perspective. In H.S.Levin, H.M.Eisenberg, & A.L.Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 157–172). New York: Oxford University Press.

Stuss, D.T. (1992). Biological and psychological development of executive functions. *Brain and Cognition, 20,* 8–23.

Stuss, D.T. & Benson, D.F. (1986) *The frontal lobes.* New York: Raven.

Stuss, D.T., Kaplan, E.F., Benson, D.F., Weir, W.S., Chiulli, S., & Sarazin F.F. (1982). Evidence for the involvement of orbito-frontal cortex in memory functions: An interference effect. *Journal of Comparative Physiological Psychiatry, 96,* 913–925.

Teuber, H.L. (1964). The riddle of frontal lobe function in man. In J.M.Warren & K.Albert (Eds.), *The frontal granular cortex and behaviour.* New York: McGraw-Hill.

Teuber, H.L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis, 32,* 615–656.

Thomsen, I.V. (1987). Late psychosocial outcome in severe blunt head trauma. *Brain Injury, 1,* 131–143.

Thomson, G.H. (1939). *The factorial analysis of human ability.* London: University of London Press.

Von Cramon, C.Y. & Von Cramon, G.M. (1994). Back to work with a chronic dysexecutive syndrome?: A case report. *Neuropsychological Rehabilitation, 4,* 399–417.

Warrington, E.K. (1984). *Recognition memory test.* Windsor: NFER-Nelson.

Warrington, E.K. & James, M. (1991). *The visual object and space perception battery.* Bury St. Edmunds, UK: Thames Valley Test Company.

Weiskrantz, L. (1992). Introduction: Dissociated issues. In A.D.Milner & M.D.Rugg (Eds.), *The neuropsychology of consciousness* (pp. 1–10). London: Academic Press.

Williams, D.M. & Hollan, J.D. (1981). The process of retrieval from very long-term memory. *Cognitive Science, 5,* 87–119.

Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome.* Bury St. Edmunds, UK: Thames Valley Test Company.

Zimmerman, I.L. & Woo-Sam, J.M. (1973). *Clinical interpretation of the Weschler Adult Intelligence Test.* New York: Grune & Stratton.

CHAPTER FIVE
Ageing and Executive Functions: A Neuroimaging Perspective

J.K.Foster *Department of Psychology, University of Manchester, Manchester, U.K.*

S.E.Black and Brian H.Buck *Department of Medicine, Research Program in Aging and Cognitive Neurology, Sunnybrook Health Science Centre, Toronto, Canada*

Michael J.Bronskill *Medical Imaging Research, Sunnybrook Health Science Centre, Toronto, Canada*

Executive functions include processes such as goal selection, planning, monitoring, sequencing, and other supervisory processes which permit the individual to impose organisation and structure upon his/her environment. There has been considerable recent neuropsychological interest in executive functions and the brain regions which are thought to subserve these capacities. Moreover, it appears that at least some executive functions may be compromised in older individuals, particularly in those people suffering from cognitive disorders such as dementia. Unfortunately, however, it has proven problematic to establish the precise brain regions which are critically implicated in the mediation of executive capacities, and the kinds of neuropsychological tests which are optimally sensitive to executive dysfunction.

Over the past decade there has been a tremendous burgeoning of interest into the potential utility of non-invasive brain imaging techniques in the delineation of brain—behaviour relationships. For those researchers interested in higher cognitive capacities, among the most significant brain regions which it is possible to probe using these newly available techniques are the frontal lobes. This large area of the cortical forebrain has been at the focus of special interest for those interested in executive and associated cognitive processes. However, the precise psychological capacities and behavioural functions subserved by prefrontal brain regions have been something of an enigma to researchers in the past. Furthermore, the eluci dation of the particular brain regions which subserve specific executive capacities has proven resistant to traditional neurobehavioural approaches. Recent findings suggest that executive deficits may emerge among the behavioural sequelae of damage to more widespread brain regions, rather than being truly specific to dysfunction of the frontal areas (see Stuss, Eskes, & Foster, 1994 for a review of the recent experimental literature). It therefore now

seems questionable whether theoretical concepts such as executive functions can be mapped neatly onto brain regions such as the frontal lobes.

The advent of new neuroimaging technologies presents an opportunity to make further progress in evaluating the relationship between widespread brain regions and diverse psychological capacities. More specifically, adopting a neuroimaging perspective permits us potentially to reappraise the question of the precise neural correlates of specific executive processes. Using such neuroimaging techniques it is possible *in vivo* to monitor the functional neural correlates of specific "executive" cognitive challenges using techniques such as PET (positron emission tomography). In addition, it is possible to evaluate the structural brain concomitants of executive impairment, utilising structural techniques such as magnetic resonance imaging (MRI). Now that the brain's "neural hardware" can be measured volumetrically, questions pertaining to the validity of executive tests can be addressed from a more rigorous and quantitative perspective.

The purpose of this empirical chapter is therefore to offer a complementary perspective to some of the other chapters in this volume by presenting MRI data pertaining to the question of the neural mediation of executive and associated functions. Furthermore, it is hoped that the data reported will cast some light on the problem of the appropriateness and methodology of several widely used tests of these capacities. It is sometimes assumed in the neuropsychological literature that established tests of executive functions are tapping specifically into the capacities of frontal brain regions. However, the ultimate test of this supposition is to determine to what extent it holds in the brains of living, functioning individuals. The key issue is the extent to which performance on executive tests correlates with known frontal pathology and/or dysfunction. By applying neuroimaging techniques in close temporal proximity to the testing of executive capacities it is now possible to address this question.

The central brain index on which the discussion presented here will be based is the relative (to overall cranial capacity) volume of particular brain regions and structures. Although it may not prove to be definitive, brain volume should at least be regarded as one indicant of level of brain function. This is particularly the case when one is considering dysfunction arising from significant brain atrophy. This investigation applied MRI in the study of the brain—behaviour status of older individuals. Specifically, detailed neuropsychological evaluation was coupled with high-resolution, volumetric brain imaging in normal older individuals and in patients suffering from Alzheimer's disease (AD). Brain—behaviour correlations were then computed in these people in order to investigate which specific brain regions could be related to the status of executive and associated functions. For comparison data pertaining to the neural mediation of non-executive capacities will also be discussed, in order to address the question of the regional specificity of the conclusions which may be drawn.

THE UTILITY OF VOLUMETRIC MRI

MRI is a powerful, non-invasive, high-resolution technique through which it is possible to image the brain *in vivo*. When this technique is harnessed to appropriate computer software, it can be used to quantify the volumes of designated brain compartments and regions of interest, and subsequently to address important questions concerning brain—behaviour relationships. Recent advances in methodology have made volumetric MR brain analysis an active research field (Jernigan et al., 1990; Press et al., 1989; Rusinek et al., 1991; Tanna et al., 1991). For example, the image analysis technique of bifeature segmentation (Cline et al., 1990), which was used to analyse the data reported here, makes use of the topographically coupled pixels derived from two matched brain images to segment brain regions into grey matter, white matter, CSF, and lesion compartments. This approach is more accurate than more conventional thresholding techniques because it is much less sensitive to spurious intensity variations in brain images and other methodological artefacts.

MR Volumetric Investigations of Regional Brain Atrophy in Normal and Pathological Ageing

Previous CT and MR studies have demonstrated significant ventricular and sulcal enlargement with accompanying cerebral shrinkage or atrophy in AD (Albert et al., 1984; Benson, 1986), but extant findings have differed with respect to relative loss in different tissue compartments (Jernigan et al, 1991; Kikinis et al, 1992; Rusinek et al., 1991). In terms of specific brain regions frontal lobe atrophy does not seem *necessarily* to occur in all patients with AD, although it is clear that there is dysfunction in the frontal regions in a significant subpopulation of AD patients, and deficits on tests of executive functions have been observed in some patients suffering from AD. Significant volume loss in the hippocampus and entorhinal cortex of AD patients compared with controls has also been widely demonstrated (Jack et al., 1992; Kesslak et al., 1991; Seab et al., 1988). Indeed, it is known from pathological studies that typical cases of AD follow a regionally distinct course of progression, initially involving parietal and temporal structures (including the hippocampal formation) and, later, the inferior and lateral frontal regions (Brun & Englund, 1981). Therefore, one would expect frontal involvement to be revealed in individuals in whom the disease has progressed to a more advanced stage.

Investigation of Brain—Behaviour Relationships

The central feature of this study concerned the investigation of brain—behaviour relationships, specifically with respect to executive and associated functions. Early attempts to correlate MR measurements of degenerative changes in the brains of AD patients with neuropsychological measures tended to use indices of

global cognitive status, such as the Mini-Mental State Exam (MMSE) (Bondareff et al., 1988; Kesslak et al., 1991). More recently, some attempts have been made to use neurobehavioural indices which are putatively region-specific (Heindel et al., 1992), but such studies have been somewhat hampered by restricted volumetric sampling of brain structures and problematic segmentation procedures. Furthermore, a number of these brain—behaviour studies have failed to consider the alternative explanation that any associations which are observed might be more parsimoniously attributed to global, non-specific brain atrophy or to spurious consequences of baseline levels of individual differential functioning.

In this study we attempted to correlate regional neuroanatomy with specific psychological deficits thought to arise from damage in particular brain areas, while simultaneously considering possible confounding factors such as an individual's general level of mental functioning, age and education. The specific hypotheses investigated were that the level of executive and attentional functions would correlate negatively with the degree of atrophy in frontal regions, whereas memory and learning capacities would correlate negatively with the level of atrophy in the hippocampal formation and associated temporal regions. In addition, performance on language and visuospatial tasks was predicted to correlate specifically with left- and right-parietotemporal atrophy respectively. These additional functions (memory, learning, language, visuospatial skills) were investigated in order to address the specificity of any significant neural correlations with executive functions. We further examined the neuropsychological test data using global and regional countercorrelations (i.e. the inverse correlations of those which had been predicted). We anticipated that the results of these supplementary analyses would not show significant brain—behaviour associations.

Participant Profiles

Participants with AD were compared with normal age-matched participants. A summary of participant demographics is shown in Table 5.1. The AD group consisted of 10 right-handed patients, all of whom met the NINCDS—ADRDA diagnostic criteria (McKhann et al., 1984) for probable AD. Ten right-handed age- and education-matched normal elderly individuals also took part in the investigation. Standard neurological and psychological exclusion criteria were applied in selecting participants and consent was required for participation in the study following appropriate ethical procedures.

Neuroimaging

Imaging was conducted with a 1.5 T MR system (SIGNA, Version 4.7; General Electric Medical Systems, Milwaukee, USA). For brain tissue volume

TABLE 5.1 Participant Demographics and Neuropsychological Test Performance

	Alzheimer Group (n = 10)	*Control Group (n = 10)*	*p**
Sex			
Male	7	6	n.s.
Female	3	4	n.s.
Age (years)	70.1 (4.9)	72.5 (5.7)	n.s.
Education (years)	13.0 (1.4)	13.6 (2.1)	n.s.
Mini-Mental State	19.4 (6.1)	27.7 (0.9)	< .001
DRS	105.6 (18.7)	140.3 (4.1)	< .001
NART	104.8 (10.3)	116.6 (6.7)	.008
Rey Copy	14.9 (13.8)	30.3 (2.7)	.003
FAS	20.9 (12.7)	47.1 (18.1)	.002
Semantic Fluency	7.4 (4.7)	21.1 (6.2)	< .001
Boston Naming	19.9 (6.7)	29.1 (0.7)	.001
Line Orientation	14.2 (10.8)	25.7 (4.6)	.007
WMS Visual Reproduction			
Immediate	11.5 (8.9)	34.0 (6.9)	< .001
Delayed	2.6 (3.7)	24.9 (11.9)	< .001
CVLT Word Recall (5 trials)	16.5 (7.7)	52.1 (10.0)	< .001
WMS Digit Span			
Forward	6.3 (1.6)	9.2 (1.7)	.001
Backward	4.0 (1.6)	6.8 (1.7)	.001
WCST Categories (64 trials)	0.6 (0.7)	2.3 (0.9)	.000

* Two-tailed student's *t* test for independent samples. n.s.=not significant.
CVLT=California Verbal Learning Test; WMS-R=Revised Wechsler Memory Scale; WCST=Wisconsin Card Sorting Test.

calculation, a two-spin echo sequence covering the whole brain was performed in the axial plane. Fifty-eight 3mm slices were obtained (half-Fourier sampling; 192 phase-encoding steps; TR/TE=3000/30, 80ms; field-of-view=20cm; imaging time=11.6m). The brain slices were contiguous and interleaved.

For hippocampal volume determination, a sagittal T1-weighted 3D volume technique was used (124, 1.3mm slices; TR/TE=35/5ms; flip angle =35°; field-of-view=22cm; imaging time=14.4m).

The bifeature segmentation procedure reported by Kikinis et al. (1992) was selected because of its high validity and reproducibility. The volume of grey matter, white matter, CSF and white matter lesions in the whole brain and in the left and right frontal and parietotemporal regions was calculated. The methodology was carefully tested and refined to achieve consistent results.

For regional analysis, six axial slices of brain tissue were used to estimate the relative volumes of frontal and parietotemporal regions. In addition to the midventricular segmentation slice, the three contiguous slices below and two contiguous slices above were used. The procedure devised to estimate the left

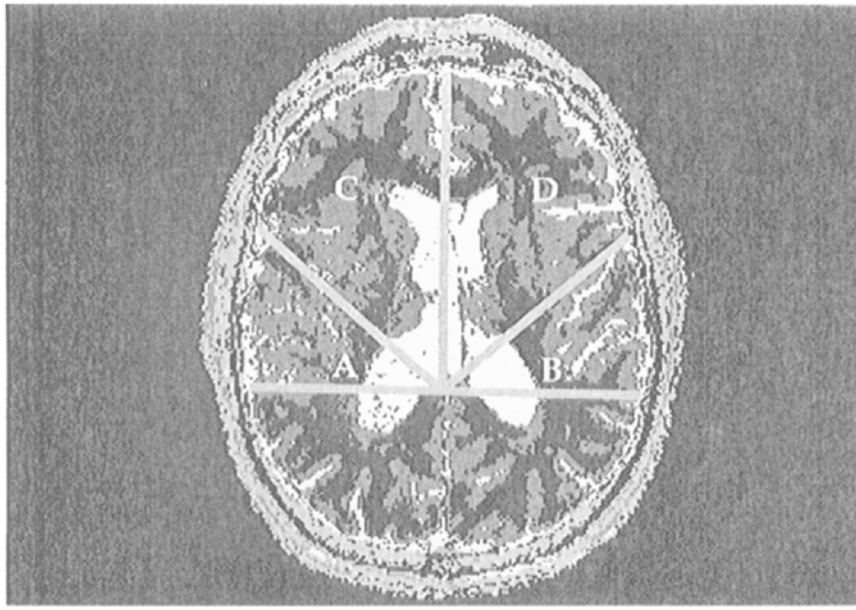


FIG. 5.1. Illustration of the procedure used to divide the axial slice into four regions: (A) left parieto-temporal; (B) right parieto-temporal; (C) left frontal; (D) right frontal.

and right frontal and parietotemporal regions in each slice is indicated in Fig. 5.1.

The reliability of the brain volumetric measurements was assessed with the intraclass correlation coefficient of reliability (Fleiss, 1986), which expresses the proportion of the total measurement variance attributable to the variability in the true brain volumetric measurements. As the amount of random error in measurements decreases, reliability increases and the intraclass correlation coefficient approaches unity.

Reliability was computed for measurements by three separate observers of all 20 brains imaged. With the exception of the white matter lesions, the intraclass correlation coefficients for the volumetric measurements of all regions and compartments were above 0.80. Moreover, percentage differences in total brain volumes for serial measurements on two normal individuals over a six-month interval were less than 1%.

Automated segmentation of the hippocampal formation proved challenging. A manual method was therefore adopted. The sagittal images were reformatted into coronal slices and the hippocampal areas in the left and right hemispheres were outlined manually by a neuroradiologist expert in MR hippocampal neuroanatomy (Fig. 5.2) (Bronen et al., 1991; Bronen & Cheung, 1991).

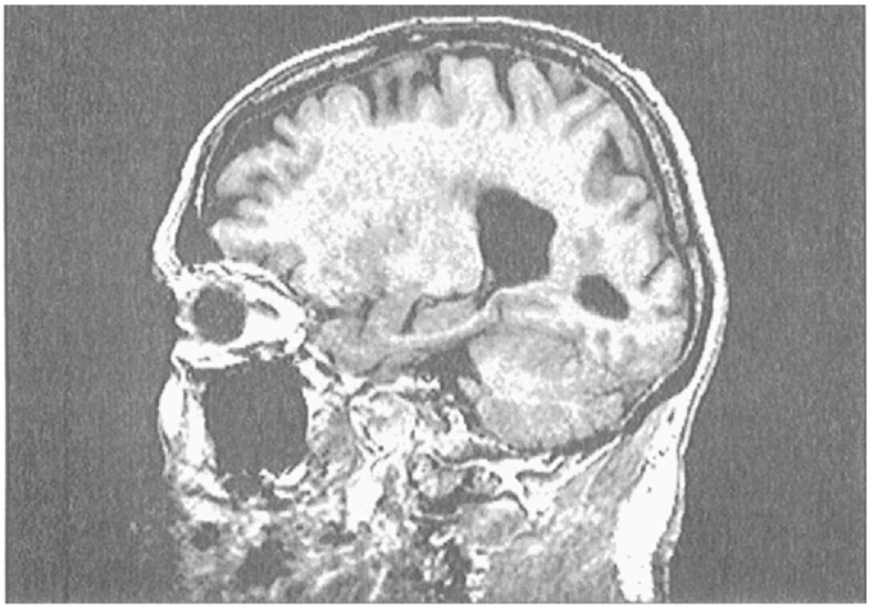


FIG. 5.2. Illustration of (1) the mid-hippocampal sagittal image used for reformatting the images perpendicular to the hippocampal long axis for the (2) coronal sections.

To account for differences in head size, all regional and hippocampal measures were corrected for overall cranial capacity (i.e. total volume of brain parenchyma plus CSF in that individual). Importantly, this procedure permitted us to derive relative measures of the volume of each tissue compartment or brain structure.

Multiple analyses of variance (MANOVAs) were then used to compare the whole brain, right and left frontal and parietotemporal tissue compartments (grey matter, white matter, and CSF), and right and left hippocampal volumes between the control and AD patients. The purpose of the MANOVAs was to evaluate group differences on the linear combination of global, regional, and hippocampal brain compartments. If the MANOVA was significant, further univariate analyses of variance (ANOVA) were then performed, in order to examine group differences in the individual compartments. By first performing MANOVA, the overall alpha error was controlled at the 0.05 level during multiple univariate comparisons.

The MANOVAs revealed significant effects due to group on the linear combination of global, frontal, parietotemporal, and hippocampal regions. Univariate ANOVAs, comparing individual compartment volumes between groups, showed significant reductions in the hippocampal volumes of the AD group. Additionally, the AD group showed a statistically significant increase in

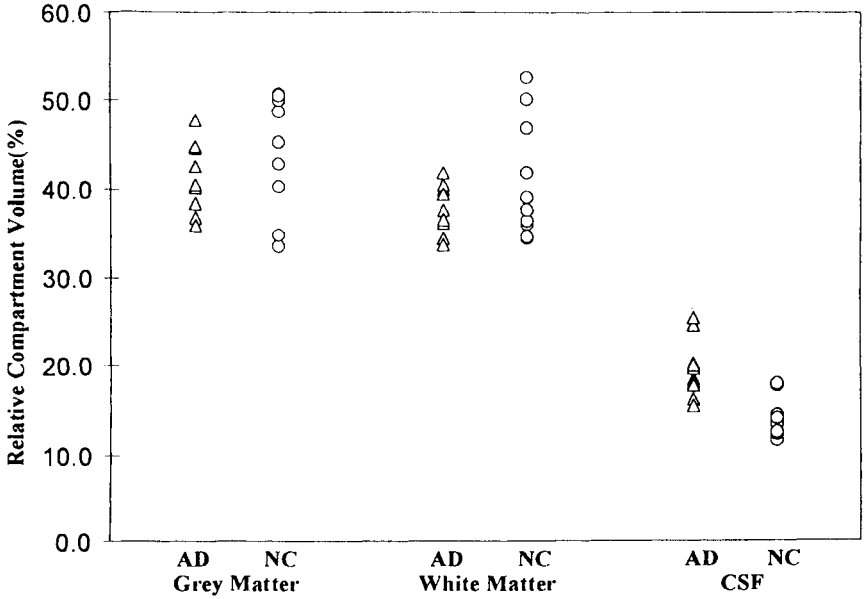


FIG. 5.3. Scatterplot illustrating the relative grey matter, white matter, and CSF volumes for Alzheimer patients and controls.

the volume of global CSF but no differences in the volumes of grey and white matter (Fig. 5.3). A similar pattern was seen in the analyses of regional CSF volumes.

Based on the outcome of the ANOVAs, a *post hoc* choice was made to conduct all subsequent statistical analyses of brain—behaviour relationships with the CSF compartment volume as the quantitative index of brain atrophy. This choice was further supported by the consistent reliability of the global and regional measurements of CSF volume (i.e. intraclass correlation coefficient>0.95).

Neuropsychological Assessment

Given recent findings in the neuropsychological literature suggesting that executive capacities are not exclusively mediated by the frontal lobes, it was considered important to evaluate executive processes in the context of a variety of functional measures. The neuropsychological tests employed were therefore selected for their putative ability to identify psychological dysfunction arising from damage to a wide range of cortical and subcortical structures, but focusing upon executive and associated capacities.

Tests were chosen which, based on previous researeh, were thought to reflect differentially focal damage to the frontal lobes, hippocampus, and

parietotemporal region (see Lezak, 1983; Spreen & Strauss, 1991). The tests were divided into six main categories: (1) executive functions; (2) short-term memory; (3) working memory; (4) learning and long-term memory; (5) language-related functions; and (6) visuospatial functions (Table 5.2).

Note from the table that, although two tests were allocated to the "executive functions" category (i.e. FAS and Wisconsin Card-Sort Task), performance on several of the other tests has been shown to be sensitive to frontal dysfunction, including immediate and delayed reproduction of the Wechsler Memory Scale (revised) figures, the Boston naming test, line orientation and Rey copy. As is becoming increasingly apparent in the neuropsychological literature the appropriate emphasis here is one of relative rather than of absolute distinctions, with performance at least partially mediated by common cognitive processes across test domains. This is especially the case when considering executive capacities.

Three measures of general functional capacity were also used. The Mattis Dementia Rating Scale (DRS) (Mattis, 1976) and the Mini-Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) enabled us to evaluate the level of severity of dementia, while the North American Adult Reading Test (NART) (Blair & Spreen, 1989) was also administered to obtain an index of premorbid cognitive ability.

Table 5.1 presents information on participant demographics, together with mean and standard deviation scores on the neurobehavioural tests. Age, sex distribution, and YOE were not significantly different between the two groups, but the mean NART score was significantly lower (p=.008) in the AD group than in the control group. As expected, performance on all other neuropsychological tests was worse for the AD group than for the control group.

In order to examine the relationship between the neuroimaging and neurobehavioural data partial correlations were then computed between seven brain volumes (global CSF volume, four regional CSF volumes [left and right frontal; left and right parietotemporal], left and right hippocampal volumes) and the neuropsychological tests (Table 5.3). To maximise statistical power, we included both groups (N=20) in our analysis, but we attempted to reduce the confounding effect of disease severity by first partialling out the DRS score. Since age and YOE could also affect neuropsychological performance and atrophy these variables were also partialled out from the correlations. In order to test the global significance of the 7×11 partial correlation matrix, a test of the omnibus null hypothesis (Cohen & Cohen, 1983) was subsequently conducted.

Using CSF and hippocampal volumes as independent measures, we entered the data from the neurobehavioural tests as dependent measures into multivariate models. The results of these analyses are presented in the correlation matrix (Table 5.3) which shows the partial correlation values between the various brain and neuropsychological variables. The partial correlation matrix as a whole was

TABLE 5.2 Putative Regional Location of Cerebral Mediation of Performance on Neuropsychological Tests

	Neuropsychological Function	*LF*	*RF*	*LPT*	*RPT*	*LH*	*RH*
Executive Functions							
FAS Fluency	phonemic word retrieval	+ +	+	+			
WCST	abstraction, mental flexibility	+ +	+ +	+	+		
Short-term Memory (STM)							
WMS-R Forward Digit Span	verbal STM			+			
WMS-R Immediate Reproduction	non-verbal STM		+		+		
Working Memory							
WMS-R Backward Digit Span	verbal working memory	+	+			+	
Learning and Long-term Memory							
CVLT Word Recall	verbal recall, recognition	+		+		+ +	
WMS-R Delayed Reproduction	non-verbal recall		+		+		+ +
Language Functions							
Boston Naming Test	visual naming	+		+ +			
Semantic Fluency	category word retrieval			+ +			
Visuospatial Functions							
Line Orientation	spatial matching		+		+ +		
Rey Copy	visuoconstructive skills		+		+ +		

+=the number of plus signs provides an index of the hypothesised strength of association between psychological and regional cerebral dysfunction.
LF=Left Frontal; RF=Right Frontal; LPT=Left Parieto-temporal; RPT=Right Parieto-temporal; LH=Left Hippocampal; RH=Right Hippocampal.

significant (2(70)=174; p<.01), indicating that it was unlikely that the observed associations between atrophy and neuropsychological performance were a result of chance. Several correlations reached significance at the .05 level, as shown in Table 5.3.

Group Differences in Cortical and Hippocampal Atrophy in AD

Of the brain measures, differences in CSF showed up strongly on the MANOVAs. There was greater atrophy in all regions sampled in the AD group compared with controls. Furthermore, there was almost no overlap in the global CSF measures obtained, with the AD group showing a mean volume of 19.6cm^3

TABLE 5.3 Partial Correlation Matrix (Age, Education, DRS Held Constant) Illustrating the Relationship Between Neuropsychological Test Performance and Global/Regional CSF/Hippocampal Volumes

	Global CSF	*LF CSF*	*RF CSF*	*LPT CSF*	*RPT CSF*	*LH*	*RH*
Executive Functions							
FAS Fluency	–0.44*	0.03	–0.12	–0.21	–0.11	0.21	0.20
WCST Categories	–0.06	0.01	–0.42*	0.03	–0.07	–0.03	0.11
Short-term Memory							
WMS-R Digits	–0.46*	–0.28	–0.18	–0.51*	–0.40	0.52*	0.37
WMS Figures: Immediate Reproduction	–0.67**	–0.35	–0.44*	–0.37	–0.36	0.37	0.29
Working Memory							
WMS-R Digits Backwards	–0.45*	–0.30	–0.47*	–0.33	–0.13	0.35	0.36
Learning and Long-term Memory							
CVLT Word Recall	–0.60**	–0.45*	–0.45*	–0.43*	–0.56**	0.49*	0.39
WMS Figures: Delayed Reproduction	–0.26	–0.38	–0.26	–0.29	–0.20	0.62**	0.51*
Language Functions							
Boston Naming	0.03	0.19	–0.06	–0.28	–0.29	–0.07	–0.04
Semantic Fluency	–0.47*	–0.29	–0.26	–0.54*	–0.41	0.39	0.19
Visuospatial Functions							
Line Orientation	–0.13	–0.16	–0.12	–0.08	0.08	0.09	0.05
Rey Copy	–0.56**	–0.39	–0.34	–0.28	–0.32	0.32	0.13

* $p<.05$ (one-tailed test); ** $p<.01$ (one-tailed test).

(SD=3.2cm^3) and the elderly control subjects showing a mean volume of 14.1cm^3 (SD=2.2cm^3) (Fig. 5.3). These findings are similar to other reports in the MR literature (Jernigan et al., 1990; Press et al., 1989; Rusinek et al., 1991).

Overall volumes of grey and white matter compartments did not differ significantly between the two groups but, importantly, there was a significant decrease in the AD group in the volume of grey and white matter in the left frontal region. Other regions showed a similar trend, which did not achieve statistical significance. Our findings therefore seem to indicate that the pathological reduction in the volume of brain parenchyma in AD is not specific to either the grey or white matter tissue compartments. However, we did find regional accentuation in the severity of brain atrophy. It is noteworthy that parietotemporal and frontal CSF measurements showed a greater relative increase in AD patients relative to controls than did the global measurement. The latter finding was especially interesting, given that frontal atrophy is not regarded as a typical pathological hallmark of AD. Furthermore, it suggested that there

was sufficient variability in the volume of more anterior brain regions to evaluate our hypotheses regarding the neural correlates of executive functions.

The mean hippocampal volume of the AD patients was significantly reduced by approximately 20%. However, there was considerable overlap between participants in the two groups. The degree of overlap is possibly related to the observation that hippocampal volume appears to decrease in normal, as well as pathological, ageing.

Group Differences in Neuropsychological Performance

As expected the performance of patients in the AD group was impaired on the vast majority of the neuropsychological tests utilised (Table 5.1). Although age and YOE did not differ significantly between the AD and control groups, there was, nevertheless, a significant difference between the NART scores of the AD patients and controls. This finding indicates a discrepancy between two frequently used premorbid indicants of intellectual ability (NART and YOE), and casts some doubt on the utility of the NART score as a valid index of premorbid functioning in AD patients.

Association between CSF Volumes and Neuropsychological Performance

This study aimed to examine brain—behaviour correlates in AD using noninvasive brain imaging techniques, coupled with neuropsychological assessment methods. A particular motivation for the study concerned the investigation of the neural mediation of executive functions.

The data indicated that a number of our *a priori* hypotheses concerning brain-behaviour associations were not supported (see Table 5.3). Moreover, this somewhat negative outcome seemed to apply to both executive and non-executive functions, indicating a possible problem with the regional specificity and/or sensitivity of extant neuropsychological tests in general rather than executive tests *per se*. However, some of the predicted correlations were significant, despite the small sample size. The most apparent correlation was that between global CSF and almost all of the neuropsychological tests. Such global correlations are perhaps not unexpected given the diffuse nature of the atrophy that can occur in AD. Moreover, any region-specific differences were, of course, incorporated into the global measures.

There was a mixed pattern of correlations for the executive tests, with few apparently showing truly selective associations with frontal regional measures. However, the category score of the Wisconsin Card-Sort Test (WCST) was correlated with right frontal CSF. Furthermore, this association appeared to be regionally specific, as it did not correlate with global or other regional measures, including the predicted correlation with left frontal CSF. Of the other tests which

were expected to reveal significant association with frontal regions, performance on digit span backwards showed a significant correlation with the right frontal area, although it also correlated significantly with the volume of global CSF. However, it did not show the predicted association with left frontal CSF. Delayed reproduction of the Wechsler figures did not show the predicted right frontal association, but this correlation was significant for the immediate reproduction of the figures, albeit in the context of an accompanying significant global correlation. Furthermore, performance on the California Verbal Learning Test (CVLT) word recall correlated with the volume of both left and right frontal regions, although this association was far from specific.

Among the other neurobehavioural measures, performance on some of the memory tests was associated with hippocampal volume. However, these associations were, again, often not regionally specific. A similar pattern emerged with the supplementary language and visuospatial measures, with respect to the brain regions thought to mediate performance on these tasks (see Tables 5.2 and 5.3).

From a neuropsychological perspective, the lack of unequivocal support for many of the hypothesized brain—behaviour correlations (including those involving executive functions) is perhaps not surprising given that the tasks employed are multifactorial and likely involve the orchestration of more than one brain region. However, these observations run counter to some traditional neuropsychological suppositions. Our findings suggest that 1:1 brain—behaviour correlations may not be a realistic outcome of neurobehavioural investigations at least in the populations investigated here. Furthermore, our data indicate that tests of executive functions are not unique in this regard, and it may be misleadingly simplistic to argue that the performance of any neuropsychological task is specifically sensitive to the integrity of a particular brain region. On the other hand, perhaps the regional brain volumes obtained in the current investigation were (with the possible exception of the hippocampal measures) anatomically too crude to establish robust associations between specific psychological capacities and the brain regions subserving them.

Other important considerations include heterogeneity within participant groups, which may well have contributed towards variability in the brain data. The statistical sample may have been too small to test reliably the large number of correlations. However, the fact that the omnibus test of the whole correlation matrix was significant suggests that those significant correlations which were observed between atrophy and psychological performance were not merely due to chance. Finally, it may also be the case that traditional neuropsychological tests do not necessarily provide the best behavioural indicators of atrophy in specific brain regions in AD; a more experimental, cognitive neuropsychological approach may be required.

However, the findings do have important implications for the question of which neuropsychological tests are the most revealing indicators of atrophy in specific brain regions in normal and pathological ageing. Specifically, in the case of executive functions, our data indicate that, of the measures used, the categories measure of the Wisconsin Card-Sort Test should be the preferred index of executive capacities, with digit span backwards and immediate reproduction of the Wechsler figures also proving useful in predicting the degree of frontal brain atrophy.

FUTURE DIRECTIONS

The evaluation of brain—behaviour relations using *in vivo* neuroimaging methods offers new possibilities to correlate functional and structural brain changes with psychological deficits. However, the present study indicates that certain neuropsychological tests (including measures of executive capacities) may be much more sensitive and specific than others to damage in particular brain regions. Some of the predicted brain—behaviour correlations did emerge, but few of the neuropsychological tests showed truly selective associations with relative atrophy in the specific brain regions thought to mediate performance on these tasks. Rather, the majority of the neuropsychological tests correlated with atrophy in other non-predicted brain regions and/or with global reduction in parenchymal volume. Specifically, within the domain of executive and associated functions, only performance on the category score of the Wisconsin Card-Sort Test, the digit span backwards task and immediate reproduction of the Wechsler figures showed a significant positive correlation with the volume of frontal brain regions. Among these tests only the WCST categories measure showed a truly regionally specific correlation, i.e. with no accompanying global correlations present.

With respect to more general brain—behaviour issues, the findings of this study indicate some of the potential failings of conventional neurobehavioural measures. There is sometimes a lack of rigorous evaluation in the neuropsychology literature concerning the validity and generalisability of psychometric measures, many of which have been developed utilising data obtained from normal individuals and/or patients who have experienced "focal" brain damage. Individuals with both distributed (e.g. dementia) and more restricted (e.g, stroke) primary lesions should be evaluated before attributions about the selectivity and specificity of particular neuropsychological measures can be confidently made. The range of problems associated with the interpretation of lesion-based data should also be considered. An extensive theoretical debate about these important questions has taken place over the past three decades (see, for example, Gregory, 1961; Webster, 1973; Weiskrantz, 1968, 1974 on interpretative issues, and Finger & Stein, 1982; Schoenfeld & Hamilton, 1977 for more practical considerations). Taken together, these issues

underline the importance of adopting a converging operations perspective, in which complementary investigatory techniques (neuroimaging, neurobehavioural, developmental, cognitive) are used to address central research questions.

A further possible implication of our findings is that structural brain measures alone may be inappropriate for determining brain—behaviour relationships. Rather, in addition to new, more experimentally motivated neuropsychological tests, complementary (specifically functional) imaging procedures are also required. The availability of such data sets would permit the coupling of morphometric brain measures with metabolic brain indices. These structural/functional brain measures can then be related (either together or separately) with the findings obtained from neuropsychological testing procedures. Furthermore, when using structural MR measures to investigate possible brain—behaviour relationships, more refined neuroanatomical mapping should be attempted through the application of detailed parcellation techniques (see, for example, Rademacher et al., 1992). Such an approach would help to address the question of whether morphometric MR techniques are conceptually suitable for addressing important neurobehavioural questions, or whether those structural indices which have been used to date have simply been anatomically too crude to inform our conceptualisation of brain—behaviour relationships.

A final central issue raised by the present investigation concerns the distribution and organisation of the brain regions involved in subserving executive capacities. It would seem questionable, from a reading of the contemporary neuropsychology literature, whether any one particular test can be accepted as uniquely tapping executive functioning, and this point would seem to be supported by our findings. The concept of anatomically distributed functional systems, of which the frontal regions form an integral part, is perhaps important in this context (Alexander, De Long, & Strick, 1986; Barbas & Pandya, 1991; Rosvold & Szwarcbart, 1964). There is accumulating evidence that performance on tests of executive function is not solely mediated via the frontal lobes, and that deficits on these tasks may also emerge following damage to non-frontal brain regions (Goldberg & Bilder, 1987; Kinsbourne, 1977; Stuss & Gow, 1992). These observations indicate the importance of performing an assessment of executive functions within the context of a broadly based neuropsychological evaluation of both frontal and other regional brain functions, as in the study reported here. Such a comprehensive neuropsychological assessment is critical before specific attributions about the location of any underlying brain dysfunction can be securely made.

ACKNOWLEDGEMENTS

The work reported in this chapter was supported by the Ontario Mental Health Foundation, Canada and a NATO Collaborative Science Grant. J.K.Foster was supported by a NATO Postdoctoral Fellowship (administered by the UK Science and Engineering Research Council) and a Rotman Research Institute Postdoctoral Fellowship. We are grateful to the following individuals for their scientific, logistic, and moral support: Peter Stanchev, John P.Szalai, Gordon Winocur, and Morris Moscovitch. We would also like to thank Gordon Cheung for performing the hippocampal measures. We are particularly grateful to Donald Stuss for his advice and to Farrell Leibovitch and Matthew Norton, who assisted in the image analysis. Jessica Haberman and Jayson Parker are thanked for typing a previous version of this manuscript and for preparing figures, respectively. The support of GE Medical Systems is also acknowledged.

Part of the work reported in this chapter was presented at the 118th Annual Meeting of the American Neurological Association, Boston, October 1993 and the 2nd Annual Meeting of the Society of Magnetic Resonance, San Francisco, August 1994.

REFERENCES

Albert, M., Naeser, M.A., Levine, H.L. et al. (1984). Ventricular size in patients with presenile dementia of the Alzheimer type. *Archives of Neurology, 41,* 1258–1263.

Alexander, G.E., De Long, M.R., & Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357– 381.

Barbas, H. & Pandya, D.N. (1991). Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. In H.S.Levin, H.M.Eisenberg, & A.L. Benton (Eds.), *Frontal lobe function and dysfunction.* New York: Oxford University Press.

Benson, D.F. (1986). Neuroimaging and dementia. *Neurologic Clinics, 4,* 341–353.

Blair, J.R. & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist, 3,* 129–136.

Bondareff, W., Raval, J., Colleti, P.M. et al. (1988). Quantitative magnetic resonance imaging and the severity of dementia in Alzheimer's disease. *American Journal of Psychiatry, 145,* 853–856.

Bronen, R.A. & Cheung, G. (1991). Relationship of hippocampus and amygdala to coronal MRI landmarks. *Magnetic Resonance Imaging, 9,* 449–457.

Bronen, R.A., Cheung, G., Charles, J.T. et al. (1991). Imaging findings in hippocampal sclerosis: Correlation with pathology. *American Journal of Neuroradiology, 12,* 933–940.

Brun, A. & Englund, E. (1981). Regional pattern of degeneration in Alzheimer's disease: Neuronal loss and histopathological grading. *Histopathology, 5,* 549–564.

Cline, H.E., Lorensen, W.E., Kikinis, R. et al. (1990). Three-dimensional segmentation of MR images of the head using probability and connectivity. *Journal of Computer Assisted Tomography, 14,* 1037–1045.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioural sciences.* London: Lawrence Erlbaum Associates Ltd.

Finger, S. & Stein, D.G. (1982). *Brain damage and recovery.* New York: Academic Press.

Fleiss, J.L. (1986). *The design and analysis of clinical experiments.* New York: John Wiley & Sons.

Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Goldberg, E.&Bilder, R.M. (1987). The frontal lobes and hierarchical organization of cognitive control. In E.Perecman (Ed.), *The frontal lobes revisited.* New York: IRBN Press.

Gregory, R.L. (1961). The brain as an engineering problem. In W.Thorpe & O.L.Zangwill, (Eds.), *Current problems in animal behaviour.* London: Cambridge University Press.

Hachinski, V.C., Sliff, L.D., Zalkha, E. et al. (1975). Cerebral blood flow in dementia. *Archives of Neurology, 32,* 632–637.

Heindel, W.C., Jernigan, T.L., Salmon, D.P. et al. (1992). Brain structural correlates of cognitive performance in Alzheimer's disease. *Journal of Clinical Experimental Neuropsychology, 14,* 67.

Jack, C.R., Petersen, R.C., O'Brien, P.C. et al. (1992). MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology, 42,* 183–188.

Jernigan, T.L., Archibald, S.L., Berhow, M.T. et al. (1991). Cerebral structures on MRI: Part I. Localization of age-related changes. *Biological Psychiatry, 29,* 55–67.

Jernigan, T.L., Press, G.A. & Hesselink, J.R. (1990). Methods for measuring brain morphologic features on magnetic resonance images: Validation and normal aging. *Archives of Neurology, 47,* 27–32.

Kesslak, J., Nalcioglu, O. & Cotman, C. (1991). Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. *Neurology, 41,* 51–54.

Kikinis, R., Shenton, M.E., Gerig, G. et al. (1992). Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *Magnetic Resonance Imaging, 2,* 619–629.

Kinsbourne, M. (1977). Cognitive decline with advancing age: An interpretation. In W.L. Smith, & M.Kinsbourae, (Eds.) *Aging and dementia.* New York: Spectrum Publications.

Lezak, M. (1983). *Neuropsychological Assessment.* New York: Oxford University Press.

Mattis, S. (1976). Mental status examination for organic mental syndrome in the elder patient. In L.Bellack & T.B.Karasu (Eds.), *Geriatric psychiatry.* New York: Grune & Stratton.

McKhann, G., Drachman, D., Folstein, M.F. et al. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services task force on Alzheimer's disease. *Neurology, 34, 939–* 945.

Press, G.A., Amaral, D.G. & Squire, L.R. (1989). Hippocampal abnormalities in amnesic patients revealed by high-resolution magnetic resonance imaging. *Nature, 341,* 54–57.

Rademacher, J., Galaburda, A.M., Kennedy, D.N. et al. (1992). Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience, 4,* 352–374.

Rosen, W.G., Terry, R.D., Fuld, P.A. et al. (1980). Pathological verification of the ischemic score in the verification of dementias. *Annals of Neurology, 7,* 486–488.

Rosvold, H.E. & Szwarcbart, M.K. (1964). Neural structures involved in delayed-response performance. In J.M.Warren & K.Akert (Eds.), *The frontal granular cortex and behavior.* New York: McGraw-Hill.

Rusinek, H., de Leon, M.J., George, A.E. et al. (1991). Alzheimer disease: Measuring loss of cerebral gray matter and MR imaging. *Radiology, 178,* 109–114.

Schoenfeld, T.A. & Hamilton, L.W. (1977). Secondary brain changes following lesions: A new paradigm for lesion experimentation. *Physiology and Behavior, 18,* 951–967.

Seab, J.P., Jagust, W.J., Wong, S.T.S. et al. (1988). Quantitative NMR measurements of hippocampal atrophy in Alzheimer's disease. *Magnetic Resonance Medicine, 8,* 200–208.

Spreen, O. & Strauss, E. (1991). *A compendium of neuropsychological tests.* New York: Oxford University Press.

Stuss, D.T., Eskes, G.A., & Foster, J.K. (1994). Experimental neuropsychological studies of frontal lobe functions. In F.Boller & J.Grafman (Eds.), *Handbook of neuropsychology* (vol. 9). Amsterdam: Elsevier.

Stuss, D.T. & Gow, C.A. (1992). "Frontal dysfunction" after traumatic brain injury. *Neuropsychiatry, Neuropsychology and Behavioral Neurology, 5,* 274–282.

Tanna, N.K., Kohn, M.I., Horwich, D.N. et al. (1991). Analysis of brain and cerebrospinal fluid volumes with MR imaging: Part II. Aging and Alzheimer dementia. *Radiology, 18,* 123–130.

Webster, W.G. (1973). Assumptions, conceptualizations, and the search for the functions of the brain. *Physiological Psychology, 1,* 346–350.

Weiskrantz, L. (1968). Treatments, inferences, and brain function. In L.Weiskrantz (Ed.), *Analysis of behavioural change.* London: Harper & Row.

Weiskrantz, L. (1974). Brain research and parallel processing. *Physiological Psychology, 2,* 53– 54.

CHAPTER SIX
Sustained Attention and the Frontal Lobes

Tom Manly *MRC-APU Rehabilitation Research Group,*
Addenbrookes Hospital, Cambridge, U.K.
Ian H.Robertson *MRC Applied Psychology Unit, Cambridge, U.K.*

Things take time. An effective cognitive system requires the ability to perpetuate process sufficiently to match the demands of goal and environment. Some functions, such as the retention of a fact once learned, are subjectively effortless and have a time course of years. For others, such as the serial addition of auditory digits presented every two seconds, continuation over very brief periods feels highly effortful and leads to increasingly error-prone performance (Gronwall & Sampson, 1974).

This chapter concerns the perpetuation of attention. Attention is a process of cognitive control which selects from a potentially huge range of possible inputs and a potentially huge range of possible actions. Such selectivity, it has been argued, is necessary for coherent and effective operation of a limited capacity system (Broadbent, 1971), whether this limitation occurs at the level of how much information can be processed simultaneously or how many actions can be performed simultaneously. Just as capacity is limited in terms of simultaneous performance, so too is the capacity to maintain an initially achievable performance over time. Of 652 respondents in a postal survey of Cambridge residents conducted by the authors, 43% felt that during the previous week their attention had wandered off during long television programmes, 48% reported failing to catch how their local weather was going to be despite having just watched the forecast, and 63% reported needing to reread paragraphs because of inattention at the first pass.

The focus of this chapter is the contribution of the human frontal lobe in maintaining attention over time. We will discuss evidence from two prin cipal sources which allow inferences about structure-function links: the examination of patterns of cognitive impairment shown by patients with brain damage and the consideration of the patterns of metabolic activity during the performance of cognitive tasks in normal participants. That the former approach may have value is suggested by the apparent exacerbation of problems in maintaining attention

seen in certain neurological patients. Occasionally this can appear to be very specific. B.E., for example, worked as a very successful computer programmer before sustaining a closed-head injury in a road traffic accident. The accident left him with a post-traumatic amnesia of between one and two weeks. A neuropsychological assessment some 15 months later revealed little of obvious concern. On a measure of "fluid intelligence", the Cattell culture fair, his score was substantially above average. On a demanding measure sensitive to cognitive deficits following closed-head injury, the Paced Auditory Serial Addition Task (Gronwall & Sampson, 1974), he scored within the normal range. No dramatic shifts in personality or motivation were reported by people who knew him. Despite this profile both B.E. and his colleagues detected a decline in his work performance. He would start jobs enthusiastically but leave them unfinished. Sometimes midway though a task he would notice that he had "drifted off." As he worked on a program, irrelevant events and noises in the workplace now sometimes seemed magnetic in distracting his attention from the task.

NORMAL "VIGILANCE" PERFORMANCE

Subjective reports of difficulties in maintaining performance in complex situations such as those of B.E. and of the survey respondents are not readily amenable to analysis. A cognitive neuropsychological approach to this area therefore requires experimental designs in which the effects of individual factors may be more easily considered. To this end work with neurological patients and in imaging studies has borrowed heavily from the experimental investigation of normal "vigilance" performance. These paradigms, developed in the 1950s, created situations which were analogous to (indeed were initially devised to test) a radar operator watching a screen over long periods for the occurrence of rare events.

In an early demonstration of limited normal capacity over time, participants were asked to monitor a clock hand moving at one increment per second for a two-hour period. The target to which they had to respond was a single movement taking in two increments. This occurred approximately 12 times in each half hour. While this rare event was relatively easy to distinguish, a decline in the number of targets detected was observed after the first half hour. Increases in the rate of such errors were noted in each subsequent half hour block, this pattern being characterised as the *vigilance decrement* (Mackworth, 1950, 1961).

It could be argued that overreliance on methodologies developed to consider performance in monotonous, repetitive tasks such as radar watching or industrial quality control could unnecessarily limit the definition of attentive behaviour over time and the scope of neuropsychological research. Certainly there are few reports of the extent to which performance on such paradigms predicts complaints of difficulties in real life. However, the intensive study of such tasks with normal participants has led to considerable clarification of the factors which

can influence performance and therefore provides a good basis for the examination of these factors in neurological patients. The use of such paradigms is also supported by the observation of generally high correlations between quite variant tasks and between similar tasks using different modalities (Davies & Parasuraman, 1982; See, 1995), suggesting that common, core processes are involved which presumably play a part in more complex prolonged situations requiring attentional control.

Following the early demonstrations of the vigilance decrement it was proposed that increasing rates of detection failure may be related both to diminishing sensitivity to characteristics of the stimulus and to changes in the participant's criteria for making a response. In an attempt to tease these factors apart components in performance were considered in a signal detection analysis (Broadbent & Gregory, 1963; Egan, Greenberg et al, 1961). A general assumption of this approach was that reduced sensitivity to target characteristics would lead to an increased number of false—positive responses, that is responding to a non-target as if it were a target, while a tendency to become more conservative in responding generally would lead to a decline in both correct and incorrect responses (Swets, 1977).

A previous general finding in the literature on normal individuals using this analysis has been that decrements in performance are almost exclusively attributable to an increased criterion for responses, that is of the development of a more conservative response strategy (Broadbent, 1971; Swets, 1977). Factors which were found to influence criteria included the prior knowledge of, or on-line assessment of, the probability of a target occurrence (Baddeley & Colquhoun, 1969), presumably enabling participants to use variations in expectation as a cue to attend to task particularly at appropriate times.

More recent manipulations of factors in vigilance tasks have indicated that an apparent reduction in participants' ability to discriminate targets from non-targets is a prevalent phenomenon. Reviewing the effect of such manipulations, Parasuraman has proposed a taxonomy of vigilance (Parasuraman & Davies, 1977; Parasuraman, Warm et al., 1987). This taxonomy classifies tasks on the factors of event rate (fast or slow), target dis crimination type (successive or simultaneous), sensory modality (auditory or visual), and source complexity (single or multisource). Subsequently, the additional classification of tasks as "sensory" or "cognitive" (perceptual judgments or judgments about familiar alphanumeric characters) has been proposed (See, 1995).

Investigations have suggested complex interactions between these factors and the postulated reasons for decrements. For example it was noted that at fast presentation rates (in excess of one stimulus per 2.5 seconds) in tasks where comparison with previous presentations in the sequence was required to establish whether or not the current stimulus was a target, decrements were related to a decline in sensitivity (false—positive responses increased). In tasks with slow

presentation rates in which all the information required to judge stimulus status was present in a single instance, decrements were largely attributed to changes in response criteria (all responses decreased; Parasuraman, 1979).

Applying Shiffrin and Schneider's distinction between controlled, attentionally demanding processing and automatic, attentionally undemanding processing (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), Fisk and Schneider found performance decrements to be greater under conditions where the development of automaticity was difficult owing to changes in the target stimulus throughout the task (Fisk & Schneider, 1981). A signal detection analysis suggested that decrements in the controlled condition were more attributable to changes in sensitivity while shifts in criterion gave a better account of decrements under automatic processing conditions. Such findings have led to the suggestion that declining performance at fast presentation rates in conjunction with a demanding task may be related to attentional capacity limitation, while decline in slow-rate, cognitively undemanding tasks may be due to changes in criterion due to target expectancy (Parasuraman, Warm et al., 1987; Vickers & Leary, 1983).

Mathews and Holley (1993) report that performance on *brief,* attentionally demanding tasks predicts performance decrements over longer periods only on tasks with fast presentations and which require difficult discriminations. In more general terms, it may be possible to draw a distinction between failures in "sustained attention" due to attentional capacity limitation under repeated, rapid and high demand—*despite those resources being primarily oriented towards the task*—and failures due to poor control of attention, that is *failing to maintain the attention on the task at all*.

FOCAL LESION STUDIES

Damage to the frontal lobe can cause profound impairment in everyday life despite many cognitive functions appearing intact when tested in isolation. Consideration of the effects of frontal damage has led to a broad agreement that these structures subserve a particular executive role in orchestrating activity to produce coherent and effective behaviour (Foster, Eskes et al., 1994; Luria, 1966; Shallice, 1988; Shallice & Burgess, 1991; Stuss & Benson, 1984). One predictable consequence of poor executive control would be a tendency for attention to be hijacked by salient but task-irrelevant events, and this has indeed been a common clinical observation (e.g. Hecaen & Albert, 1978; Luria, 1966). While the neuropsychological picture of attentional deficits following frontal damage is far from consistent, abnormal performance of selective attention tasks (Bench, Frith et al., 1993; Corbetta, Miezin et al, 1991; Janer & Pardo, 1991; Pardo, Pardo et al, 1990) and in the spatial allocation of visual attention (Mesulam, 1981) have been reported. There is good reason to suppose, therefore,

that adequate prefrontal function may be a necessary, if not sufficient, condition for sustaining attention over time.

Perhaps the clearest demonstration of prefrontal involvement in maintaining attention to undemanding tasks was provided by Wilkins et al. (Wilkins, Shallice et al., 1987). The measurement of abilities in perpetuating performance requires some "leading edge" task against which fluctuations or decrements with time can be judged. While it would be reasonable to consider variation in performance on almost any task as an interesting parameter in itself, the core process in examining sustained attention has, generally, been one of serial contingent response. Wilkins et al., however, adopted an alternative paradigm, that of counting events. In their task participants were required either to count the number of identical tones they heard or to count the number of times their finger received a tactile stimulus during strings of up to 11 such stimuli.[1]

At the slowest rates of presentation patients with lesions to right anterior structures were disproportionately impaired in both auditory and right-hand tactile conditions relative to left-frontal and right- and left-temporal lobe patients. The absence of a group effect at faster rates of stimulus presentation made it unlikely that differences in numerosity per se could account for this result. That impairment was apparent in both auditory and tactile conditions despite adequate basic stimulus detection implicated dysfunction at the level of common crossmodal system sensitive to inter-stimulus duration. This system was interpreted by the authors as necessary for voluntarily imposing attention on an uninteresting task.[2]

A similar lateralised dimension to performance on a visual attentional measure was recently reported by Rueckert and colleagues (Rueckert & Grafman, 1996). When compared with patients with focal left-frontal lesions, patients with right frontal lesions, predominantly acquired as the result of penetrating gunshot wounds, showed impairment on a more conventional measure of sustained attention, the Continuous Performance Test (Rosvold, Mirsley, Sarason, Bransome, & Beck, 1956). In this task participants were asked to respond as quickly as possible to the appearance of the letter X on a screen but not to the appearance of any of 10 distractor letters. Not only were right frontal patients

[1] Few of the many experimental procedures described in this chapter are normed and standardised for clinical practice. Rosvold and Mirsky's Continuous Performance Test is available commercially, though this requires computer hardware to administer. Other classic vigilance tests based on Mackworth-type paradigms are available, but again these are lengthy and computer-administered (e.g. Zimmermann, 1993). It is increasingly apparent that measurement of sustained attention is possible over relatively short periods, and indeed those paradigms with the strongest validation against blood flow, functional imaging or lesion studies have almost without exception been relatively brief procedures. Wilkins et al.'s counting procedure, validated against lesion groups, has been adapted, standardised and normed as a clinical test of sustained attention—the elevator counting subtest of the Test of Everyday Attention (Robertson, Ward et al., 1994, 1996).

significantly slower in target responses than left frontal and control groups, they also showed a disproportionate slowing of RT during the second half of the test relative to the first half and made significantly more omissions. As we will discuss further, simple decrements such as that described by Rueckert et al. may be rather difficult to provoke, and reliance on such decrements in order to claim deficits in sustaining attention may underrepresent such difficulties. Godefroy et al. (1994), for example, did not detect decrements in the performance of patients with mixed focal frontal lesions. Patients with lesions following surgery to the anterior communicating artery performed an extended selective attention task. Responses had to be made to one class of three possible visual stimuli presented at randomly varying interstimulus intervals of between 5 and 15 seconds. No difference in performance was found between the two 12.5 minute halves of the test. This was interpreted by the authors as a failure to replicate Wilkins et al.'s result. However, the frontal group had increased reaction times (RT) to targets, greater errors of omission and commission (false—positives) and showed significantly greater variation in RT than did controls. As we will discuss further, this latter factor has been interpreted by Stuss et al. (Stuss, Stethem et al., 1989) as indicative of greater fluctuations in attention.

Woods and Knight (Woods & Knight, 1986), in reanalysing the data of Knight et al. (Knight, Hillyard et al., 1981), moved away from reliance on purely behavioural data. They considered elements within monitored electrophysiological responses (event-related potentials, ERPs) which are thought to form useful correlates of attentional function. On an auditory discrimination task five patients with focal lesions to right dorsolateral prefrontal cortex of various aetiology showed significantly reduced ERP negativity at longer interstimulus intervals relative to eight left dorsolateral patients and normal controls. The amplitude of ERP negativity has been associated with anticipation of a stimulus such as might occur in the interval between a cue and a target (Naatanen, 1982; Tecce, 1972). The reduced negativity in this group was accompanied by significantly more failures to detect target stimuli than either the left frontal patients or the controls.

Differences in the performance of executive tasks between patients with predominantly left or right focal frontal lesions were examined by Glosser and Goodglass (1990). Two versions of a non-verbal continuous performance task were included in the experimental measures. In the first participants were asked to respond to only the occurrence of the letter/symbol X when it occurred in a sequence of X and O symbols. In the second response to the X was made contingent upon its being preceded by the sequence XO. The results suggested a

[2] While it is clear that the task contained features which made performance for patients with right frontal lesions difficult, it is not yet clear to what extent "interest" or motivation, such as might be modulated by reward characteristics, contributes to failure.

somewhat complex laterality interaction. Twelve patients with frontal lesions were impaired relative to patients with posterior lesions on both tasks. On the simple task, left frontal patients showed a pattern of declining performance while right frontal patients appeared to be able to maintain initial performance, albeit at an impaired level. On the complex, successive task right frontal patients performed more poorly than all other groups, their failures being particularly characterised by an inability to withhold a response following the appearance of an XO sequence.

The studies of attentional function in patients with focal anterior damage described so far have all considered the maintenance of repeated stimulus—response couplings, whether this was implied by selective responses to a class of stimuli or try retaining a count of stimuli. Kertesz et al. (Kertesz, Nicholson et al., 1985) considered the maintenance of motor output to a single initial instruction. Included in these motor impersistence tasks were keeping the eyes closed, keeping the mouth open, and protruding the tongue. Each task required participants to maintain the instructed posture for 20 seconds. Again, patients with focal damage to the right hemisphere (from stroke) showed significantly greater impairment on such tasks than left-hemisphere patients matched on lesion size and age-matched controls. Although the number of patients with lesions exclusively to right anterior structures was small such damage appeared to be associated with particular difficulties in these tasks.

FUNCTIONAL IMAGING STUDIES

Studies of sustained attention in patients with focal lesions generally indicate that anterior damage leads to more difficulty than posterior damage and suggest that the integrity of right prefrontal structures is more impor tant than of those of the left. This localisation has received additional support from functional imaging studies of normal performance. In a positron emission tomography (PET) study, Cohen et al. compared the patterns of cerebral activation in normal individuals when making prolonged discriminations between auditory stimuli and when at rest (Cohen & Semple, 1988). The attentional condition was associated with increased activation in right prefrontal cortex. This activation was accompanied by *decreased* activation in anterior cingulate and posterior parietal areas. The extent of interconnectivity between these regions makes such a finding consistent with an inhibitory relationship. This finding was confirmed in a subsequent study by Cohen and colleagues (Cohen, Semple et al., 1992) which further localised the increase in right prefrontal cortex activity to the middle prefrontal gyrus.

In a task somewhat analogous to that used by Wilkins et al. (Wilkins, Shallice et al., 1987), Pardo et al. asked 23 normal participants to detect and keep a count of brief pauses in stimulation applied to the toe while changes in cerebral blood flow were monitored by a PET camera (Pardo, Fox et al., 1991). Relative to a

control condition of rest, significant activation occurred not only in somatosensory areas but also in right prefrontal and right superior parietal foci. The right prefrontal signal was apparent when participants were remaining alert for pauses in stimulation to either left or right foot. Left frontal regions showed no significant change in activity in either condition. In a further task participants were asked to monitor a visual spot at fixation and to detect and count incidence of dimming. While no actual dimming took place most participants reported false alarms. Again, right frontal and superior parietal areas were activated relative to rest, suggesting that this region was implicated in maintaining attention irrespective of stimulus modality.

Deutsch et al. examined patterns of cerebral activation in normal participants performing a variety of attention-demanding tasks. Conditions included monitoring a string of words to respond to any word which was repeated, to any occurrence of the phoneme cluster /br/ and to any bird name. The striking finding across these separate conditions requiring sustained attention to quite distinct aspects of the signal was a consistently asymmetric pattern of prefrontal activation, again with right prefrontal cortex showing the greatest activity (Deutsch, Papanicolaou et al., 1987).

The lesion and imaging studies discussed have differed in terms of whether the core task was one of selection—response or counting, whether attention was for simple or complex features of the stimulus and in stimulus presentation modality. Despite this diversity the results have been relatively consistent and generally support the notion that: (1) common processes necessary for maintaining alertness to relevant but possibly infrequent signals have been tapped; (2) that these appear relatively independent of the task details or stimulus modality; and (3) that these appear predominantly reliant on adequate right frontal function. The relative rarity of patients with focal frontal lesions means that further development in this area will tend to be rather slow. PET and MRI studies offer one route to a more detailed analysis of structure—function links across a more systematic variation of sustained attention task parameters. Another approach is to consider a much more ubiquitous group in whom frontal dysfunction is prevalent.

CLOSED-HEAD INJURY STUDIES

The forces exerted on the brain as a result of blunt impact or rapid deceleration can result in widespread diffuse damage due to stretching and shearing of nerve fibres in addition to more focal contusions and haemorrhages (Mattson & Levin, 1990). The potential for such widespread damage, including damage to the connections between structures, undermines precise functional localisation in this group. However, owing to the mechanics of such closed-head injuries (CHI), frontal regions and the connections between frontal cortex and posterior and

subcortical structures are disproportionately vulnerable to damage (Mattson & Levin, 1990; Stuss & Gow, 1992). These anatomical factors have been held to account for the similarities in behavioural deficits exhibited by both closed-head injury patients and by patients with focal lesions to the frontal lobes and have led to this group's being studied in terms of further exploring prefrontal function (Stuss & Gow, 1992; Gansler, Covall et al., 1996). If, as focal lesion and imaging studies suggest, the frontal cortex is centrally involved in the maintenance of attention it then becomes theoretically and clinically important to consider whether CHI survivors are particularly vulnerable to problems in this area.

Clinically, observations of problems in maintaining concentration following CHI have been widespread (e.g. Conkey, 1938; McKinlay, 1981; Rimel, Giordani et al., 1981, 1982). Experimental studies with vigilance paradigms have produced more varied results.

Parasuraman et al. considered the performance of 10 patients with a mild head injury on a task in which response was required for one of 10 visual digits presented at the fast rate of 60 per minute (Parasuraman, Mutter et al., 1991). Over a period of eight minutes performance did not differ significantly from that of controls, both being essentially at ceiling. When, however, the task was repeated under conditions where the stimuli were perceptually degraded following a procedure described by Nuechterlein et al. (Nuechterlein, Parasuraman et al., 1983), the patient group's performance in terms of target detection and false-positives was significantly below that of the controls. This was interpreted as reflecting increased demands for effortful, controlled processing in the degraded condition. While both groups showed a decrement in performance over the task there was no significant difference between them in the rate of decline relative to initial performance.

A similar failure to demonstrate disproportionate decrements in attentional performance has pervaded other studies (Brouwer & Van Wolffelaar, 1985; Ponsford & Kinsella, 1992; Spikman, Van Zomeren et al., 1996) but has not been universal. Whyte et al. sought to examine the question of decrement by manipulating the sustained attention task so that CHI patients and controls would start at a comparable level (Whyte, Polansky et al., 1995). To this end, the duration of stimulus exposure was adjusted for each participant to achieve a 75% success rate criterion. Using a very slow event rate of 10 items per minute over a relatively brief task duration of 14 minutes, 26 CHI patients of mixed severity showed a disproportionate decrement in detection in addition to being slower and less accurate overall than controls. This result is perhaps all the more surprising given the target to non-target ratio of 1:1, which would tend to act against criterion shifts related to target probability (Baddeley & Colquhoun, 1969).

Loken et al. considered the performance of 20 severely head-injured patients between one and nine months post-injury in a serial selection task at three levels of complexity (Loken, Thornton et al., 1995). The majority of the patients had

diffuse damage with five showing discernible contusions to frontal structures. The tasks, administered separately, involved increasing complexity of discrimination, from responding to a filled and not to an outline circle to responding when three filled circles were present in an array of eight circles. As with the Whyte et al. study the rate of stimulus presentation was very low at approximately six per minute with variable interstimulus intervals and target probability was high at 0.6. While, interestingly, there was no disproportionate effect of complexity in terms of the vigilance decrement for the CHI group, consideration of performance across tasks demonstrated that the patients not only performed more poorly (latency and accuracy) than controls but that their performance declined more rapidly than controls.

It is tempting to speculate that the much slower stimulus presentation rates in the Whyte et al. and Loken et al. studies, relative to those which did not show disproportionate decrements, account for this difference between their results. This picture is somewhat complicated, however, by the use of a warning cue between 2 and 5s prior to stimulus onset in the Loken et al. study (Loken, Thornton et al., 1995). We might expect that failure due to poor attentional maintenance on the task encouraged by the long inter-stimulus intervals would be somewhat ameliorated by such a cue. This speculation may still be entertained, given that the interval between cue and stimulus still exceeded the actual interstimulus intervals of the previous studies and that the cue did not predict of target status. It is possible, therefore, that despite the relatively high target probability in these studies, long unfilled intervals between targets placed increasingly unmet demands on an endogenous system for controlling attention in the absence of constant or frequent environmental elicitation.

The consistent finding from all of the studies discussed is that CHI patients are more likely to be impaired in terms of accuracy, speed, and consistency in performing these often ostensibly simple but prolonged tasks, whether or not that impairment is reflected in a disproportionate decrement. Moving away from abnormal *decline* in performance as the hallmark of sustained attention deficits, Stuss et al. have argued that poor attentional control may well be reflected in periods of adequate task performance, when attentional resources are directed towards the task, and periods of poor or absent performance, when resources are not so directed (Stuss, Stethem et al., 1989). While this distinction is difficult to make on the basis of accuracy alone, consideration of reaction times allows some inferences to be drawn. As slowness of response has been a frequently observed characteristic of head-injured groups, so too has increased variability in speed of response (Van Zomeren & Deelman, 1978; Stuss, Stethem et al., 1989; Whyte, Rose et al., 1994). It is often the case that reaction time data are smoothed to reduce the influence of outlying points. Stuss et al. (Stuss, Stethem et al., 1989) argued that such transformations in the context of attentional performance following CHI may miss a deficit in consistency. Considering the effects of

fatigue, selective, divided and sustained attention on a series of RT tasks, Stuss et al. demonstrated a significantly increased variability in performance by CHI patients which could not be accounted for by the overall difference in average RT.

Recently, we have reversed the usual target response characteristics of sustained attention tasks in order to consider the patterns of fluctuating attention following closed-head injury (Robertson, Manly et al., 1997). In conventional paradigms, targets require a response while non-targets are ignored. When targets are rare this situation considerably limits the amount of information available on ongoing performance. In the task (Sustained Attention to Response Task; SART), regularly presented non-target visual digits (at a presentation rate of 52 per minute) require responses while responses should be withheld for a rare target digit of three. In conventional paradigms, the effect of slowed responses may be confounded by the general slowing of all responses apparent following CHI. In this task, we reasoned, the effects of attentional fluctuation would be in the opposite direction. Owing to the regular, rhythmic pacing and the rarity of targets, the task encourages a strategy of fast, anticipatory, automatic responding. We argue that sustained attention to one's response pattern is demanded to resist this and so to withhold a response at an appropriate and unpredictable moment.

The results showed that errors were highly predictable from significantly faster RTs to non-targets in the trials immediately preceding the occurrence of a target. Thirty-four CHI patients, selected on the basis of being between 9 and 18 months post-injury, made significantly more errors than age-, sex-, and IQ-matched controls and this pattern was accompanied by significantly greater variations in RT across the task. As has been noted, the relationship between performance on sustained attention measures and clinical observations of concentration problems following CHI has been largely untested. We found that performance on the SART was a significant predictor of the complaints of relatives of CHI survivors, and of subjective complaints in normals on two standardised measures, the Dysexecutive Syndrome Questionnaire (DEX; Burgess, Alderman et al., 1996) and the Cognitive Failures Questionnaire (CFQ; Broadbent, Cooper et al, 1982).

The drawing of inferences about particular stages in processing and responding to stimuli on the basis of multiply determined RTs has been criticised (Sternberg, 1969). Consideration of event-related potentials (ERPs) provides an additional methodology for monitoring correlates of preparation and response to stimuli. Segalowitz, Dywan and Unsal (1997) monitored electrophysiological and behavioural responses to stimuli in 20 CHI patients of mixed injury severity and matched controls. Behaviourally, CHI patients again demonstrated significantly greater target RTs than controls and showed significantly more variation in RT than controls, independent of overall RT level.

Electrophysiologically, the amplitude of P300, occurring between 300 and 500msc following a stimulus and interpreted as reflecting attentional allocation, was significantly greater in the control group. For the patients, there was a significant correlation between P300 amplitude and variability in reaction time (higher attentional allocation resulted in greater consistency across the task). CNV amplitude also differed between patient and control groups. In the patients the extent of the negativity was a correlate of variability in RT (heightened anticipation of the stimulus resulted in greater consistency across the task).

THE CONTRIBUTION OF THE FRONTAL LOBES TO SUSTAINING ATTENTION

It has been proposed that the attentional functions of the human brain are subserved by subsystem networks which are anatomically separable from the processing systems they influence and composed of distinct areas which perform different functions (Posner, Petersen et al., 1988; Posner & Petersen, 1990). On the basis of the evidence discussed inferences about anatomical location of functions necessary for sustained performance must remain somewhat one-sided. Populations who have been exposed to the necessary diversity of measures have either had no lesion or have experienced diffuse damage from closed head injuries. Patients with focal lesions and normal participants in imaging studies have been exposed to a narrower range of tasks. However, in these cases, the importance of prefrontal cortex, in particular of the right hemisphere, has been a consistent finding.

Evidence from the study of normal performance in tasks of long duration requiring selection between responses appears to point to at least two principal and putatively separable causes for failure. The first lies in the capacity to discriminate stimuli of importance from noise under conditions of fast presentation with a requirement for effortful, controlled processing, which may be characterised as resource limitation. The second lies in the ability to maintain attentional resources on the task for long enough periods to perform it adequately, which may be characterised as impairment of attentional control. This ability appears most taxed in situations where stimuli are separated by long and unpredictable gaps and where the discrimination is rather undemanding.

In normals, the allocation of attention in vigilance tasks has been probed using a secondary task (Parasuraman, 1984). RT to these probes were significantly slowed when the main task was fast and demanding. Such a slowing may be considered as analogous to resistance to distraction in that events that competed with the main task for attention took longer to develop sufficient excitation to provoke a response. It may be considered, therefore, that in such paradigms the designed environment facilitates the process of staying on task by making a continual demand for sustained attention.

Where tasks are slow, irregular, and involve undemanding discrimination, such secondary probes attract faster RTs (Parasuraman, 1984). Allocating only such resources as are sufficient for the perceived demands of the task and retaining spare capacity for the processing of other, potentially important, environmental events may be highly adaptive. However, the tendency to miss targets, despite the apparent ease of the discrimination of targets from noise in these measures indicates that this strategy is imperfectly maintained. In these tasks, the designed environment facilitates attentional wandering because demands are very intermittent.

As we have noted, Mathews and Holley found that estimates of controlled processing resources derived from brief tasks, such as visual search, formed good predictors of performance over much longer durations only when the core, repeated task required effortful, controlled processing (Mathews & Holley, 1993). Deficits in selective and divided attention have been observed following frontal lesion (e.g. Bench, Frith et al., 1993; Corbetta, Miezin et al., 1991; Janer & Pardo, 1991; Pardo, Pardo et al., 1990). Reduced speed of processing has been the most ubiquitous finding in studies of cognitive function following closed-head injury (e.g. Van Zomeren & Deelman, 1978; Stuss, Stethem et al., 1989; Whyte, Polansky et al, 1995). Such findings make failure in *demanding* and *rapid* sustained attention tasks predictable consequences of frontal lesions. But to what extent are deficits in tasks of long duration which make apparently *low* and *intermittent* demands on such processing capacity predictable consequences of frontal injury?

The prefrontal cortex is densely connected to other areas of the brain. The nature of these connections provides some clues to a further contribution of the prefrontal cortex in sustaining attention. In functional imaging studies of sustained attention tasks, activation of right prefrontal cortex has been associated with a reduction of activity in the anterior cingulate and the posterior parietal cortex (Cohen & Semple, 1988; Cohen, Semple et al., 1992). These areas have been associated with attentional selection (e.g. Bench, Frith et al., 1993) and with spatial attention (Corbetta, Miezin et al., 1993; Mesulam, 1981; Posner, Inhoff et al., 1987), respectively. Posner and Petersen have equated the inhibition of function in these areas with the subjective sense of "clearing one's mind," which may be necessary to remain alert for an anticipated signal and resist becoming sidetracked by external or internal competitors for attention (Posner & Petersen, 1990). The prefrontal cortex also receives projections from and projects to the reticular structures of the brain stem. This provides a basis for the endogenous modulation of arousal (Heilman, Watson, Valenstein, & Goldberg, 1987). Via such mechanisms frontal signals may both act to damp arousal responses to salient but irrelevant stimuli and to promote arousal to stimuli which, for example because of repeated exposure, do not provoke such a response automatically but nevertheless remain relevant to the task at hand. Such a system

could also act to maintain anticipatory excitation in the absence of any external stimuli at all.

Such anatomical mechanisms appear to map well onto a cognitive model of prefrontal executive function proposed by Norman and Shallice (1980) and Shallice (1988). According to this model, many situations elicit well-learned, routine responses. Such responses or schemata (which may control highly complex combinations of motor and cognitive acts) are activated by an associated triggering input. The selection of a given schema will depend upon its level of activation relative to the level of competing signals and the influence of higher level controlling schemata. Via such a system apparently complex activities such as those involved in driving a car can be performed appropriately and in an automatic fashion (Shallice, 1988). There are, however, times when this system is not effective, for example when a situation is novel or when the most highly activated schema is inappropriate to an overall goal. In such situations, Norman and Shallice have argued, supervisory control must intervene to bias the competition between response schemata through excitatory or inhibitory modulation (Shallice, 1988). Recently, Stuss et al. have applied the concepts of supervisory control to the process of sustaining attention (Stuss, Shallice et al., 1995). They argue that, in a sustained attention task where relevant events are separated by long intervals, a selected schema (e.g. respond to red) will gradually lose activation and thereby control over the processes necessary to make the perceptual response selection. As the activation decays, the schemata will become increasingly vulnerable to competition from other irrelevant schemata which may be triggered by environmental or internal events. To prevent this happening, they argue, a supervisory attentional system would be required to monitor actively the level of activation of the relevant schema, to provide "top-down" re-energising of the schema when the monitoring detects excessive deactivation and to inhibit the activation of irrelevant schemata if they interfere with the overall goal.

The concept that sustained attention in slow and undemanding tasks is one function of a supervisory system offers a number of advantages. It allows for a more detailed analysis of task failure through manipulation of monitoring, activation, and inhibitory requirements. If such processes are common across tasks demanding supervisory control it should also be possible to predict sustained attention deficits from much briefer measures inasmuch as they make similar demands. This type of analysis also raises the possibility of targeting rehabilitation/compensation strategies at specific areas, for example, the use of periodic external cues to monitor the extent to which one is "on task."

CONCLUSION

The studies discussed indicate that: (1) problems in maintaining attention are prevalent in groups with frontal lesions; (2) variation in certain task parameters such as information processing speed and the demands of the discrimination required may elicit separable forms of impairment, one best characterised as a resource limitation under high demand, the other as a failure of attentional control under low or intermittent demand; and (3) in studies in which functional localisation is possible, mainly using slow and undemanding tasks, the predominant role of right hemisphere anterior structures, in particular the dorsolateral region, has been a common finding.

It would appear, therefore, that there exists a frontally based system, one of whose roles it is to maintain a state of task-specific alertness under circumstances where habituation and deselection of the current goal state would otherwise occur. Assessment of integrity of this function is complicated by the use of a wide range of paradigms and analytic approaches. While it seems probable that the involvement of such a system would be required to a greater or lesser degree in a variety of activities, its function is most apparent when demands on other cognitive systems are at a minimum, when inactive intervals without stimuli must be bridged and when little more is required than keeping one's mind on a single, simple goal. Conversely, in situations of challenging demands and where stimulus salience is maintained (through variety, rate or intensity), it could be anticipated that the demands on an internal activation system would be less, and failure on such tasks may be related to other causes than inability to maintain activation and attention.

The time course of the loss of attention during a task may be much shorter than formerly thought, and may be characterised by an intermittent and inefficient activation of task goals leading to fluctuating performance over time, rather than simply to a steady temporal decrement. Clinical tests now exist which tap this function, and these are proving useful in identifying hitherto unrecognised deficits in lesioned groups. The cognitive neuropsychology of the sustained attention system in relation to other attentional and cognitive systems is as yet barely elucidated. Already, however, the usefulness of assessing these functions has been demonstrated in rehabilitation (Robertson, Tegnér et al., 1995), as well as in constraining the understanding of deficits and pathologies as diverse as progressive supranuclear palsy (Esmonde, Giles et al., 1996) and unilateral neglect (Hjaltason, Tegnér, Tham, Levander, & Ericson, 1996; Robertson, Manly et al., in press).

REFERENCES

Baddeley, A.D. & Colquhoun, W.P. (1969). Signal probability and vigilance: A reappraisal of the signal-rate effect. *British Journal of Psychology, 60(2),* 169–178.

Bench, C.J., Frith, C.D. et al. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 31,* 907–922.

Broadbent, D.B. (1971). *Decision and stress.* London: Academic Press.

Broadbent, D.B., Cooper, P.F. et al. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1–16.

Broadbent, D.E. & Gregory, M. (1963). Vigilance considered as a statistical decision. *British Journal of Psychology, 54,* 309–323.

Brouwer, W.H. & Van Wolffelaar, P.C. (1985). Sustained attention and sustained effort after closed head injury: Detection and 0.10 Hz heart rate variability in a low event rate vigilance task. *Cortex, 21,* 111–119.

Burgess, P.W., Alderman, N. et al. (1996). The Dysexecutive Questionnaire (DEX). In B.A. Wilson, N.Alderman, P.W.Burgess, H.Emslie, & J.J.Evans (Eds.), *Behavioural assessment of the dysexecutive syndrome.* Bury St. Edmunds: Thames Valley Test Company.

Cohen, R.M. & Semple, W.E. (1988). Functional localization of sustained attention. *Neuropsychiatry, Neuropsychology and Behavioural Neurology, 1,* 3–20.

Cohen, R.M., Semple, W.E. et al. (1992). Metabolic brain pattern of sustained auditory discrimination. *Experimental Brain Research, 92(1),* 165–172.

Conkey, R.C. (1938). Psychological changes associated with head injuries. *Archives of Psychology, 232,* 1–62.

Corbetta, M., Miezin, F.M. et al. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience, 11,* 2383–2402.

Davies, D.R. & Parasuraman, R. (1982). *The psychology of vigilance.* New York: Academic Press.

Deutsch, G., Papanicolaou, A.C. et al. (1987). Cerebral blood flow evidence of right frontal activation in attention demanding tasks. *International Journal of Neuroscience, 36,* 23–28.

Egan, J.P., Greenberg, G.Z. et al. (1961). Operating characteristics, signal detectability and the method of free response. *Journal of the Acoustical Society of America, 33*, 993–1007.

Esmonde, T., Giles, E., Gibson, M., & Hodges, J.R. (1996). Neuropsychological performance, disease severity and depression in progressive supranuclear palsy. *Journal of Neurology, 243(9),* 638–643.

Fisk, A.D. & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors, 23,* 737–750.

Foster, J.K., Eskes, G.A. et al. (1994). The cognitive neuropsychology of attention: A frontal lobe perspective. *Cognitive Neuropsychology, 11(2),* 133–147.

Gansler, D.A., Covall, S. et al. (1996). Measures of prefrontal function after closed head injury. *Brain and Cognition, 30,* 194–204.

Glosser, G. & Goodglass, H. (1990). Disorders in executive control functions among aphasic and other brain damaged patients. *Journal of Clinical and Experimental Neuropsychology, 12,* 485–510.

Godefroy, O., Caberet, M., &Rousseaux, M. (1994). Vigilance and the effects of fatigability practice and motivation on simple reaction time tests in patients with lesion of the frontal lobe. *Neuropsychologia, 32,* 983–990.

Godefroy, O. & Rousseaux, M. (1996). Divided and focused attention in patients with lesion of the prefrontal cortex. *Brain and cognition, 30,* 155–174.

Gronwall, D.M.A. & Sampson, H. (1974). *The psychological effects of concussion.* Auckland: Auckland University Press.

Hecaen, H. & Albert, M.L. (1978). *Human neuropsychology.* New York: Wiley.

Heilman, K.M., Watson, R.T., Valenstein, E., & Goldberg, M.E. (1987). Attention: Behavioural and neural mechanisms. In F.Plum (Ed.), *Handbook of physiology, section 1: The nervous system* (pp. 461–481).

Hjaltason, H., Tegnér, R., Tham, K., Levander, M., & Ericson, K. (1996). Sustained attention and awareness of disability in chronic neglect. *Neuropsychologia,* 34(12), 1229–1233.

Janer, K.W. & Pardo, J. (1991). Deficits in selective attention following bilateral anterior cingulotomy. *Journal of Cognitive Neuroscience, 3,* 231–234.

Kertesz, A., Nicholson, I.et al. (1985). Motor impersistence: A right hemisphere syndrome. *Neurology, 35,* 662–666.

Knight, R.T., Hillyard, S.A. et al. (1981). The effects of frontal cortical lesions on event-related potentials during auditory selective attention. *Electroencephalography and Clinical Neurophysiology, 52,* 571–582.

Loken, W.J., Thornton, A.E. et al. (1995). Sustained attention after severe closed head injury. *Neuropsychology, 9(4),* 592–598.

Luria, A.R. (1966). *Higher cortical functions in man.* London: Tavistock.

Mackworth, N.H. (1950, 1961). Researches in the measurement of human performance. In H.A.Sinaiko (Ed.), *Selected papers on human factors in the design and use of control systems.* Dover: Dover Publications.

Mathews, G. & Holley, P.J. (1993). Cognitive predictors of vigilance. *Human Factors, 35 (1),* 3– 24.

Mattson, A.J. & Levin, H.S. (1990). Frontal lobe dysfunction following closed head injury: A review of the literature. *Journal of Nervous and Mental Disease, 178,* 282–291.

McKinlay, W.M. (1981). The short-term outcome of severe blunt head injury as reported by relatives of the injured persons. *Journal of Neurology, Neurosurgery and Psychiatry, 44,* 527– 533.

Mesulam, M.M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology, 10,* 309–325.

Naatanen, R. (1982). Processing negativity: An evoked potential reflection on selective attention. *Psychological Bulletin, 92,* 605–40.

Norman, D.A. & Shallice, T. (1980). *Attention to action: Willed and automatic control of behaviour.* Centre for Human Information Processing (Technical Report No. 99). (Reprinted in revised form in R.J.Davidson, G.G.Schwartz, & D.Shapiro [Eds.] [1986] *Consciousness and self-regulation* [Vol. 4]. New York: Plenum Press.)

Nuechterlein, K.H., Parasuraman, R. et al. (1983). Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science, 220,* 327–329.

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science, 205,* 924–927.

Parasuraman, R. (1984). Sustained attention: A multifactorial approach. *Attention and Performance, XI,* 493–511.

Parasuraman, R. & Davies, D.R. (1977). A taxonomic analysis of vigilance. In R.R.Mackie (Ed.), *Vigilance: Theory, operational performance and physiological correlates.* New York: Plenum Press.

Parasuraman, R., Mutter, S.A. et al. (1991). Sustained attention following mild closed-head injury. *Journal of Clinical and Experimental Neuropsychology, 13(5),* 789–811.

Parasuraman, R., Warm, J.S. et al. (1987). Vigilance: Taxonomy and utility. In L.Mark, J.S. Warm, & R.L.Huston (Eds.), *Ergonomics and human factors: Recent research.* New York: Springer.

Pardo, J.V., Fox, P.T. et al. (1991). Localization of a human system for sustained attention by positron emission tomography. *Nature, 349,* 61–64.

Pardo, J.V., Pardo, P. et al. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Science USA, 87,* 256–259.

Ponsford, J. & Kinsella, G. (1992). Attentional deficits following closed-head injury. *Journal of Clinical and Experimental Neuropsychology, 14(5),* 822–838.

Posner, M.I., Inhoff, A.W. et al. (1987). Isolating attentional systems: A cognitive-anatomical analysis. *Psychobiology, 15,* 107–121.

Posner, M.I., & Petersen, S.E. et al. (1988). Localization of cognitive operations in the human brain. *Science, 240,* 1627–1631.

Posner, M.I. & Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Rimel, R.W., Giordani, B. et al. (1981). Disability caused by head injury. *Neurosurgery, 9,* 221–228.

Rimel, R.W., Giordani, B. et al. (1982). Moderate head injury: Completing the clinical spectrum of brain trauma. *Neurosurgery, 11(3),* 344–351.

Robertson, I.H., Manly, T., Beschin, N., Daini, R., Haeske-Dewick, H., Hömberg, V., Jehkonen, M., Pizzamiglio, L., Shiel, A., & Weber, E. (in press). Auditory sustained attention is a marker of unilateral spatial neglect. *Neuropsychologia.*

Robertson, I.H., Manly, T., et al. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35(6),* 747–758.

Robertson, I.H., Tegnér, R., Tham, K., Lo, A., & Nimmo-Smith, I. (1995). Sustained attention training for unilateral neglect: Theoretical and rehabilitation implications. *Journal of Clinical and Experimental Neuropsychology, 17,* 416–430.

Robertson, I.H., Ward, A., Ridgeway, V., & Nimmo-Smith, I. (1994). *Test of Everyday Attention,* Flempton: Thames Valley Test Company.

Robertson, I.H., Ward, A., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The test of everyday attention. *Journal of the International Neuropsychological Society, 2,* 523–534.

Rosvold, H.E., Mirsley, A.F., Sarason, I., Bransome, E.D., & Beck, L.M. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20,* 343–350.

Rueckert, L. & Grafman, J. (1996). Sustained attention deficits in patients with right frontal lesions. *Neuropsychologia,* 54(10), 953–963.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

See, J. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117,* 230–249.

Segalowitz, S.J., Dywan, J. & Unsal, A. (1997). Attentional factors in response time variability after traumatic brain injury: An ERP study. *Journal of the International Neuropsychological Society, 3,* 95–107.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shallice, T. & Burgess, P. (1991). Deficit in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Smith, A. & Nutt, D. (1996). Noradrenaline and attention lapses. *Nature, 380,* 291.

Spikman, J.M., Van Zomeren, A.H. et al. (1996). Deficits of attention after closed head injury: Slowness only. *Journal of Clinical and Experimental Neuropsychology, 18 (1),* 1–13.

Sternberg, S. (1969). On the discovery of the processing stage. *Acta Psychologica, 30,* 276–315.

Stuss, D.T. & Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Buttetin, 95,* 3–28.

Stuss, D.T. & Gow, A. (1992). Frontal dysfunction after traumatic brain injury. *Neuropsychiatry, Neuropsychology and Behavioural Neurology, 5(4),* 272–282.

Stuss, D.T., Shallice, T. et al. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences, 769,* 191–209.

Stuss, D.T., Stethem, L.L. et al. (1989). Reaction time after head injury: Fatigue, divided and focused attention and consistency of performance. *Journal of Neurology, Neurosurgery and Psychiatry, 79,* 81–90.

Swets, J.A. (1977). Signal detection theory applied to vigilance. In R.R.Mackie (Ed.), *Vigilance: Theory, operational performance and physiological correlates.* New York: Plenum Press.

Tecce, J.J. (1972). Contingent negative variation and psychological processes in man. *Psychological Bulletin, 77,* 73–108.

Van Zomeren, A.H. & Deelman, B.G. (1978). Long term recovery of visual reaction time after closed head injury. *Journal of Neurology, Neurosurgery and Psychiatry, 48,* 21–28.

Vickers, D. & Leary, J.N. (1983). Criterion control in signal detection. *Human Factors, 25,* 283–296.

Whyte, J., Polansky, M. et al. (1995). Sustained arousal and attention after traumatic brain injury. *Neuropsychologia, 33(7),* 797–813.

Whyte, J., Rose, T. et al. (1994). Quantification of attention-related behaviours in individuals with traumatic brain injury. *American Journal of Physical Medicine and Rehabilitation, 73(1),* 2–9.

Wilkins, A.J., Shallice, T. et al. (1987). Frontal lesions and sustained attention. *Neuropsychologia, 25,* 359–365.

Woods, D.L. & Knight, R.T. (1986). Electrophysiologic evidence of increased distractibility after dorsolateral prefrontal lesions. *Neurology, 36,* 212–216.

Zimmermann, P. & Fimm, B. (1993). *The Test for Attentional Performance.* Würselen: Psytest.

CHAPTER SEVEN
How Specific are the Memory and Other Cognitive Deficits Caused by Frontal Lobe Lesions?

Andrew R.Mayes *Department of Clinical Neurology, Royal Hallamshire Hospital, Sheffield, U.K.*

Irene Daum *Clinical Neuropsychology Unit, Faculty of Psychology, Ruhr-University of Bochum, Bochum, Germany*

INTRODUCTION

The prefrontal cortex constitutes a considerable proportion of the human cortical mantle, over 29% according to Fuster (1989). It seems probable, therefore, that different prefrontal regions will mediate distinct psychological functions. Indeed, lesions of the prefrontal cortex have been shown to disrupt a variety of psychological functions to produce disorders of movement control, language, problem solving, memory, and affect. Nevertheless, it is widely believed that the prefrontal cortex has a primary, unitary function. This broad unitary function is supposed to involve the supervision or execution of other cognitive processes so as to achieve intended goals (see Luria, 1972; and Shallice, 1982, 1988 for different accounts of this view). Deficits produced by lesions of the prefrontal cortex are interpreted as resulting from various kinds of executive disturbance. The view is clearly compatible with the idea that different prefrontal lesions cause functionally distinct deficits with the proviso that all of these are forms of "dysexecutiveness". There is, however, surprisingly little evidence to show that different prefrontal lesions cause distinct psychological deficits and, in particular, distinct forms of dysexecutiveness (see Kolb & Whishaw, 1990; Passingham, 1993 for reviews of the deficits found, but see Shallice & Burgess, 1996, for a consideration of how dysexecutiveness may be fractionated).

The prefrontal cortex is connected to structures in many other parts of the brain including the posterior association cortex, limbic structures, the basal ganglia, and other subcortical structures. Many of these connections are either reciprocal or form part of a circuit. Damage to other structures in these circuits may, therefore, produce similar deficits to those caused by lesions to the prefrontal cortex. This raises the question of the extent to which deficits caused by lesions to the prefrontal cortex are selective to that region. There are at least two reasons why lesions of other brain regions might disrupt performance in a

way similar to that found after prefrontal lesions. The first is the reason just given, namely, that the disrupted function is mediated by a neural circuit comprising the prefrontal cortex and other brain regions. If this reason applies the non-frontal lesion would be disrupting exactly the same function as the prefrontal lesion. The second reason is somewhat different in nature and would mean that the other lesion disrupts another function even though it disturbs performance on the same task. This would occur because performance on many complex tasks depends on using basic cognitive processes (such as memory) in a planned and organised fashion. To do this effectively one must have an adequate level of efficiency of those basic cognitive processes and an adequate ability to execute plans. If intact planning ability has to be used to control impaired basic cognitive processing, the resulting output may look like the product of poor planning. Therefore, even if damage to the prefrontal cortex specifically disrupts planning ability and damage to another brain region specifically disrupts a basic cognitive process, the end result of both kinds of damage may be a similar disruption of performance on a complex task. In principle, more subtle measures should be able to distinguish between the two kinds of disturbance, but this may require further testing.

NON-MEMORY COGNITIVE TESTS

The Wisconsin Card-Sort Test (WCST) is sensitive to the ability to identify abstract categories and shift cognitive set when this becomes appropriate. It has long been argued that this test is relatively selective to prefrontal cortex lesions and, in particular, that such lesions produce abnormal levels of perseverative responding. However, prefrontal cortex lesions do not always disrupt performance on this task (Kolb & Whishaw, 1990). Anderson et al. (1991) compared preformance on this task in 91 patients with stable focal brain lesions in whom the extent of brain damage was estimated using magnetic resonance imaging (MRI) and computerised tomography (CT) procedures. No significant differences were found on any WCST performance measures between those with frontal and non-frontal lesions. Other patients who had extensive frontal cortex lesions performed well on the task and some with lesions outside this region performed very badly. Indeed, it was found that some patients with focal lesions in the occipital, temporal, subcortical, or parietal regions, identified by MRI and CT, achieved very few categories and made a large number of perseverative responses. Anderson et al. (1991) also found no evidence of performance differences based on differently located prefrontal cortex lesions, and found no association between the size of prefrontal cortex lesions and the severity of the deficit on the WCST. Their results differed from previous work in this area which had suggested that dorsolateral prefrontal cortex lesions cause patients to make more perseverative responses than do orbitofrontal cortex lesions (see Shallice, 1988 for a review). In discussing their study the authors argued that

chronic prefrontal cortex lesions may not cause a selectively severe deficit on the WCST but that recent lesions of this region, whether resulting from lobectomy (Milner, 1963) or from multiple sclerosis (Arnett et al., 1994), may at least cause a more severe deficit than would result from a non-frontal cortex lesion. Nevertheless, most patients considered in the literature have chronic lesions of the frontal cortex like those of the patients studied by Anderson et al. (1991) and should not show WCST deficits that are selectively more severe than those found in patients with chronic lesions of several non-frontal brain regions. If non-frontal lesions are disrupting something other than some aspect of planning ability, this is not apparent from the WCST because all aspects of performance on the task are disrupted to the same extent in patients with these lesions as they are in patients with frontal lobe lesions. To demonstrate that the two groups have different kinds of deficit would, therefore, require the use of further tests.

A recent positron emission tomography (PET) challenge task (Berman et al., 1995) has addressed the question of the extent to which performing the WCST activates brain regions other than the frontal cortex. This task compared the brain activation produced by performing the WCST with that produced by performing a specially designed sensorimotor control task. Performing the WCST was found not only to activate several frontal regions and, in particular, the dorsolateral prefrontal cortex, but also the inferior parietal lobe, the visual association cortex, the inferior temporal lobe and portions of the cerebellum. This work seems to be broadly consistent with the lesion study of Anderson et al. (1991), but it is also consistent with that of Milner (1963, 1964), which found that WCST performance was affected by dorsolateral prefrontal cortex lesions but not by orbitomedial lesions. Presumably, with chronic lesions some functional reorganisation may take place such that different frontal regions play a more equivalent functional role in WCST performance. Performance of the WCST clearly involves many processes so, in order to reduce these, Berman et al. (1995) compared naive performance with performance when participants were experienced with the task so that factors such as anxiety, rule learning, and formation of conceptual sets were probably no longer engaged. At both naive and experienced stages, however, working memory was still relevant to perfor mance so it was interesting that the regions that remained active in both conditions included the parietal and frontal sites that have been activated in PET challenge studies which examined working memory (for example, Paulescu, Frith, & Frackowiak, 1993). It remains unclear, however, whether these activations reflect the operation of a short-term storage or rehearsal process, or the operation of a willed action. In other words, Berman et al.'s PET study makes clear that we still do not know why patients with frontal cortex lesions are impaired at the WCST.

There has been no systematic exploration of whether the full range of "executive" tests that are believed to be selectively sensitive to prefrontal cortex lesions really are so. Such an exploration should clearly include fluency tests

(Benton, 1973), the cognitive estimates test (1978), the Tower of London or its variants (Shallice, 1982) and other perhaps less widely used tasks such as the Stroop test. These various tests probably tap different aspects of executive function although this remains to be properly elaborated, and they should be disrupted by damage to different regions of the prefrontal cortex. Being complex tasks, it is likely that they will be disrupted by lesions in other brain regions. If this occurs one might hope that the other lesions will disrupt task performance in different and characteristic ways. This would be expected if the other lesions were disturbing the task-critical underlying processes. If there is no evidence that non-frontal lesions have produced an impairment for different reasons than frontal lesions as with Anderson et al.'s (1991) results, this interpretation remains possible. However, further work would be needed to eliminate the other interpretation which is that both kinds of lesion are disrupting similar aspects of executive function. This second interpretation may also become more plausible when the prefrontal cortex and the other brain region are closely and intimately interconnected as is the case with the prefrontal cortex, parts of the basal ganglia such as the caudate, and parts of the thalamus. One might well believe that lesions of all these structures may produce similar kinds of "dysexecutiveness".

MEMORY DEFICITS CAUSED BY FRONTAL LESIONS

Prefrontal association cortex lesions might cause memory deficits for three reasons. First, as Mishkin (1982) has argued, lesions of the ventromedial frontal cortex (the posterior part of the orbital and medial frontal lobe), which includes the projection sites of the medial temporal lobe limbic system structures, might cause a form of the organic amnesia syndrome. Direct evidence for this in humans is, however, still hard to find, although Petrides (1989) cites observations of amnesia in cases with basal frontal tumours (for example, Luria, 1976). The problem with such cases is that the damage may well extend beyond the frontal cortex into structures in the basal forebrain. Second, frontal lesions may disrupt memory as a secondary effect of their disruption of the coordination, monitoring, and organising of online information processing (Luria, 1972; Shallice, 1988; Shimamura, 1994). If differently located frontal lesions can be shown to disrupt different aspects of planning ability, then one might also expect differently located frontal lesions to disrupt somewhat different aspects of memory. The third possibility is that, as Goldman-Rakic (1988) has argued, frontal lesions disrupt working memory in a material-specific manner that depends on the precise location of the lesion. This kind of disruption might also be expected to disrupt the ability to organise information on-line and so have a secondary effect on memory very similar to that of the second account. What is plain is that whether or not some frontal cortex lesions cause an amnesia, lesions in this region disrupt strategic organisation of cognitive processes, although whether this disruption is caused by disturbing the region's primary executive

function or whether it is secondary to a disturbance of working memory will not be considered critically in this chapter.

If at least some frontal cortex lesions cause an amnesia, then they should disrupt recognition as well as recall for pre- as well as postmorbidly experienced facts and events while not disturbing intelligence as well as immediate memory (see Mayes, 1988). Typically, however, frontal cortex lesions do not have this effect. Although the issue of whether or not frontal lesions disrupt intelligence has had a controversial history, Duncan, Burgess & Emslie (1995) have shown recently that patients with frontal cortex lesions who have superior intelligence as assessed by the Wechsler Adult Intelligence Scale may have impairments of from 20 to 60 points on measures of fluid intelligence that tap novel problem-solving ability. Frontal cortex lesions also seem to cause a mild disruption of immediate memory. Thus, Janowsky et al. (1989) found that patients with highly focal frontal cortex lesions showed immediate memory deficits as indicated by their impaired digit span performance, and Pigott and Milner (1995) found that right frontal lobe excisions resulted in impaired immediate memory for a complex pattern. Owen et al. (1995) have also shown that frontal patients were impaired at a spatial working memory task, a test of immediate spatial memory. This impairment (and perhaps the other immediate memory deficits that have been observed in frontal patients) may have resulted from the inefficient search strategies that Owen et al. noted in their frontal patients. More critically for the amnesia hypothesis of frontal cortex memory deficits is the frequent finding that patients with selective frontal cortex lesions show impairments in free recall, but not in recognition of postmorbidly encountered information (Hirst & Volpe, 1988; Incisa Della Rocchetta, 1986; Incisa Della Rocchetta & Milner, 1993; Janowsky et al., 1989; Jetter et al, 1986; Stuss et al., 1994; Gershberg & Shimamura, 1995; Daum et al., 1995).

Although amnesics often show severe impairments of recognition memory, Aggleton and Saunders (1997) have recently performed a meta-analysis which suggests that patients with selective damage to the hippocampus, fornix, mammillary bodies, or anterior thalamus may be markedly impaired on measures of free recall, but only marginally, if indeed significantly, impaired at their performance on Warrington's (1984) Recognition Memory Test. Some amnesics may, therefore, show a similar pattern of performance on anterograde tests of recall and recognition to some patients with frontal cortex lesions. The amnesics' pattern of performance is not, however, likely to result from their inability to organise their encoding and retrieval processes effectively unless they have additional damage to the frontal cortex. There is, in fact, evidence that amnesics encode information at a normal rate and in a normal way (Mayes et al., 1993). In contrast, there is good reason to suppose that the free recall deficit seen in patients with frontal cortex lesions is secondary to their impaired use of organisational strategies at encoding and retrieval. Indirect support for this claim

is provided by Daum et al. (1995), who found that patients with frontal cortex lesions were impaired at freely recalling once-presented word lists that contained randomly ordered, but semantically categorised, words. They were not, however, impaired at recalling either unrelated lists of words or lists comprising semantically categorised and organised words. Normal individuals are probably unable to exploit their superior organisational skills to encode semantically the once-presented unrelated word lists, and these skills are probably not required for semantically encoding the organised, semantically categorised lists. With the random, but categorised lists, however, these organisational skills are valuable even with one presentation trial.

Performance on tests of free recall is relatively unguided during study and at test little guidance is provided by cues. For this reason, the way individuals use encoding and retrieval strategies is likely to be more important than with recognition where far more guidance is provided at test and where elaboration during study generally has little beneficial effect (in contrast, free recall depends critically on the ability to form links between items during study). There is direct evidence that patients with frontal cortex lesions do not use organisational strategies well. Thus, Gershberg and Shimamura (1995) have shown that these patients displayed less subjective organisation of words during free recall tests of unrelated items, and less subjective organisation and category clustering during free recall tests of related items. Consistent with this, the patients were less likely than normal controls to report using strategies that would lead to the formation of associations between items. Gershberg and Shimamura also found that the frontal patients benefited from strategy instruction either at study or at test which suggests that the patients were poor at organising both their encoding and their retrieval processes, but that external instruction could to some degree replace the organising role normally played by the frontal cortex.

Interestingly, there is some evidence from memory challenge studies with PET that the left frontal cortex is activated during elaborative encoding particularly of semantic information, whereas the right frontal cortex is activated during retrieval (Shallice et al., 1994; Tulving et al., 1994). These studies do not indicate what the frontal cortex is doing during encoding or retrieval, but in a single photon emission computed tomography (SPECT) study that examined the activation processes accompanying recall of verbal associates learnt in a previous study episode, the performance of two groups of normals was compared using a regions of interest analysis (Montaldi et al., 1995). One group had near ceiling levels of recall whereas conditions were made harder for the second group so that its recall levels fell between 50 and 60% correct. It is reasonable to suppose that this second group had to engage in a more effortful search process. The second group showed significantly more right frontal cortex activation than the first group. As it had retrieved far fewer items, this activation could not be a reflection of retrieval success, but probably indicates the greater degree of search

effort that the second group had to exert during the cued recall test. This is consistent with the right frontal cortex playing a role in organising retrieval search processes.

Although the strategic organisation of encoding and retrieval processes is likely to be much more important for the efficiency of free recall than of recognition, it is likely that performance on some recognition tests would be impaired by frontal cortex lesions to the extent that they disrupt strategic organisation. Kesner et al. (1995) have systematically examined recognition in a group of patients with frontal cortex lesions. These patients were compared with controls on both item and temporal order recognition memory or spatial location, word, abstract picture, and hand position information. The patients showed no deficits on item recognition memory except for hand position information. In clear contrast, the patients had severe temporal order recognition memory deficits for all the kinds of information used. Interestingly, although left and right as well as bilateral lesions disrupted order recognition memory for words, abstract pictures and hand positions, left-sided lesions had no effect on spatial information order recognition memory. Kesner et al. (1992) found that patients with relatively focal hippocampal lesions did not show this kind of dissociation between item and order recognition memory, but had equivalent impairments for both kinds of recognition. If the deficit in item recognition memory proves replicable, then it can be said that this kind of temporal lobe lesion causes a different pattern of memory breakdown from that caused by frontal cortex lesions.

The findings of Kesner and his colleagues agree with extensive previous work that indicates an effect of frontal cortex lesions on temporal order memory (McAndrews & Milner, 1991; Milner et al., 1985; Shimamura et al., 1990), but has found no effect of these lesions on long-term spatial memory (Smith & Milner, 1984). Several comments need to be made about the effects of frontal lobe lesions on recognition memory. First, item recognition memory is minimally affected, probably because it depends to only a negligible extent on the formation of associations between items, and it is this encoding and retrieval of associations (critical for free recall) that probably requires strategic planning. Impairment of hand position recognition memory very likely constitutes an exception because the prefrontal cortex may selectively encode and represent feedback from the kinds of motor movement information relevant to encoding hand position, but is not concerned with the encoding of word, picture and spatial information in the same way. As Kesner and his colleagues discuss, frontal patients have been shown to have problems with egocentric localisation in other studies. It is also possible that, in so far as an impairment in monitoring one's own responses underlies the deficit, a similar functional deficit also underlies the impairment in self-ordered pointing that has been observed in frontal patients (Petrides & Milner, 1982). This ability is tapped by presenting patients with a series of cards

on each of which are the same 12 items in different randomly assigned positions. They are presented the cards one at a time and have to point to each item in turn (one per card) in a self-determined order. Clearly the poor ability to monitor their own responses leads to the poor use of strategy shown by frontal patients, which probably leads to their poor memory performance on this self-ordered pointing task.

Second, temporal order recognition may be relatively severely impaired because this kind of memory probably depends on the effortful and organised encoding and retrieval of associative information of a kind essential for retrieving the temporal order in which items were presented. Very little is known about the kinds of associative encoding most beneficial for temporal order memory, but it seems intuitively very likely that these forms of encoding play an important role. Remembering the temporal location of remote public events such as the Chernobyl disaster involves in most cases retrieving related public and personal events that eventually give a temporal fix. Similarly, remembering the temporal location of items in lists or other recently experienced events depends on retrieving similar public and personal associations. Evidence that a failure of strategic organisation may underlie the temporal order recognition memory deficit shown by frontal patients is provided by the finding that the deficit can be completely overcome if patients have to perform actions with common objects that are presented in a specific temporal order (Butters et al., 1994; McAndrews & Milner, 1991). The use of naming, visual imagery, experimenter-performed actions and verbal elaboration did not have this effect so the encoding that is enforced by patient-performed actions must involve the kind of associations that are critical to temporal order memory. They presumably normally require frontally mediated strategic organisation.

Third, if spatial recognition memory is not impaired, then one must suppose that the kinds of associative encoding and retrieval required for this form of memory do not require more than a negligible level of strategic organisation. Furthermore, one must suppose that source memory does, as this form of memory is impaired by frontal lobe lesions (Janowsky, Shimamura, Kritchevsky, & Squire, 1989). This is slightly puzzling because spatial as well as temporal information seems relevant for identifying the source of an experienced item of information. Nevertheless, current evidence is consistent with the view that frontal contextual memory as well as item free recall deficits all arise as secondary consequences of failures of effortful strategic organisation rather than of automatic processes as has been suggested in the past (see Mayes, 1988 for a discussion).

As already indicated, frontal cortex lesions are not the only ones that produce relatively selective free recall deficits, as lesions of at least some parts of the hippocampal circuit (such as the fornix) may disrupt free recall, but have relatively little effect on forced-choice item recognition memory (Aggleton & Saunders,

1997). However, unlike the frontal free recall deficit, the pattern of the hippocampal circuit deficit does not include a set of problems with strategic organisation. But severe and selective free recall deficits have also been reported in patients with lesions in other non-frontal sites. For example, Hanley et al. (1994) found that a patient with relatively selective damage to the head of the left caudate nucleus following the rupture and repair of an anterior communicating artery aneurysm had a severe deficit in verbal free recall tasks, but performed completely normally on verbal recognition tests as well as non-verbal recall and recognition tests. As the head of the caudate nucleus is connected in a tight serial circuit with the prefrontal cortex, the nature of the deficit in this patient may be the same as that seen after frontal cortex lesions. This was not formally tested, and so remains a possibility. However, although, like frontal patients, this patient showed a mild recall deficit for some premorbidly acquired information (Shimamura et al., 1990), she showed no evidence of having an immediate memory deficit or a problem with temporal order recognition memory, and her free recall deficit was not reversed by giving her imagery or levels of processing instructions during study. Even so, if the different measures and manipulations used by Gershberg and Shimamura (1995) had been employed, this patient might have behaved similarly to their frontal patients.

Failure to find that the patient of Hanley and his colleagues was exactly like the frontal patients of Janowsky and his colleagues in all respects might simply be a result of the functional heterogeneity of different parts of the frontal cortex and the non-frontal regions with which they form circuits. If frontal memory deficits are caused by distinct execu tive impairments, then these functionally distinct memory deficits may be produced by differently placed lesions that disrupt different executive processes. This issue has not been systematically explored, but functional dissociation nevertheless almost certainly occurs. For example, Delbecq-Derouesne et al. (1990) have reported the currently unique case of a patient who, following an anterior communicating artery aneurysm that caused frontal cortex damage, developed a severe impairment on recognition memory tests, but seemed to perform relatively normally on tests of free recall. More detailed inspection suggests that this patient did not perform entirely normally on free recall tests, because he made a high number of intrusion errors. According to Shallice's (1988) account of the mode of operation of the frontal cortex supervisory system, it can break down in at least two ways. One kind of breakdown would lead to an impairment in the accurate evaluation of any retrieved information. This would lead to a problem with recognition because items should not be easily assessable as old or new, but would not necessarily affect the number of items that a patient could successfully recall although it would probably lead to an increase in intrusion errors. Interestingly, it should also lead to confabulation, which has frequently been reported after frontal cortex lesions (Stuss & Benson, 1984), and possibly to difficulties with judging

the frequencies of recently encountered items, which has also been reported to be impaired after frontal cortex lesions (Milner et al., 1985). The second kind of breakdown might take two forms, one of which might make it difficult to set up search strategies and the other of which might make it difficult to employ associative encoding strategies. If either one or both of these problems applied, then free recall might be relatively selectively affected. A breakdown in the use of either appropriate encoding strategies, appropriate search strategies, or both, may be involved in the anterograde free recall deficits shown by frontal patients, but only a breakdown in the use of search strategies can be implicated in the selective deficits in recalling premorbidly experienced information (recognition is unaffected) shown by frontal patients (Mangels, Gershberg, Shimamura, & Knight, 1996). As already indicated, it could be that disturbances of encoding strategies are more associated with left frontal cortex lesions whereas problems with search strategies are more associated with right frontal cortex lesions (for example, see Shallice et al., 1994).

Frontal patients also show other memory deficits that may prove dissociable from the memory deficits discussed so far. First, they show meta-memorial deficiencies (see Mayes, 1988), such as a pathologically inaccurate feeling of knowing (Shimamura et al., 1991). As metamemory can be conceived of as the evaluation of what knowledge is held in memory and must depend on the retrieval of information within a knowledge-based context, it is easy to see how it involves strategic planning processes. It is also likely that accuracy of metamemorial judgements depends on evaluative processes that are also mediated by the frontal cortex.

Second, it has been claimed that some patients with frontal lobe lesions have problems with prospective memory (see Shimamura et al., 1991; Burgess, personal communication). They have difficulty remembering to carry out physical or mental actions either at some specified future time or at a future time when a specific signal is presented. This may apply even when their memory for recent events seems normal so that they can describe what they have been trying to remember for prospective purposes.

Third, frontal patients show abnormal sensitivity to interference in various memory situations. These may be of several different kinds. For example, it is unclear that poor ability to perform under the dual-task conditions of the Brown —Peterson task results from the same deficit as does failure to show release from proactive interference following a shift of the semantic category of to-be-remembered items (see Mayes, 1988 for a discussion). Moscovitch (1982) found that failure to show release in frontal patients was associated with poor performance on the Wisconsin Card-Sort Test. As he found that patients with temporal lobe lesions managed to show successful release, it would be interesting to find out whether the degrees of release shown by patients with other brain lesions do or do not correlate with the extent of their impairment on

the Wisconsin Card-Sort Test. Shimamura et al. (1995) have also shown that patients with frontal lobe lesions are abnormally susceptible to proactive interference in A—B, A—C learning paradigms of two kinds. In the first, they are required to learn related word pair lists (for example, "lion-hunter, lion-circus") and, in the second, they are required to learn unrelated word pairs, for which the response words are recombined with the stimulus words in the A—C list. With both learning paradigms, the patients exhibited particularly poor performance during the initial trial of the second list as well as an abnormal tendency to make prior list intrusions. They also showed evidence of abnormally high levels of retroactive interference when required to relearn the A—B list.

Shimamura and his colleagues considered the interesting possibility that the abnormal susceptibility of frontal patients to A—B, A—C interference may underlie their other memory impairments in free recall, source memory, and memory for temporal order. These forms of memory would all be likely to be disrupted if many wrong associations could not be inhibited at retrieval. Shimamura (1994) has argued that the frontal cortex may be concerned with a general gating function and that disruption of this gating function may be responsible for many of the executive disorders associated with frontal cortex lesions. The specific deficit observed would depend on which other brain region is not appropriately gated as a result of a frontal cortex lesion. Thus, basic perceptual functions may be disrupted after some frontal lesions because such lesions have been shown to potentiate middle latency evoked potentials (Knight et al., 1989), and inability to inhibit incorrect competing responses at retrieval may occur when frontal lesions disrupt a modulatory input to the medial temporal lobe regions implicated in amnesia. Two points need to be made in relation to this hypothesis. The first is whether lesions in the target site or in the connecting fibres between the frontal cortex and the target site have the same effect, in which case medial temporal cortex lesions or lesions that disconnect this region from the frontal cortex should increase susceptibility to interference. The second point is that the kinds of memory disorder resulting from frontal cortex lesions may often have two or more potential underlying causes. For example, temporal order recognition memory deficits could result from frontal cortex lesions because patients have a gating problem that increases their susceptibility to interference, a strategic organisation deficit that affects both encoding and retrieval, or a deficit in evaluating retrieved responses, or some combination of these.

If there are several distinct memory disorders caused by frontal lobe lesions, then it should be possible to observe multiple dissociations in patients who have frontal lobe damage. Single dissociations, i.e. failure to find certain memory deficits in some frontal patients, are of little interest because these deficits may merely require a greater degree of disturbance of a single underlying function before they become apparent. A single dissociation by itself, therefore, provides no evidence for the existence of two or more processes. Support for this comes

only from double or multiple dissociations. There has unfortunately been no systematic exploration of whether frontal cortex lesions lead to memory deficits that multiply dissociate. Janowsky and his colleagues have, however, examined the same group of frontal patients on several memory tasks. As a group, these patients were impaired on free recall tasks under some, but not all conditions, at temporal order memory for both pre-and postmorbidly encountered information, at digit span, and at metamemory as indicated by accuracy of feeling of knowing. In contrast, as a group, the patients performed normally on recognition tests and showed normal release from proactive interference, and no reference was made to their confabulating (see Shimamura et al., 1991). These areas of preservation may merely reflect the need for a greater degree of impairment of the same function before they also show impairment, as indicated above. The observation of Delbecq-Derouesne et al. (1990) of relatively selective recognition impairment in a patient with a frontal cortex lesion does mean, however, if replicable, that there are at least two kinds of memory impairment produced by frontal cortex lesions, one of which is perhaps caused by a problem with evaluation of retrieved information and the other of which is caused by a difficulty in setting up appropriate encoding and retrieval operations.

The most convincing demonstration of the functional heterogeneity of the frontal cortex with respect to memory suggests that different parts of the lateral frontal lobes are implicated in self-ordered pointing and conditional learning. Self-ordered pointing has already been discussed and presumably depends on the ability to monitor and remember one's own responses. Frontal cortex lesions have also been shown to impair conditional learning in which patients have to learn to select an appropriate response for each of a set of distinctive stimuli (Petrides, 1985). As Petrides (1989) discusses, these two kinds of impairment can be dissociated from each other to a certain extent in humans. Work with monkeys shows this more dramatically and also indicates that lesions of somewhat different lateral frontal regions produce the two kinds of deficit. It seems likely that other dissociations of memory deficits will be found in humans when they are looked for more systematically because the frontal lobes comprise many different cyto-architectonic regions each with their own cortical and subcortical connections.

However many frontal memory deficits there are, it seems likely that many (and possibly all) of them are secondary to disturbances in the strategic organisation of cognitive processes. This is only beginning to be explored systematically. If it is true, then the severity of the memory impairment should correlate with the severity of the responsible executive deficit. Also, in many cases, it should be possible to reverse the deficit by externally providing the guidance that the damaged supervisory system can no longer itself give. It may be that some frontal memory deficits are not secondary to executive failures. This might be the case if ventromedial frontal cortex lesions really do cause an

amnesia, and it might apply to the deficit in hand position recognition memory reported by Kesner et al. (1995).

The issue of how selectively the memory deficits are related to frontal lesions has also not been adequately addressed. With respect to temporal order memory, for example, it is certainly the case that amnesics, regardless of the site of their underlying lesion, show impaired performance (Mayes, unpublished observations) and lesions in other non-frontal brain regions that are not associated with amnesia may also possibly disrupt this form of memory. If lesions in non-frontal brain regions also disrupt the kinds of memory that are also disrupted by some frontal cortex lesions, this does not mean that they are affecting the same functions. But it does mean that the memory deficit by itself cannot be taken as an exclusive marker of a frontal cortex lesion. Performance on other tests that identify the specific frontal function affected would need to be assessed. The next section describes some preliminary work that explores the extent to which some memory deficits are selectively frontal.

COMPARISON OF THE EFFECTS OF LESIONS ON MEMORY AND COGNITIVE FUNCTIONS

Our study compared the performance on a battery of cognitive and memory tests of 10 patients with circumscribed frontal cortex damage, 10 patients with damage to the posterior cortex, and 10 healthy controls. These three groups were matched with respect to age, gender, intelligence, and a measure of affect arousal, which indicated that the patient groups were normally alert. Matching on intelligence is particularly important because there is evidence that performance on executive tests such as verbal fluency correlates with intelligence (for example, see Miller, 1984). In all cases the locations of lesions were established on the basis of CT or MRI scans.

The participants were given two executive tests: the Nelson (1976) form of the WCST and several word fluency tests that required them to produce as many items as they could in one minute that were examples of: countries, words beginning with the letter “b”, and alternately a man’s first name and the name of a vegetable. The memory battery included the following: (1) A test of memory for temporal order in which 10 abstract words were presented and after a one-minute filled delay participants were asked to reproduce the original order. In a parallel test at the same delay they were given a five-choice forced-choice recognition test for similar abstract words. (2) A different kind of test of temporal order memory was given in which participants were presented two lists of 12 face photographs two minutes apart. They were then given a Yes/No recognition test with the 24 target faces and 24 foil faces. If a face was correctly recognised, then subjects were asked to indicate from which list it came. Identification of the list in which an item appeared depends on temporal order memory because list identity is temporally defined. (3) A spatial memory task

was given in which 16 miniature models of everyday objects were presented in designated squares of a 49-square grid. At study, participants were required to guess the price of each item and then, after a five-minute delay, they were given the objects and asked to place them back in their original positions. (4) The participants were given a test of prospective memory in which they were given three kinds of instructions such as "When I stand up, please put the pencil in the mug." Fifteen minutes, 30 minutes and 45 minutes later, the experimenter satisfied the trigger conditions of each instruction. At the end of the session, the participants were asked to recall what they had been instructed to do. (5) Story recall was tested after a five-minute delay and a test of both recall and recognition of a categorised 16-item word list was also given. The story recall test, like the other recall tests used, gave the chance to take a measure of confabulation in the form of the number of items reproduced by participants that were not part of the original story, i.e. the number of commission errors. (6) A test of recall and recognition of geometrical patterns was given. In the recall condition, they copied the patterns from a dot template at study and, 30 minutes later, at test, they were asked to reproduce the patterns from memory using the dot templates as triggers. In an equivalent recognition test, they were given a five-choice forced-choice recognition test at the 30-minute delay.

With respect to the WCST, the patient groups only differed from the control group with respect to number of perseverative errors. The frontal group did not, however, do significantly worse than the posterior group although they did show a slight tendency to do so. With the fluency tasks there was a general tendency for both lesioned groups to perform worse than the control group but, if anything, the frontal group scored marginally although insignificantly better than the posterior group. Group differences or their absence cannot be explained in terms of factors such as intelligence since these were matched across groups. These results, therefore, increase the suspicion that executive test performance may not be selectively and distinctively affected by frontal cortex lesions.

Both lesioned groups performed normally on the recognition test for the abstract words. However, it was also the case that neither group was impaired at temporal order memory for the abstract words, which also does not fit with previous findings of a frontal deficit in this kind of task. A possible explanation is that the normal group performed worse on the temporal ordering task than they have done previously when given explicit instructions during the study phase to remember the order of the items. If such instructions had been given, then the frontal patients might have been significantly impaired, but we have no reason to suppose that their deficit would have been any worse than that of the posterior group.

The results with the list discrimination temporal order memory test were more interesting. On the list discrimination phase of this test, both lesioned groups were impaired to the same degree at identifying from which list the photographs

came. Thus, although the frontal group was impaired at this form of temporal order memory, the impairment it showed was not selective to frontal lobe lesions. Consistent with these results are those of Kopelman (1989), who found that performance on a between-list test of temporal order memory, similar to the one we used, did not correlate with a neuroradiological measure of frontal atrophy and only very weakly with performance on cognitive tests sensitive to frontal cortex lesions. It is, therefore, possible that frontal cortex lesions only cause a mild deficit in between-lists temporal order memory. Even so, this deficit seems to be no more severe than the deficit caused by posterior cortex lesions. It should also be said that our within-list test of temporal order memory provides no evidence that frontal cortex lesions lead to worse performance than posterior cortex lesions. Against this conclusion that both within and between list temporal order memory is equally disrupted by frontal and posterior cortex lesions, it might be argued that our patients did not have damage to the frontal cortex region which is critical for temporal order memory. However, nearly all our patients with frontal cortex lesions had sustained damage to the dorsolateral prefrontal cortex and Milner et al. (1985) have argued that it is precisely damage to this area, rather than the orbitofrontal cortex, which is primarily responsible for temporal order memory deficits.

To our surprise, the group with frontal cortex lesions performed worse than their normal controls on the spatial memory test, which is not consistent with the findings of Smith and Milner (1984). For this reason, a replication is desirable, although the different result may reflect the fact that the aetiologies of our patients differed from those of Smith and Milner's. There is certainly evidence that frontal cortex lesions produce visuospatial working memory deficits in humans (Owen et al., 1995), and recordings from single neurones within the dorsolateral frontal cortex in monkeys show that some neurones in this region increase their firing rates during the delay period (see Petrides, 1989). It remains to be seen whether the immediate spatial memory deficit caused by frontal cortex lesions is accompanied by a spatial long-term memory deficit. If it is, it needs to be determined whether the deficit is specific to tests of recall or whether it also applies to recognition. As frontal patients are known to be impaired at many tests of recall, the degree of specificity of a selective recall deficit is likely to be rather low. Long-term spatial memory deficits are likely to accompany short-term ones unless the information stored for the two kinds of memory is subtly different. Whether or not both kinds of spatial memory deficit are caused by frontal cortex lesions, it should be noted that in our study, although the frontal patients were impaired at spatial memory, the posterior patients also showed a tendency to be impaired and their performance did not differ significantly from that of the frontal group.

With prospective memory, our patient groups completed a smaller number of the instructions that they also subsequently remembered than did their controls.

They did not differ from the controls with respect to the number of instructions that were neither followed nor remembered. Although this replicates what has been found with some frontal patients previously (Burgess, personal communication), it was striking that the frontal patients did not differ from the posteriorly lesioned patients, who were equivalently impaired. In general, therefore, although the frontal patients were impaired on two tests of executive function and on memory tests that tapped temporal order, spatial, and prospective memory, these deficits were not selective because the patients with posterior cortex lesions were equally impaired.

A similar pattern of results emerged for the measure of recall success because both patient groups were impaired to the same degree at story recall. Thus, although frontal lesions may have disrupted this form of recall for different reasons from posterior lesions, the deficit was not selective. There were, however, also group differences with respect to the number of errors of commission or confabulations made. This time only the frontal group made significantly more errors than the control group and they showed a trend that approached significance also to make more commission errors than the posterior group. Recall of the word lists produced a somewhat similar pattern of results with the lesioned groups recalling less and the frontal group tending to make more commission errors not only than the control group, but also than the posterior group. Interestingly, the two lesioned groups were also equivalently and significantly impaired on the word recognition test.

The above result with recognition is contrary to what has usually been reported of frontal groups (for example, see Kesner et al., 1995) and contrary to what we found with our other recognition tests including recognition of the dot patterns, at which the two lesioned groups performed as well as the control group. Both patient groups were, however, impaired at correctly recalling the patterns although, once again, they showed the same degree of impairment. The frontal group did show a strong trend towards making more recall false alarms than either of the other groups and the posterior group did not differ from the normal control group.

There is, therefore, a fairly clear sign that the frontal group alone tends to make recall commission errors and that these errors are not produced at pathologically high levels by patients with posterior cortical lesions. It cannot easily be argued that this "confabulatory" tendency results from poor memory because otherwise it should have been found to a similar extent in the posterior group whose free recall was equivalently impaired. The tendency to make false alarm responses was not only found in the frontal patients with free recall, but also with recognition. With the recognition component of the between-lists temporal order memory test, none of the groups differed with respect to their hit rate on the recognition test for the studied faces, but the frontal patients made

significantly more false alarms than either of the other groups, the mean scores of which were closely similar to each other.

The implications of the study are clear. Lesions of the frontal cortex disrupt a range of executive functions and disturb those kinds of memory that probably depend on the effective use of the disrupted executive functions, but these deficits may also be produced by lesions of other parts of the brain and, in particular, of several regions of the posterior cortex. Interestingly, however, frontal lesions alone in our study increased the tendency to make false alarm responses in tests of both recognition and free recall. If no other non-frontal brain lesion causes this tendency, then a pathologically strong tendency to make false alarms could be a reliable marker of frontal cortex lesions. The impairment is probably an effect of dorsolateral frontal lesions, because most of our patients had such lesions, but it is unclear whether other memory deficits also display the same degree of frontal selectivity. It is also unclear exactly what underlies the pathological tendency to make false alarm errors although it could relate to the kind of sensitivity to interference considered by Shimamura and associates (1995). In other words, frontal patients may make too many false alarm errors because they cannot easily choose between correct and incorrect items that also come to mind during memory retrieval.

CONCLUSION

This chapter is a plea that more attention should be paid to determining how selective are the effects of frontal cortex lesions on executive functions and related memory task performance. As both are tapped by complex tasks, performance on which depends both on basic cognitive processes and the ability to control those processes online through the use of strategic planning, breakdown may occur for many reasons. Lesions of the prefrontal association cortex seem to disrupt various aspects of planning ability either because that region mediates planning processes directly or because it mediates working memory, disruptions of which disturb planning ability. This would explain the effects of frontal lesions on executive functions and related memory functions. Non-frontal lesions may have similar effects either because they lie within a neural circuit with the frontal cortex that mediates planning or working memory ability, or because they mediate the basic cognitive functions vital for the complex task in question. This can only be determined if several tests are given. In other words, a specifically frontal pattern of breakdown may typically only be identifiable through the use of several tests which reveal a specific pattern of breakdown. The only possible exception to this guiding principle that we have found so far is the frontal tendency to make abnormal numbers of false alarms. Further work will be needed, however, to determine whether any other non-frontal brain lesions have the same effect and, if so, whether they have it for the same or different reasons. Currently, the best way of knowing whether there has

been frontal lobe damage as opposed to temporal lobe or some other kind of brain damage to give patients MRI or CT scans.

REFERENCES

Aggleton, J.P. & Saunders, R.C. (1997). The relationships between temporal lobe and diencephalic structures implicated in anterograde amnesia. *Memory, 5,* 49–71.

Anderson, S.W., Damasio, H., Jones, R.D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13,* 909–922.

Arnett, P.A., Rao, S.M., Bernardin, M.S., Grafman, J., Yetkin, F.Z. & Lobeck, L. (1994). Relationship between frontal lobe lesions and Wisconsin Card Sorting Test performance in patients with multiple sclerosis. *Neurology, 44,* 420–425.

Benton, A.L. (1973). The measurement of aphasic disorders. In A.Caceres Velasquez (Ed.), *Aspectos patologicos del lenguaje.* Lima: Centro Neuropsicologico.

Berman, K.F., Ostrem, J.L., Randolph, C, Gold, J., Goldberg, T.E., Coppola, R., Carson, R.E., Herscovitch, P., & Weinberger, D.R. (1995). Physiological activation of a cortical network during performance of the Wisconsin Card Sorting Test: A positron emission tomography study. *Neuropsychologia, 33,* 1027–1046.

Butters, M.A., Kaszniak, A.W., Glisky, E.L., Eslinger, P.J., & Schacter, D.L. (1994). Recency discrimination deficits in frontal lobe patients. *Neuropsychology, 8,* 343–353.

Daum, I., Schugens, M.M., Spieker, S., Poser, U., Schoenle, P.W., & Birbaumer, N. (1995). Memory and skill acquisition in Parkinson's disease and frontal lobe dysfunction. *Cortex, 31,* 413–432.

Delbecq-Derouesne, J., Beauvois, M.F., & Shallice, T. (1990). Preserved recall versus impaired recognition. *Brain, 113,* 1045–1074.

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

Fuster, J.M. (1989). *The prefrontal cortex* (2nd ed.). New York: Raven Press.

Gershberg, F.B. & Shimamura, A.P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychology, 13,* 1305–1333.

Goldman-Rakic, P.S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Review of Neuroscience, 11,* 137–156.

Hanley, J.R., Davies, A.D.M., Downes, J.J., & Mayes, A.R. (1994). Impaired recall of verbal material following rupture and repair of an anterior artery aneurysm. *Cognitive Neuropsychology, 11,* 543–578.

Hirst, W. & Volpe, B.T. (1988). Memory strategies with brain damage. *Brain and Cognition, 8,* 379–408.

Incisa Della Rocchetta, A. (1986). Classification and recall of pictures after unilateral frontal or temporal lobectomy. *Cortex, 22,* 189–211.

Incisa Della Rocchetta, A. & Milner, B. (1993). Strategic search and retrieval inhibition: The role of the frontal lobes. *Neuropsychologia, 31,* 503–524.

Janowsky, J.S., Shimamura, A.P., Kritchevsky, M., & Squire, L.R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103,* 548–560.

Jetter, W., Poser, U., Freeman, R.B.J., & Markowitsch, H.J. (1986). A verbal long term memory deficit in frontal lobe damaged patients. *Cortex, 22,* 229–242.

Kesner, R.P., Hopkins, R.O., & Chiba, A.A. (1992). Learning and memory in humans with an emphasis on the role of the hippocampus. In L.Squire & N.Butters (Eds.), *Neuropsychology of Memory* (2nd ed.). New York: Guilford Press.

Kesner, R.P., Hopkins, R.O., & Fineman, B. (1995). Item and order dissociation in humans with prefrontal cortex damage. *Neuropsychologia, 32,* 881–891.

Knight, R.T., Scabini, D. & Woods, D.I. (1989). Prefrontal gating of auditory transmission in humans. *Brain Research, 504,* 338–342.

Kolb, B. & Whishaw, I.Q. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: Freeman and Company.

Kopelman, M.D. (1989). Remote and autobiographical memory, temporal context memory, and frontal atrophy in Korsakoff and Alzheimer patients. *Neuropsychologia, 27,* 437–460.

Luria, A.R. (1972). *The working brain.* Harmondsworth: Penguin.

Luria, A.R. (1976). *The neuropsychology of memory.* New York: Wiley.

Mangels, J.A., Gershberg, F.B., Shimamura, A.P., & Knight, R.T. (1996). Impaired retrieval from remote memory in patients with frontal lobe damage *10,* 32–41.

Mayes, A.R. (1988). *Human organic memory disorders.* Cambridge: Cambridge University Press.

Mayes, A.R., Downes, J.J., Shoqeirat, M., Hall, C., & Sagar, H.J. (1993). Encoding ability is preserved in amnesia: Evidence from a direct test of encoding. *Neuropsychologia, 31,* 745– 759.

McAndrews, M.P. & Milner, B. (1991). The frontal cortex and memory for temporal order. *Neuropsychologia, 29,* 849–859.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathologies. *British Journal of Clinical Psychology, 23,* 53–57.

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

Milner, B. (1964). Some effects of frontal lobotomy on man. In J.M.Warren & G.Akert (Eds.), *The frontal granular cortex and behavior.* New York: McGraw-Hill.

Milner, B., Petrides, M., & Smith, M. (1985). Frontal lobes and the temporal organization of memory. *Human Neurobiology, 4,* 137–142.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London B, 298,* 85–95.

Montaldi, D., Mayes, A.R., Barnes, Patterson, Wilson, L., Hadley, D., & Wyper, D. (1995). Findings with a memory challenge task using SPECT. First Conference on functional mapping of the human brain. June 27–30, Paris.

Moscovitch, M. (1982). Multiple dissociations of function in amnesia. In L.S.Cermak (Ed.), *Human memory and amnesia.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313–324.

Owen, A.M., Sahakian, B.J., Semple, J., Polkey, C.E., & Robbins, T. (1995). Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia, 33,* 1–24.

Passingham, R. (1993). *The frontal lobes and voluntary action.* Oxford: Oxford University Press.

Paulescu, E., Frith, C.D., & Frackowiak, R.S. (1993). The neural correlates of working memory. *Nature, 362,* 342–345.

Petrides, M. (1985). Deficits on conditional associative-learning tasks after frontal and temporal-lobe lesions in man. *Neuropsychologia, 23,* 601–614.

Petrides, M. (1989). Frontal lobes and memory. In L.Squire& G.Gainotti (Eds), *Handbook of neuropsychology, Volume 3* (pp. 75–90). Amsterdam: Elsevier.

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

Pigott, S. & Milner, B. (1995). Capacity of visual short-term memory after unilateral frontal or anterior temporal-lobe resection. *Neuropsychologia, 32,* 969–981.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298,* 199–209.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shallice, T. & Burgess, P. (1996). The domain of supervisory processes and temporal organisation of behaviour. *Philosophical Transactions of the Royal Society of London B, 351,* 1405– 1412.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 14,* 294–303.

Shallice, T., Fletcher, P., Frith, C.D., Grasby, P., Frackowiak, R.S.J., & Dolan, R.J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature, 368,* 633–635.

Shimamura, A.P. (1994). Frontal lobes and memory. In M.S.Gazzaniga (Ed.), *The cognitive neurosciences.* Cambridge, MA: MIT Press.

Shimamura, A.P., Janowsky, J.S., & Squire, L.R. (1990). Memory for temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia, 28,* 803– 813.

Shimamura, A.P., Janowsky, J.S., & Squire, L.R. (1991). What is the role of frontal lobe damage in memory disorders? In H.D.Levin, H.M.Eisenberg, & A.L.Benton (Eds.), *Frontal lobe functioning and dysfunction.* New York: Oxford University Press.

Shimamura, A.P., Jurica, P.J., Mangels, J.A., & Gershberg, F.B. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience, 7,* 144–152.

Smith, M. & Milner, B. (1984). Differential effects of frontal-lobe lesions on cognitive estimation and spatial memory. *Neuropsychologia, 22,* 697–705.

Stuss, D.T., Alexander, M.P., Palumbo, C.L., Buckle, L., Sayer, L., & Pogue, J. (1994). Organizational strategies of patients with unilateral or bilateral frontal lobe injury in word list learning tasks. *Neuropsychology, 8,* 355–373.

Stuss, D.T. & Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95,* 3–28.

Tulving, E., Kapur, S., Craik, F.I.M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Science, USA, 91,* 2016–2020.

Warrington, E.K. (1984). *The Recognition Memory Test.* Windsor: NFER-Nelson.

CHAPTER EIGHT
Normal Age-related Memory Loss and its Relation to Frontal Lobe Dysfunction

Alan J.Parkin *Laboratory of Experimental Psychology, University of Sussex, Brighton, U.K.*

There is now abundant evidence from a variety of sources, and forms of measurement, indicating that the prefrontal cortex undergoes a marked deterioration as a function of normal ageing. This is true not only in humans but in other primates and even rodents (Fuster, 1989). It is extremely unlikely that these biological facts do not have significant implications for understanding cognitive decline in the normal elderly population.

Neuropsychological studies now confirm that ageing produces a marked decline in performance on tests considered sensitive to various aspects of frontal lobe function. Mittenberg, Seidenberg, O'Leary, and DiGiulio (1989) gave a neuropsychological test battery to people of varying ages (20 to 75 years) and found that the strongest correlations were between age and measures of frontal performance. Thus the authors considered that frontal lobe dysfunction most accurately characterised the pattern of cognitive decline in the normal elderly.

This finding is supported by other studies that have examined the influence of age on specific frontal lobe tests. The Wisconsin Card-Sorting Test (WCST; Grant & Berg, 1948) and its variants (e.g. Nelson, 1976) has been used quite extensively with elderly groups. With WCST it is generally thought that perseverative errors are the main index of frontal dysfunction and a number of studies have indicated significant age-based increase in the occurrence of these errors (Braun & Lalonde, 1990; Daigneault, Braun, & Whitaker, 1992; Haaland, Vranes, Goodwin, & Garry, 1987). However, other studies, while replicating this result, also indicate age-related declines in total errors and number of categories achieved (Parkin & Walter, 1992; Spencer & Raz, 1994). A further study has shown a significant decline in these latter measures but not perseverative responding (Boone, Miller, Lesser, Hill, & D'Elia, 1990).

The Stroop test is also considered to be a frontal test (Bench, Frith, Grasby, Friston, Paulescu, Frackowiak et al., 1993) and a number of studies have reported age-related declines on this task (Boone et al., 1990; Whelihan &

Lesher, 1985). The influence of age on fluency measures is less clear. "Letter fluency" (most commonly administered as the FAS test) does not seem particularly sensitive to age (e.g. Daigneault et al., 1992; Parkin & Walter, 1992; although see Veroff, 1980; Whelihan & Lesher, 1985). However, more demanding fluency tasks such as category fluency and alternating fluency do reliably show pronounced age effects (Parkin, Walter, & Hunkin, 1995). In addition Parkin and Lawrence (1994) have shown very dramatic age effects using the Alternate Uses Test of flexible thinking.

Recently Loveday (1996) has intensively investigated frontal abilities in the elderly. Her data show significant age-related declines on various measures of fluency. She also found a significant age effect on WCST. In addition she explored the effects of age on two components of the CANTAB (Sahakian & Owen, 1992) sensitive to frontal function, ID/ED shift (a similar type of task to WCST) and the Tower of London. Highly significant age-related declines were also found here.

The above review is not meant to be exhaustive but merely an indication of the substantial evidence for age-related neurobehavioural decline in the elderly centred on measures of frontal lobe function. In the remaining parts of this chapter I will address two more recent issues. First, does this pattern of frontal decline in the elderly represent loss of a single "frontal" ability which is common to a range of frontal tasks, or does it represent loss of a number of different abilities? Second, does this pattern of frontal decline have implications for understanding the most prominent cognitive change with age—loss of memory?

ONE OR MORE FRONTAL FUNCTIONS?

Given the extent of the prefrontal cortex it is likely, on intuitive grounds, that it subsumes more than one function. However, it is notable that some theorists have argued that the tendency for various tests of frontal function to intercorrelate is evidence for a single frontal function. Most recently it has been proposed that fluid intelligence, as characterised by Spearman's *g,* may in a large part be a reflection of frontal lobe function (Duncan, Burgess, & Emslie, 1995).

Eslinger and Grattan (1993) drew a distinction between two forms of frontal lobe function: *reactive* and *spontaneous* flexibility. Reactive flexibility was defined as a readiness to shift cognition and behaviour freely in response to changing demands of an external situation. It is exemplified by the WCST in which the participants continually have to alter their response pattern in response to instruction. Spontaneous flexibility describes the ready flow of ideas and answers in response to a question. This is typically measured by fluency tasks such as the Alternate Uses Test and word fluency, where participants are required to think divergently. Applying these tests to a group of neurological

patients, they found that the two types of test were associated with different underlying lesions.

In two experimental studies of ageing we have examined the validity of Eslinger and Grattan's dichotomy within an ageing population. In the first of these (Parkin & Lawrence, 1994), we found strong correlations between two measures of spontaneous fluency (FAS and Alternate Uses Test) but no significant relationship between these and errors on WCST. In the second study (Parkin et al., 1995) no correlation was obtained between a category fluency test and WCST errors but a correlation was found between WCST errors and an alternating fluency task. In this latter task two target categories are specified and the participant must give exemplars alternately from each one. The alternating fluency task also correlated with category fluency and it was thus assumed that alternating fluency had both a spontaneous fluency element (retrieving exemplars) and a reactive element (switching categories).

Loveday (1996) also explored the relation between her various measures of frontal function. Tests of fluency (FAS, category fluency, alternating fluency) all correlated significantly with one another. However, there was no significant correlation between WCST and either FAS or category fluency. However, in line with Parkin et al. (1995) she did find a correlation between WCST and alternating fluency, thus supporting the view that this task has both reactive and spontaneous flexibility components.

FRONTAL FUNCTION AND NORMAL AGE-RELATED MEMORY LOSS

The above studies have established that there is a significant decline in frontal lobe function with age and that this decline appears to represent the loss of at least two separate dimensions of frontal performance. With this in mind we will now move to our second new issue: is the age-related decline in frontal function associated with the loss of memory typical in later life?

The first investigation of the relation between memory function and frontal lobe dysfunction in the elderly involved a study of source amnesia. In a source memory experiment participants are presented with a series of facts emanating from two or more sources (e.g. different speakers) and are subsequently asked to retrieve both the facts and their source. Studies such as that of Janowsky, Shimamura, and Squire (1989) indicate that frontal lesions disproportionately impair memory for source relative to fact memory. Craik, Morris, Morris, and Loewen (1990) tested older adults on their ability to remember where they had been given factual information using the source-forgetting paradigm designed by Schacter, Harbluk, and McLachlan (1984). In the study participants were presented with both true and false facts about well-known personalities (e.g. "Bob Hope's father was a fireman;" "Jane Fonda always eats oatmeal for

breakfast"). One week later memory for the facts was tested by asking questions pertaining to both the made-up facts and true facts. When they answered a question they were also required to state the source of their knowledge. The measure of source amnesia was based on the number of newly acquired (i.e. false) facts presented during the learning phase which participants claimed to have learned from some source other than the experiment. The extent of source amnesia was then correlated with participants' scores on two frontal tests, WCST and FAS. This showed a strong relationship between the extent of source amnesia and impairment on both FAS and WCST.

Glisky, Polster, and Routhieaux (1995) carried out an interesting source amnesia study in which a group of elderly participants were first given a range of neuropsychological tests. On *a priori* grounds, a group of five tests—WCST, FAS, the mental arithmetic subtest of WAIS, and the mental control and backward digit span subtests from WMS—R—were assumed to load on frontal function. The remaining tests—logical memory, verbal paired associates, and visual paired associates (from WMS—R) plus delayed cued recall from the California Verbal Learning Test—were all considered to be tests reflecting medial temporal lobe function. A factor analysis was then carried out on these test scores and this revealed two factors which confirmed *a priori* considerations.[1]

Glisky et al. then conducted an experiment in which participants listened to a series of sentences spoken in either a male or a female voice and rated how likely they would be to hear each sentence on the radio. Following this they were given two types of memory test. In the item memory test they decided which of two sentences they had heard before, whereas in the source test they heard each target sentence spoken by both the male and female speaker and had to decide in which voice the sentence had originally been presented.

The subsequent data analysis was somewhat unusual in that participants were then divided, on the basis of their scores relative to the mean, into "high" and "low" frontal, indicating relatively good and relatively poor frontal function respectively, and in the same manner into "high" and "low" medial temporal lobe function. Source and item memory in these pairs of contrasting groups was then examined. In the frontal contrast no difference was found between high and low groups on item memory but the low group were significantly worse (at chance in fact) on source memory. In the other analysis those participants

[1] The identification of a single frontal factor by Glisky et al. would appear to contradict my earlier conclusions about more than one frontal factor in neuropsychological test profiles of the elderly. However, Glisky et al. only included two tests, WCST and FAS, which directly reflect the contrast between reactive and spontaneous fluency. More importantly, FAS is a test which we have found least sensitive to the effects of ageing perhaps because it is more sensitive to IQ. Since WCST is also IQ-sensitive the relation between them in Glisky et al.'s data could simply reflect an association based on IQ.

deemed to have more efficient medial temporal lobe function scored better than their low counterparts on item memory but no group difference in source memory was found.

Glisky et al.'s study thus confirms other studies we have considered in which source memory appears related to frontal lobe function. In addition, it shows that the other component of the task, item memory, loads on a factor which they assume relates to some more primary memory ability mediated by the medial temporal lobe. However, this consistent picture has been challenged in a study by Spencer and Raz (1994) in which young and elderly participants were evaluated using a variation on the paradigm introduced by Schacter et al. (1984). Here they were read fictitious statements about famous people and examined these statements simultaneously on either pink or blue cards and in one of two different rooms. Subsequent testing evaluated fact recall by asking a pertinent question, and source memory was measured in terms of participants' ability to attribute correctly knowledge of that fact to its prior presentation in the experiment. Memory for contextual detail was assessed by asking about further details of presentation, e.g. room used for presentation.

Spencer and Raz (1994) administered a range of frontal tasks and were thus able to examine how performance on these three types of memory related to frontal ability—the frontal tests used were a computerised version of WCST, the Stroop task, and a task measuring memory for temporal order (the latter involved participants attempting to place, in correct temporal order, 18 activities that they had taken part in as a consequence of being experimental volunteers). Large age effects on source memory were found but the extent of this did not correlate with measures of frontal function. However, there was evidence of a relation between impaired frontal function and memory for contextual detail.

The Spencer and Raz (1994) study appears to contradict the association of frontal dysfunction with impaired source memory in the elderly indicated by both the Glisky and Craik studies. However, it is notable that source memory ability in the elderly appeared to be very good (at around 80%) in the elderly sample, whereas performance was considerably lower in the other studies especially that of Glisky where participants averaged around only 60% accuracy. It may thus be the case that the measure of source memory used by Spencer and Raz was insensitive to potential impairments in some individuals. In connection with this it is notable that memory for contextual detail was found to be more difficult and this did show a relationship with frontal ability. As we shall see later there are reasonable grounds for assuming that memory for this information does not differ qualitatively from that defined in a stricter sense as "source."

Other paradigms have also been used to explore the relationship between memory and frontal function in the normal elderly. Parkin and Walter (1991) compared old and young participants on the Brown-Peterson test of short-term forgetting. Three items were presented on each trial and forgetting assessed over

distraction intervals ranging from 18 to 60sec. As in previous studies, they found poorer performance in the older group but no interaction between the extent of this age effect and retention interval. Correlations were also carried out between memory performance and two tests of frontal function, FAS and WCST, and, in both cases, memory performance was correlated with frontal ability.

Parkin and Walter (1992) compared the accuracy of yes—no recognition memory in three age groups and found a small but significant age effect. In addition, using a technique which is now termed the recognition and conscious awareness (RCA) paradigm (Gardiner, 1988), participants also had to classify each recognised target on a subjective basis. If recognition involved specific recollection of a target's prior occurrence (e.g. an image it evoked) it was classified as a "remember" (R) response, whereas recognition based simply on familiarity was classified as a "know" (K) response. The effects of age on the distribution of R and K responses were very clear-cut: with increasing age participants made substantially fewer R responses and correspondingly higher numbers of K responses. This held even in subgroup comparisons where age had no effect on recognition accuracy. This changed pattern of responding was not due to lowered confidence in the elderly. Instead this pattern appeared to be related to WCST performance, with those individuals showing greatest reliance on K responding tending to have higher error rates.

Parkin et al. (1995) compared young and elderly groups on the list discrimination task (LDT). Here participants studied two sets of target sentences separated by a three-minute interval. Retention was tested by requiring them to identify target sentences they had seen and, contingent upon this, to state in which of the two lists the target had occurred. No age differences in recognition were found but the elderly performed much worse on memory for temporal context. Furthermore, the extent of this temporal context deficit correlated with the extent of impaired performance on measures of spontaneous fluency but not reactive fluency.

Finally, a study we reviewed earlier (Parkin & Lawrence, 1994) gave elderly participants two memory tasks considered to have a component sensitive to the role of the frontal cortex in memory function. One task, release from proactive interference, is a variant of the Brown—Peterson paradigm in which participants study successive triads of items drawn from the same taxonomic category and recall them after a short retention interval. As the trials progress, an interference effect builds up, with poorer recall overall and intrusions from earlier trials. However, this proactive interference effect can be released if a trial involving items from a different taxonomic category is introduced. Several studies have suggested that this release effect depends, at least in part, on the integrity of certain frontal lobe functions (Squire, 1982).

The second task used by Parkin and Lawrence was the matched recall and recognition task first used by Calev (1984). Usually recognition is less impaired

than recall in any neurologically compromised population but, very often, this discrepancy is an artefact of the recognition task being easier. The Calev test avoids this because its construction ensures that, within a young sample, performance on the recall and recognition components is equal. Thus any disruption of recall relative to recognition in an elderly sample could not be attributed to test sensitivity artefacts. Motivation for its inclusion in a study of the elderly was an experiment by Hanley, Davies, Downes, and Mayes (1994) in which it was shown that a lesion to the frontal cortex selectively disrupted recall relative to recognition, a finding that has been confirmed in a second study of a frontally damaged patient (Parkin, Yeomans, & Bindschaedler, 1994).

The results of the study showed that the extent of release from PI was correlated with measures of spontaneous fluency but that greater discrepancies between recall and recognition were associated with increased error rates on the WCST. Parkin and Lawrence thus concluded that there was a relationship between frontal function and age-related memory loss but that this relationship was not a unitary one. Since then, however, both Boubert (1995) and Java and Parkin (submitted) have confirmed that normal ageing does not result in a disproportionate decline in recall relative to recognition. Moreover, Java and Parkin (submitted) have failed to find a relationship between WCST and recall—recognition discrepancy. They attribute this to population differences in that there is *prima facie* evidence that the Parkin and Lawrence elderly were more neurologically compromised.

Recently Small, Okonek, Mandelkern, La Rue, Chang, Khonsary et al. (1994) have described the first study to examine the relationship between neuroimaging indices of frontal dysfunction and normal age-related memory loss. In this study brain metabolism was related to a range of neuropsychological measures including several standardised memory tests. Overall the results showed very few correlations but it was of considerable interest that evidence of decreased frontal metabolism was highly correlated with increased usage of mnemonic aids as measured by a self-assessment questionnaire.

THEORETICAL CONSIDERATIONS

The above account enables us to conclude three basic facts:

Normal ageing results in a pattern of cognitive decline that particularly reflects loss of functions associated with the frontal cortex.

This decline in function does not involve a unitary frontal factor and is more adequately characterised in terms of a dichotomy between tests of spontaneous and reactive fluency.

Age-related memory loss appears to be related to the decline in frontal lobe function as indicated by the association between certain measures of memory function and aspects of frontal functioning.

There would seem little basis for disputing the first conclusion, so our concern will be with discussing the second and third. Taking the non-unitary concept of frontal functioning first, it might be argued that the dissociation between frontal tasks is not due to the lack of an essential frontal function but due to dissociations between other, more peripheral, task elements. Thus, as Duncan et al. (1995) have suggested, it might well be that the frontal cortex is characterised in terms of *g* and that the absence of significant correlations between tests of spontaneous and reactive fluency arises from dissociation between "non-*g*" components of the tasks. This view is difficult to evaluate on one major ground. If one argues that all tasks share a common frontally mediated *g* component then this component must be operationally specified. At present the concept of *g* remains vague and it therefore seems unlikely that it will explain patterns of relation and non-relation between measures of frontal function.

An alternative view of the relation between frontal functions has been put forward by Goldman-Rakic (1987), who has proposed that the frontal cortex acts as a working memory system which provides cohesion between the various elements of any complex task. To account for the lack of association between frontal tasks, Goldman-Rakic argues that there are a number of working memory systems each concerned with a specific type of cognitive ability. This approach has recently been used in a connectionist account of human frontal function proposed by Kimberg and Farah (1993) in which a number of frontal tasks were modelled, each with their own working memory system. Further, when these networks were "lesioned" they produced distorted outputs similar in kind to the effects of frontal lobe damage in humans.

The attraction of the Kimberg and Farah model is that the efficiency of frontal functions is predicted in terms of the cohesion between task elements without recourse to the concept of a single central executive. However, within modern neuropsychology the most commonly accepted approach to frontal lobe function is to propose the existence of a "central executive" which has ultimate responsibility for the behaviour of an individual. Within this class of theory the most widely discussed is the supervisory activating system (SAS; Norman & Shallice, 1986). In this system the majority of actions are carried out by routines which do not require conscious intervention. However, when these routine activities will not suffice, the SAS must intervene and initiate an appropriate response. Typically the SAS would be called into action in tasks which, because of their novelty, require planning, or where a strong response tendency needs to be inhibited because it is inappropriate.

At present the data on frontal lobe test dissociation in the elderly could be accommodated by either the multiple working memory approach or in terms of some version of the SAS system in which different components of executive function are identified. My own inclination is towards an account based upon a multifacted central executive system. This is partly based on intuition but, more

importantly, I feel this theory gains greater credibility when the issue of executive control of memory function is considered.

FRONTAL LOBES AND MEMORY

As we have seen, it is now widely recognised that the frontal cortex plays an important role in memory function. However, the exact basis by which the frontal cortex achieves this is far from understood. One currently influential idea is based on an extension of the SAS model of executive control. This accounts stems from the work of Norman and Bobrow (1979) and proposes that executive processes are involved at encoding, where they are responsible for setting up retrieval routes for stored information, and at retrieval where they verify the appropriateness of retrieved information.

From a neuropsychological perspective this account of executive memory function predicts a possible dissociation between deficits arising because of deficient encoding and those that would arise from impairments of the verification process. Hanley et al. (1994) described a patient, R.O.B., who showed highly defective recall but normal recognition when tested on the Calev matched recall and recognition test (see above). They proposed that R.O.B.'s recall deficit stemmed from an impaired ability to set up retrieval strategies. However, because recognition obviated the need for an effective retrieval route, it was unaffected. To provide the double dissociation they highlighted the earlier case of R.W. (Delbecq-Derouesne, Beauvois, & Shallice, 1990), in which recognition was characterised by a normal hit rate but excessive false alarms and a pattern of recall in which high levels of intrusions occurred. Here, they argued, was evidence for a defective verification process. This type of defective verification following frontal lobe damage has recently been shown in two other frontal patients, J.B. (Parkin, Bindschaedler, Harsent, & Metzler, 1996) and B.G. (Schacter, Curran, & Galluccio, 1996). Also of note is that all these patients had a history of fantastic confabulation, a deficit that is most readily explained in terms of a defective verification process.

Neuroimaging studies have also supported a distinction between encoding and verification processes. Fletcher, Frith, Grasby, Shallice, Frackowiak, and Dolan (1995), for example, took PET measurements of regional cerebral blood flow during the encoding and retrieval of paired associates. The finding of most interest was that the encoding of word pairs was associated with greater activation of the left frontal cortex (along with the cingulate cortex) while retrieval was associated with increased activity of the right frontal region and also the precuneus (a parietal lobe structure) bilaterally. This study thus suggests that the frontal cortex contains separable brain systems involved in the encoding and retrieval of memories (see also Kapur, Craik, Tulving, Wilson, Houle, & Brown, 1994; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994). These

findings seem to square nicely with the executive account of memory function put forward by Hanley et al. and is further supported by the fact that patient B.G. (Schacter et al., 1996) is known to have a right frontal lesion.

The next step is to examine whether this theoretical framework can be used to account for age-related memory loss. It is clear that, despite superficial differences, many of the tasks on which older people fail have a crucial contextual encoding component. Thus an inability to retrieve the source of information, incidental presentation details, information about temporal order, and specific aspects of an occurrence (as required for the generation of an R response in the RCA paradigm), all represent some failure to encode context. Impaired recall relative to recognition could also be explained in terms of defective contextual encoding because the latter task is far less dependent on contextual cues. Finally, the elderly's impairment on release from proactive interference might also constitute a contextual encoding deficit in that attributes differentiating exemplars of different categories may fail to be encoded.

One can, of course, make similar arguments for an explanation of age-related memory deficits in terms of defective retrieval. My own studies of ageing have shown, for example, that differences between the old and young on tests of implicit memory appear to arise because the older ones fail to utilise explicit memory for target information as effectively as the young even when the amount of explicit memory available to the two groups is matched (Russo & Parkin, 1993). On the same basis, for example, poor source memory or memory for temporal order could reflect some failure to retrieve information effectively.

At present there has been little attempt to discover whether defective encoding or retrieval underlies the pattern of age-related memory loss we have examined in this chapter. However, a recent study (Loveday & Parkin, in prep.) compared the pattern of R and K responses generated in the standard RCA procedure, in which target words are exposed for 2sec during learning, and a condition where target exposure was increased to 4sec. The results showed that increased exposure time increased the probability of R and reduced to the probability of K responses relative to the 2sec condition. This finding suggests that defective encoding may be responsible for the lower level of recollective memory following two-second exposures.

The above discussion hopefully indicates that normal age-related memory loss can be studied from the perspective of disturbed frontal lobe function. However, investigations so far have merely shown associations between tests of memory with "frontal" components and standardised tests of frontal lobe function such as WCST and FAS. This has undoubtedly produced a somewhat confusing picture. Some studies (e.g. Glisky et al., 1995) have treated all frontal tests as representing a single frontal factor whereas our own work suggests that a distinction between spontaneous and reactive fluency needs to be maintained (Parkin & Lawrence, 1994).

While it is the case that the relation between some frontal tests and types of memory impairment can be understood, the relations obtained are not always clear. Associations between memory for contextual detail and fluency, for example, both seem amenable to an explanation in terms of an ability to organise retrieval effectively. However, other relationships are less clear. Why, for example, should increased R responding and a lowered discrepancy between recall and recognition be associated with better performance on WCST? The basis of this problem lies in the fact that the frontal tests utilised were not driven by the theory being used to explore frontal involvement in memory. Research into executive memory deficits might thus benefit by using executive tests that more closely match hypothetical components of the frontally mediated memory tasks. Recently we have attempted this by examining, in the elderly, the relationship between intrusions in free recall and the random number generation task (Spatt & Goldenberg, 1993). In the latter task participants are required to generate random numbers and the critical variable is the extent to which they break from randomness and produce consecutive sequences of numbers (e.g. 2, 4, 6, 8). Spatt and Goldenberg (1993) showed that increased incidence of this type of response was linked to frontal lobe damage and suggested that it reflected an inability of the executive process to monitor and inhibit inappropriate responses. We considered that this type of monitoring process might be similar to that specified in the verification process and so we examined the relationship between impairment of random number generation and intrusions in free recall. Our intuitions were confirmed, when across two recall experiments, we found higher incidence of intrusion errors associated with poorer performance on the random number generation task (correlations were .583 and .523).

CONCLUSION

The work described in this chapter has shown that ageing produces a neurobehavioural profile characterised by dysfunction of the frontal cortex. In addition it is argued that this dysfunction underlies much of the memory impairment observed as a normal consequence of ageing. It has been proposed that this deficit is best understood in terms of an executive memory impairment in which a primary distinction is drawn between deficits of encoding and retrieval. However, experiments have only just begun to explore this theory further. To achieve this it is argued that attempts to relate executive impairments per se to memory deficits should utilise tasks which attempt to mirror the hypothetical involvement of executive processes in actual memory mechanisms.

ACKNOWLEDGEMENT

Preparation of this chapter was partly supported by a grant from the Wellcome Trust.

REFERENCES

Bench, C.J., Frith, C.D., Grasby, P.M., Friston, K.J., Paulescu, E., Frackowiak, R.S.J., & Dolan, R.J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 31(9),* 907–922.

Boone, K.B., Miller, B.L., Lesser, I.M., Hill, E., & D'Elia, L. (1990). Performance on frontal lobe tests in healthy, older individuals. *Developmental Neuropsychology, 6,* 215–223.

Boubert, L. (1995). Personal communication.

Braun, C.M.J. & Lalonde, R. (1990). Le déclin des fonctions cognitives chez la personne âgée: Une perspective neuropsychologique. *Revue Canadienne du Vieillissement, 9,* 135–158.

Calev, A. (1984). Recall and recognition in mildly disturbed schizophrenics: The use of matched tasks. *Journal of Abnormal Psychology, 93,* 172–177.

Craik, F.I.M., Morris, L.W., Morris, R.G., & Loewen, E.R. (1990). Relations between source amnesia and frontal functioning in older adults. *Psychology of Aging,* 5, 148–151.

Daigneault, S., Braun, C.M.J., & Whitaker, H.A. (1992). Early effects of normal aging on perseverative and non-perseverative prefrontal measures. *Developmental Neuropsychology, 8,* 99–114.

Delbecq-Derouesne, J., Beauvois, M.F., & Shallice, T. (1990). Preserved recall versus impaired recognition. *Brain, 113,* 1045–1074.

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

Eslinger, P.J. & Grattan, L.M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia, 31(1),* 17–28.

Fletcher, P.C., Frith, C.D., Grasby, P.M., Shallice, T., Frackowiak, R.S.J., & Dolan, R.J. (1995). Brain systems for encoding and retrieval of auditory-verbal memory. *Brain, 118,* 401–416.

Fuster, J. (1989). *The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe* (2nd ed.). New York: Raven Press.

Gardiner, J.M. (1988). Functional aspects of recollective experience. *Memory and Cognition, 16,* 309–313.

Glisky, E.L., Polster, M.R., & Routhieaux, B.C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9,* 229–235.

Goldman-Rakic, P.S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational knowledge. In F.Plum & V.B.Mountcastle (Eds.), *Handbook of physiology* (pp. 373–417). Bethesda, MD: American Physiological Society.

Grant, D.A. & Berg, E.A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 38, 404–411.*

Haaland, K.Y., Vranes, L.F., Goodwin, J.S., & Garry, P.J. (1987). Differences in neuropsychological test performance in a healthy elderly population. *Journal of Gerontology, 42,* 345–346.

Hanley, J.R., Davies, A.D.M., Downes, J.J., & Mayes, A.R. (1994). Impaired recall of verbal material following rupture and repair of an anterior communicating artery aneurysm. *Cognitive Neuropsychology, 11,* 543–578.

Janowsky, J.S., Shimamura, A.P., & Squire, L.R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology, 17,* 3–11.

Java, R.I. & Parkin, A.J. (submitted). Why does digit-symbol substitution speed attenuate age-related memory variance? The role of fluid intelligence.

Kapur, S., Craik, F.I.M., Tulving, E., Wilson, A.A., Houle, S., & Brown, G.M. (1994). Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proceedings of the National Academy of Sciences USA, 91,* 2088–2011.

Kimberg, D.Y. & Farah, M.J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General, 122(4),* 411–428.

Loveday, C. (1996). The involvement of the frontal lobes in normal age-related memory loss. PhD, University of Westminster.

Loveday, C. & Parkin, A.J. (in prep.). The influence of study time on recollective experience in the elderly.

Mittenberg, W., Seidenberg, M., O'Leary, D.S., & DiGiulio, D.V. (1989). Changes in cerebral functioning associated with normal aging. *Journal of Clinical and Experimental Neuropsychology, 11,* 918–932.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313–324.

Norman, D.A. & Bobrow, D.G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology, 11,* 107–123.

Norman, D.A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R.J.Davidson, G.E.Schwartz, & D.E.Shapiro (Eds.), *Consciousness and self-regulation.* New York: Plenum Press.

Parkin, A.J., Bindschaedler, C., Harsent, L., & Metzler, C. (1996). Verification impairment following frontal lobe damage. *Brain and Cognition, 32,* 14–27.

Parkin, A.J. & Lawrence, A. (1994). A dissociation in the relation between memory tasks and frontal lobe tests in the normal elderly. *Neuropsychologia, 32,* 1523–1532.

Parkin, A.J. & Walter, B.M. (1991). Aging, short-term memory, and frontal dysfunction. *Psychobiology, 19,* 175–179.

Parkin, A.J. & Walter, B.M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7,* 290–298.

Parkin, A.J. Walter, B.M. & Hunkin, N.M. (1995). Normal aging, frontal lobe function, and memory for temporal and spatial information. *Neuropsychology, 9,* 304–312.

Parkin, A.J., Yeomans, J. & Bindschaedler (1994). Further characterization of the executive memory impairment following frontal lobe lesions. *Brain and Cognition, 26, 23–42.*

Russo, R. & Parkin, A.J. (1993). Age differences in implicit memory: More apparent than real. *Memory and Cognition, 21,* 73–80.

Sahakian, B.J. & Owen, A.M. (1992). Computerised assessment in neuropsychiatry using CANTAB. *Journal of the Royal Society of Medicine, 85,* 399–402.

Schacter, D.L., Curran, T., & Galluccio, L. (1996). False recognition and the right frontal lobe. *Neuropsychologia, 34,* 793–808.

Schacter, D.L., Harbluk, J.L. & McLachlan, D.R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behaviour, 25,* 593–611.

Small, G.W., Okonek, M.A., Mandelkern, M.A., La Rue, A., Chang, L., Khonsary, A., Ropchan, J.R., & Blahd, W.H. (1994). Age associated memory loss: Initial neuropsychological and cerebral metabolic findings of a longitudinal study. *International Psychogeriatrics, 6,* 23–44.

Spatt, J. & Goldenberg, G. (1993). Components of random generation by normal subjects and patients with dysexecutive syndrome. *Brain and Cognition, 23,* 231–242.

Spencer, W.D. & Raz, N. (1994). Memory for facts, source and context: Can frontal function explain age related differences? *Psychology and Aging, 9,* 149–159.

Squire, L.R. (1982). Comparisons between forms of amnesia: Some deficits are unique to Korsakoff's syndrome. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8,* 560–571.

Tulving, E., Kapur, S., Craik, F.I.M., Moscovitch, M. & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences USA, 91,* 2016–2020.

Veroff, A.E. (1980). The neuropsychology of aging: Qualitative analyses of visual reproduction. *Psychological Research, 41,* 1259–1268.

Whelihan, W.M. & Lesher, E.L. (1985). Neuropsychological changes in frontal functions with aging. *Developmental Neuropsychology, 1,* 371–380.

CHAPTER NINE
Do "Frontal Tests" Measure Executive Function? Issues of Assessment and Evidence from Fluency Tests

Louise H.Phillips *Psychology Department, Aberdeen University, Old Aberdeen, U.K.*

The executive processes which control and regulate cognition have been the subject of intense interest in recent years. In particular there has been a growing consensus that the frontal lobes of the brain are intimately involved in executive function (e.g. Duncan, 1986; Shallice, 1988; Stuss & Benson, 1986). There have also been a large number of studies into childhood development of frontal and executive function (e.g. Riccio, Hall, Morgan, Hynd, Gonzalez, & Marshall, 1994); the operation of executive processes in adulthood and normal ageing (e.g. Parkin & Walter, 1992); and the effects of degenerative diseases on executive function (e.g. Bhutani, Montaldi, Brooks, & McCulloch, 1992). Most of this research involves tasks which are interchangeably called "frontal lobe" or "executive" tests, commonly accepted within neuropsychology as measures of executive function. However, the term "frontal lobe tests" is problematic because such measures are often not sensitive to, or specific to, lesions in that particular brain area (Reitan & Wolfson, 1994). Also, the use of the term "executive tests" seems somewhat premature, given that little is known about the reasons for poor performance, and few studies have attempted to determine whether these tasks really assess executive function. Nevertheless, these "executive" tests are widely used in both clinical and healthy populations, so it is imperative that we increase our understanding of the factors influencing performance upon them.

This chapter focuses on the assessment of executive function in non-clinical populations and has three aims: (1) an outline of current knowledge about executive function and methods of assessment; (2) an investigation into whether performance on one particular "executive" test, fluency, is best explained in executive terms; and (3) some examples of the problems involved in current methods of assessing executive function.

THEORY AND MEASUREMENT OF EXECUTIVE FUNCTION

Executive processes are responsible for the control of cognition, and the regulation of behaviour and thought. As such, intact executive function is critical in everyday life in order to initiate, monitor, and terminate behaviour patterns. Executive processes are conscious, effortful and (thus far at least) poorly understood, in contrast to automatic, modular, non-executive processes, of which we have quite extensive knowledge (Fodor, 1983). Despite the importance of executive function in real-life situations it has proved rather difficult to measure in experimental or clinical settings. Some have argued that executive control processes are actually impossible to study scientifically (Fodor, 1983), whereas others have admitted that so far our understanding of executive function is thoroughly inadequate (Baddeley, 1990). Certainly most current theories of executive function are not fully specified, and diverge on such issues as the degree of fractionation of control processes.

Most theories of executive function have their roots in the neuropsychological study of the frontal lobes, and it is due to attempts during the past couple of decades to understand this previously mysterious area of the brain that interest in executive function has been so potent. Lesions of the frontal lobes result in the apparently paradoxical pattern of severely impaired problem solving in real-life situations, but intact ability to carry out many complex cognitive tests. In recent literature there has been a growing consensus that the underlying deficit in frontal lobe patients is one of executive control processes.

Executive processes are intertwined with the notion of volition: the freedom to make appropriate (or inappropriate) choices from a set of possible actions and the overall control that we have over processes of cognition. Although it is an important aspect of our subjective impression of cognition, volition has not been widely studied within cognitive psychology. Traditionally, cognitive research has tended to deal with the processes involved in attempting simple tasks, rather than grappling with the more slippery problem of how such processes are initiated and coordinated. This reluctance to study executive processes probably has multiple causes: unwillingness to confront homuncular ideas which might imply Cartesian duality; post-behaviourist aversion to the notion of higher cognitive processes; and weary recognition that the operation of executive function is likely to be extremely complex.

Most authors agree that executive function consists of a number of interconnecting control processes. When initially faced with a new task, the first stage of executive processing is likely to be the initial task analysis in order to identify the goals and subgoals that have to be achieved (Duncan, 1986; Lezak, 1983; Shallice & Burgess, 1991b). Those goals must then be weighted hierarchically to formulate a strategic plan of action (Denckla, 1994; Duncan, 1986; Lezak, 1983; Shallice & Burgess, 1991b), which must then be carried out

effectively. Task performance must be monitored throughout in order to suppress inappropriate automatic actions or thoughts (Denckla, 1994; Duncan, 1986; Roberts, Hager, & Heron, 1994). Performance is also evaluated in order to allow the modification of plans (Shallice & Burgess, 1991b), with goal weightings being revised on line if necessary. All of these aspects of executive functioning are likely to be interdependent: for example, good strategic planning may reduce the potency of interference.

Any test that is hypothesised to measure executive function should therefore demand at least some of the processes described above. Executive tests should be *novel* in order to tap goal identification and strategic planning, because any well-practised tasks could be carried out using previously formulated strategies. Such tests should also be *effortful,* in terms of goal planning and subsequent stages of prepotent response inhibition and monitoring. Executive tests may also involve *working memory* (Denckla, 1994; Roberts et al., 1994) in order to coordinate the concurrent mnemonic and online processing activities. In order to have face validity, therefore, executive tests must involve some combination of novelty, effortfulness and working memory demands.

Appropriate methods of assessing executive function are essential, because executive control processes are seen as central in recent theories of memory (Baddeley, 1986), attention (Norman & Shallice, 1986), and intelligence (Sternberg, 1985). The concept of working memory (Baddeley, 1986; Baddeley & Hitch, 1974) anchors memory to real-life tasks, such as problem solving and reasoning. The working memory model has at its literal centre a component which controls and coordinates mnemonic processes, the central executive. So far, little is known of the operation of the central executive (Baddeley, 1990), although it is thought to be important in a large number of cognitive tasks (e.g. reasoning, Gilhooly, Logie, Wetherick, & Wynn, 1993; and mental arithmetic, Logie, Gilhooly, & Wynn, 1994); and also in age differences in cognition (Baddeley, 1986; Craik, Morris, & Gick, 1990). The central executive component of working memory has been theoretically linked by Baddeley (1990) to the Norman and Shallice (1986) model of attentional processes. In this model a supervisory attentional system modulates and controls action plans. Shallice (1988) proposes that deficits in supervisory attentional control reflect the executive dysfunction found after frontal lobe damage.

Executive control processes are also seen as important in many theories of intelligence. In some ways the current search for satisfactory measures of executive function parallels the early history of intelligence testing (Reitan & Wolfson, 1994). Definitions of intelligence (still not agreed upon after a hundred years, perhaps an ominous sign for the study of executive function) often call upon the notion of executive control of cognition (Sternberg & Detterman, 1986). Duncan (1994) and Carpenter, Just and Shell (1990) argue that individual differences in intelligence depend upon executive properties, such as the ability

to manage goal hierarchies successfully. The conceptualisation of fluid intelligence as a measure of effortful processing of novel stimuli has much in common with current ideas about executive functioning.

So, executive processes are thought to be important in relation to a number of influential theories of cognition. It is therefore important that we have reliable, valid methods of assessment. Measuring executive function is also an urgent issue in relation to recent interest in: children's development of executive function (Denckla, 1994); declines in executive function with adult age (Baddeley, 1986); and deficits in executive function in patients with Alzheimer's disease (Della Sala, Logie, & Spinnler, 1992) and many other clinical conditions (e.g. schizophrenia, autism, Parkinson's disease, depression…). There are currently two methods of executive assessment which are widely used: the dual-task paradigm and "frontal lobe tests"; these will now be discussed.

The Dual-task Paradigm

In normal populations executive function is often studied experimentally by looking at interference between concurrent tasks (e.g. Baddeley, 1986; Gilhooly et al., 1993; Logie et al, 1994; Pearson, Logie, & Green, 1996). In the dual-task paradigm, a target task is performed simultaneously with a secondary task thought to load executive function (such as a memory preload, mental addition, or random generation of numbers). If the two tasks substantially interfere the target task is concluded to involve executive function (e.g. Logie et al., 1994). The most common secondary task used to assess executive function (at least within the working memory framework) is verbal random generation (Baddeley, 1986), in which participants must generate a stream of numbers or letters in as random an order as possible. Random generation has face validity as an executive task: it is a fairly novel task for most people; it is effortful in that it requires inhibition of stereotyped sequences in order to attain a high degree of randomness; and it involves working memory to keep track of recent responses.

There are some advantages to the dual-task method of assessing executive function: it can be used to investigate whether executive components are involved in a variety of different target tasks (and as such it is surprising that there have not been more investigations of the susceptibility of "frontal" tasks to concurrent random generation). Also, it offers the possibility that different types of secondary task might be used to tap different aspects of executive function, although thus far most dual-task literature has assumed that a "general" executive resource is coordinating the two tasks. Finally, it offers the possibility that executive processes can be dissociated from non-executive verbal and visuospatial task components through the use of a range of secondary tasks. There are also problems with the dual-task method. Little is understood about either the processes involved in random generation itself, or those involved in coordinating two tasks. For example, the two tasks might be carried out

simultaneously, or by allocating alternating short time slots to first one, and then the other (Pashler, 1994).

Frontal Lobe Tests

The most common method of assessing executive function within patient populations is to use batteries of tasks interchangeably called "frontal lobe" or "executive" tests. These tests are also increasingly used in normal populations, especially to measure developmental changes in executive function. Because there is currently a fairly widespread consensus that the frontal lobes are the seat of executive functioning, any test found to show specific sensitivity to lesions in that area is accepted as an executive test. Fortunately, most frontal tests do have some face validity as measures of executive processes: they are generally novel (in fact often so novel that participants find them bizarre) and require effort in terms of online monitoring or response inhibition, and most make demands on working memory.

Typical examples of frontal lobe tests include the Tower of London (TOL), the Wisconsin Card-sort Test (WCST) and measures of fluency. In the TOL participants are required to manipulate a pattern of disks on rods to reach a specified goal pattern. This task requires the forward planning of sequences of goals (Shallice, 1982). The WCST requires participants to discover a rule by which to place a series of cards into categories; the rule then changes and they must identify the new rule. This task is likely to depend on the monitoring of internal and external goals and inhibition of the tendency to perseverate (Riccio et al, 1994). Fluency tasks require the rapid production of a sequence of words or designs which fit specified criteria, and deficits have been attributed to an inability to generate appropriate task strategies (Lezak, 1995).

Poor performance on these tests is widely interpreted to reflect executive deficits. However, there are some fundamental problems with the assumption that these tests measure executive function. Not everyone accepts the current *Zeitgeist* that the major (or only) function of the frontal lobes is executive control processes (Damasio, 1994; Reitan & Wolfson, 1994). Also, there are frequent findings of frontal patients who perform perfectly well on such tests (e.g. Shallice & Burgess, 1991a), and patients with non-frontal lesions who perform poorly on the supposedly frontal-specific tests (Anderson, Damasio, Jones, & Tranel, 1991). More importantly in terms of understanding the cognitive processes involved in frontal tests, there is a lack of evidence that poor performance is due to executive deficits. There are a number of potential reasons for performing poorly on any of these tests, and the assumption that any deficit can be explained in executive terms does not seem justified without supporting evidence.

FLUENCY AS A MEASURE OF EXECUTIVE FUNCTION

One of the most widely used frontal tests in both clinical and normal populations is letter fluency. This test originated from Thurstone's largely unsuccessful endeavour to find independent factors of intelligence. Letter fluency assessment requires participants to generate as many words as they can beginning with a specified letter in a short time period. A number of authors particularly recommend the use of fluency tests to tap frontal or executive function, because such tests are easy to use and reliable, and have good discriminatory power (Denckla, 1994; Lezak, 1983; Parker & Crawford, 1992). Poor performance on fluency tests is generally interpreted to reflect executive dysfunction. However, little is known of the cognitive processes involved in tackling fluency tasks (Light, 1992; Randolph, Braun, Goldberg, & Chase, 1993; Sincoff & Sternberg, 1987). Indeed, because only one measure (number of words generated) is usually taken from fluency performance, we lose out on executive process information that might be gained by more in-depth analysis of the output produced (Denckla, 1994). It is therefore important to ask whether fluency tests really do assess executive function.

The most common method of assessing validity of executive tests within the neuropsychological literature is to establish sensitivity to frontal lobe lesions. Such a strong assumption of functional localisation to a specific (and nebulous) site is undoubtedly questionable; but by this criterion fluency fares as well as any other "executive" test. A number of studies report relationships between frontal lesions and letter fluency scores (Benton, 1968; Milner, 1964; Perret, 1974); although some frontal lobe patients perform well on fluency tests (Shallice & Burgess, 1991a), and non-frontal left hemispheric lesions may also cause fluency deficits (Reitan & Wolfson, 1994).

There is a plausible cognitive story to support the assumption that letter fluency tests assess executive function. Most searches through knowledge are guided by semantic constraints; therefore the request to retrieve words by phonemic criteria is relatively novel (Perret, 1974). Effortful retrieval strategies are needed to do well on letter fluency tests, so that attention can be focused on appropriate words (Baddeley, 1990). There is widespread agreement that letter fluency performance is largely determined by the generation and utilisation of effective retrieval strategies (Baddeley & Wilson, 1988; Crowe, 1992; Estes, 1974; Lezak, 1995; Monsch, Bondi, Butters, Paulsen, Salmon, Brugger et al., 1994; Parker & Crawford, 1992). Quite specific ideas have been proposed about the type of retrieval strategies that might influence fluency performance. Lezak (1995) suggests that retrieving a series of words with the same initial syllable, or with semantic links, would lead to effective performance; Salthouse (1988) proposes that an alphabetical strategy or one based on acoustic features might be adopted to retrieve words. However, despite this widespread agreement on the role of strategy use in fluency tests, there has been little attempt to investigate whether

differences in strategy use relate to variation in fluency. Without such information, it is difficult to assess whether fluency deficits can plausibly be interpreted in terms of executive dysfunction.

The assumption that fluency reflects executive function can also be questioned because fluency tests are used to assess quite different aspects of cognition in other areas of the psychological literature. For example, in Alzheimer's disease, low fluency scores are attributed to deterioration in the storage of verbal information (e.g. Binetti, Magni, Padovani, Cappa, Bianchetti, & Trabucchi, 1993; Chan, Butters, Paulsen, Salmon, Swenson, & Maloney, 1993; Martin & Fedio, 1983; Monsch et al., 1994). In the psychometric literature, letter fluency has been used as a measure of cognitive speed (McCrae, Arenberg, & Costa, 1987) or divergent thinking (Sincoff & Sternberg, 1987). This presents serious problems when we need to interpret low fluency scores: do they represent poor retrieval strategies, poor verbal storage, slow search speed, or low creativity? Without in-depth analysis of the processes involved in fluency performance, the assumption that poor scores reflect executive dysfunction seems unwarranted.

Take, for example, the changes in fluency with increased age. It is known that letter fluency performance declines in the course of normal ageing (Daigneault, Braun, & Whitaker, 1992; McCrae et al., 1987; Mittenberg, Seidenberg, O'Leary, & DiGiulio, 1989; Parkin & Walter, 1992; Whelihan & Lesher, 1985). These deficits have been assumed to reflect a deterioration in in frontal lobe executive function with age (Axelrod & Henry, 1992; Parkin & Walter, 1992; Whelihan & Lesher, 1985). This suggests that the elderly are less able to generate effective retrieval strategies in response to fluency tests. However, there has been little qualitative evaluation of the causes of age differences in fluency, as Light highlights: "Surprisingly, fine-grained analyses comparing the protocols of young and older adults…have not been conducted for fluency tasks, so that we have no way of knowing exactly why older adults produce fewer words" (Light, 1992, p. 114).

Phillips (in press) looked at whether the retrieval strategies which have been proposed to be effective in letter fluency tests (Lezak, 1983; Salthouse, 1988; Sincoff & Sternberg, 1987) changed with age. Strategy use was assessed by looking for phonemic or semantic links between sequences of at least three consecutive words. Most individuals showed some evidence of phonemic strategy use to retrieve words, with the occasional use of a systematic alphabetic strategy, or semantic links between words. Age correlated with fluency scores; however, there was no relationship between age and strategy use in letter fluency. This suggested that age differences in letter fluency were unlikely to be due to executive influences on word retrieval patterns; instead more peripheral factors were implicated, because age differences in fluency strongly overlapped with writing speed.

As previously mentioned, fluency tests were originally conceptualised within theories of intelligence; therefore it comes as no surprise to find that letter fluency scores relate to intelligence (e.g. McCrae et al., 1987; Miller, 1984). However, it is often reported that frontal patients have little impairment in intelligence (e.g. Shallice & Burgess, 1991a), and therefore in the neuropsychological literature intelligence is not normally associated with executive functioning. Also, recent theories of intelligence have often emphasised low level resource limitations rather than metacognitive influences as the basis of individual differences in scores (Anderson, 1992; Eysenck, 1986; Jensen, 1982). This would suggest that the relationship between intelligence test performance and letter fluency in normals is unlikely to be due to the influence of executive strategic factors. However, there have also been contrasting claims that individual differences in intelligence largely reflect the efficiency of executive control over cognitive function (Carpenter et al., 1990; Duncan, 1994; Sternberg, 1985). If so, those of higher intelligence should show more evidence of well-structured retrieval strategies in their generation of words in letter fluency tests. Phillips (in press) found evidence to support this: intelligence test score predicted the extent of retrieval strategy use in letter fluency, with those of higher intelligence being more likely, for example, to utilise spontaneously an alphabetic retrieval strategy. Even when only the first 10 words produced by each participant were considered, those with higher intelligence test scores showed evidence of more structured retrieval.

So, age and intelligence both predicted the number of words generated in letter fluency, but seemed to do so for quite different reasons. The evidence did not support an executive explanation for age-related fluency deficits; but did suggest that the relationship between intelligence and letter fluency might be due to differences in the use of retrieval strategies. It is therefore clear that letter fluency deficits should not *automatically* be assumed to reflect executive dysfunction. Additional protocol analysis may be a useful tool to investigate whether such deficits are strategically based in any particular group.

Non-verbal fluency tests have also been used to assess executive function. In the Ruff Figural Fluency Test (Ruff, Light, & Evans, 1987), a page is filled with boxes, each of which contains five dots arranged symmetrically in a pentagonal shape. Participants are asked to design a different figure in each box by joining two or more dots, using straight lines, and must complete as many distinct figures as possible within a minute. This test may be more dependent on intact executive function than letter fluency because it requires the generation of novel figures, rather than retrieval from long-term memory, and therefore is less susceptible to previously used strategies.

Vik and Ruff (1988) outline two easily identifiable production strategies which should result in good performance if utilised to tackle the figural fluency test. This provides an ideal method by which to assess whether differences in

fluency performance might be attributable to executive factors. Ruff (1989) reports that frontal damage results in poorer figural fluency scores. In one patient with right frontal damage, no identifiable strategy use was found at all, and this patient produced a very low fluency score, with a high number of perseverations. Ruff (1989) interprets this as evidence that poor figural fluency in frontal patients is attributable to deficits in strategic planning. In normal individuals both age and intelligence predict figural fluency scores (Ruff et al., 1987), so again it is of interest to question whether these relationships might be due to differences in the efficiency of executive function. Phillips (in press) found that both age and intelligence predicted the use of production strategies in the figural fluency test, with younger and more intelligent participants more likely to generate coherently related sequences of figures. It therefore seems as though executive factors may be implicated in both age and individual differences in the figural fluency task.

Correlational analyses therefore provide provisional support for interpretation of figural fluency as an executive test. However, it is always problematic to try to attribute causal connections using correlational data. In order to draw stronger conclusions about whether fluency correlations are actually caused by differences in effortful executive function, a more experimental approach is needed.

FLUENCY AND THE DUAL-TASK PARADIGM

As outlined earlier, both frontal lobe tests and dual-task methodology are used as ways of assessing executive function. It therefore seems logical that one way of investigating the involvement of executive processes in fluency tests is by looking at the susceptibility of fluency performance to secondary tasks. There is some evidence that verbal fluency is disrupted by concurrent secondary task performance (Baddeley, Lewis, Eldridge, & Thomson, 1984; Martin, Wiggs, Lalonde, & Mack, 1994), but no investigation has been carried out into whether letter and figural fluency are affected by the most widely used executive-loading secondary task: random generation. Martin et al. (1994) found that letter fluency performance was affected by a "frontal" motor task, but not a semantic secondary task. However, the "frontal" task used, sequence tapping, is often specifically employed within the working memory paradigm to load *non-executive* spatial retention (Logie et al., 1994). Baddeley et al. (1984) found that memory preload significantly interfered with word production in a semantic fluency task, and concluded that executive processes were involved in performance. In the experiment about to be described, letter and figural fluency tasks were given to normal individuals both with and without concurrent random generation. It was predicted that random generation should disrupt fluency performance, especially in the figural fluency task, which is more novel and therefore likely to require greater executive involvement for good performance.

As well as investigating the overall effect of concurrent random generation on fluency scores, it is of interest to look at how individual differences in fluency interact with the imposition of the secondary task. Engle & Rosen (1994) looked at semantic fluency (generating animal names) in two groups of participants: those with high and low working memory spans. Those with high working memory spans produced more words in the control fluency task. When a secondary task involving memory load was imposed, the fluency performance of high-span participants was severely affected. In contrast, those with low working memory spans were barely affected by the imposition of the secondary task. These results can be interpreted as follows: the high working memory span group may have been utilising executive search processes extensively to achieve a high score in the control fluency condition. However, when the secondary task was imposed, executive processing on the fluency task would have been disrupted, resulting in a poorer score. The low-span group may have relied on a rather passive, non-executive search to attempt the fluency task in the control condition, attaining only low scores, but not being so susceptible to the executive-demanding secondary task. This would suggest that high-span participants take a high-effort strategic search approach to fluency tasks, while low-span participants take a more passive automatic search approach.

However, Parks et al. (1988) argue for quite the opposite relationship between fluency scores and effortfulness. They suggest that those who attain high fluency scores do so with less effortful activity. They analysed positron emission tomography (PET) scans of brain metabolic activity during fluency performance and found that those who achieved higher fluency scores had less metabolic activity in the frontal lobes during the testing session. Parks et al. interpret their results as evidence that high fluency scores are attained through the use of effective, but cognitively undemanding, strategies. They make an analogy with swimming proficiency, suggesting that high fluency scorers (like experienced swimmers) can go further on less effort. This would suggest that high fluency represents utilisation of relatively automatic acquired metastrategies, rather than the effortful use of online executive processes.

Either interpretation of the relationship between fluency performance and effortfulness might underlie the correlation discussed in the previous section between intelligence test performance and fluency strategy use. This leads to two possibilities: (1) intelligence predicts fluency strategy use because higher-intelligence individuals take a more effortful, strategic approach to fluency tests. This would fit in with the Engle and Rosen findings. Previous studies support the idea that poor intelligence test performance reflects passive, non-analytical approaches to tasks, and poor planning and monitoring of goals (Carpenter et al., 1990; Duncan, 1994; Hunt, 1974). If this is the case then concurrent random generation would interfere more with fluency performance in *higher*-intelligence individuals because their effortful approach to the task would be more disrupted

by an executive load. (2) Intelligence predicts fluency strategy use because those with higher intelligence have more expertise at verbal tasks, and thus more effective metastrategies to attempt fluency tests. This would fit in with the Parks et al. (1988) conclusion. The high strategy use of high intelligence participants would in this case require little expenditure of cognitive resources. Concurrent random generation would therefore cause a greater fluency reduction in those with *lower* intelligence because they have to work harder to generate fewer words.

So differential predictions of the effects of random generation on fluency performance can be made, depending on whether high fluency reflects effortful executive processing, or the application of flexible, but relatively effortless metastrategies. It is worth noting that a greater fluency reduction in lower-intelligence individuals would also be expected if the intelligence/ fluency relationship reflects differences in general speed of processing or working memory resources rather than the influence of executive function. Those of lower intelligence are often found to be disproportionately affected by any increase in task complexity (Cohn, Carlson, & Jensen, 1985), including the introduction of a secondary task (Stankov, 1983), and this is interpreted to reflect limitations on general cognitive resources (Eysenck, 1982; Jensen, 1982).

Method

Participants. Thirty-two members of the Aberdeen Psychology Department participant panel were recruited, ranging in age from 21 to 74. For each participant, scores on the Weschler Adult Intelligence Scale, revised (WAIS) had been obtained one to three years previously. Participants were classified as either low or high WAIS, depending on whether they scored below or above the group mean scaled score (109) across all verbal and performance tests. Note that these scores are not age-corrected IQ measures.

Procedure. All participants completed four fluency conditions, each with a time limit of one minute: (1) control letter fluency: oral production of words beginning with a specified letter; (2) control figural fluency: producing as many different figures as possible by joining at least two dots in a set pattern of five, using straight lines (for designs see Fig. 9.1); (3) letter fluency+RG: oral production of words beginning with a specified letter, while simultaneously generating random keypresses (see below); (4) figural fluency+RG: producing figures by joining at least two dots with straight lines, while simultaneously generating random numbers out loud (see below).

Performance in each condition was assessed as the number of valid items (words or figures) produced, with subtraction of any repetitions. The order of test presentation was counterbalanced to prevent practice effects confounding interpretation. Half of the participants from each WAIS group carried out the

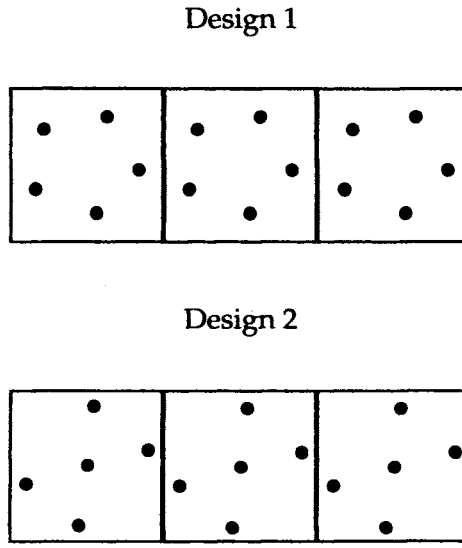


FIG. 9.1. Designs used in figural fluency tests. In the tests a whole A4 page was filled with the boxes. Participants were instructed to join two or more dots in each box, using only straight lines.

tests in order 1 (see below), the other half in order 2. The letters "B" and "F" were used because they had previously been reported to be approximately matched for difficulty (Borkowski, Benton, & Spreen, 1967). Two different figural fluency designs were used (see Fig. 9.1). Instructions given for the figural fluency tasks were taken from the Ruff test manual (Ruff, 1989).

Order 1

Letter fluency using letter "F"
Figural fluency using design 1
Letter fluency using letter "B"+RG
Figural fluency using design 2+RG

Order 2

Figural fluency using design 1+RG
Letter fluency using letter "F"+RG
Figural fluency using design 2
Letter fluency using letter "B"

In the random generation tasks, participants were instructed to generate one keypress or number every 0.8 seconds (having previously heard a metronome playing at that speed) and to produce sequences in as random an order as

possible. In the keypress condition, they rested each of their fingers above a key and had to press one of those keys every 0.8 seconds. In the oral random generation condition, they had to say a number between one and ten every 0.8 seconds. (For further details of random generation tasks see Pearson et al., 1996.) Control random generation performance was also assessed, but is not reported here.

Results

Performance on the fluency tests (in terms of mean number of acceptable items produced in a minute) was analysed using a three-factor analysis of variance, with fluency type (letter/figural) and task condition (control/ +RG) as within-subjects factors, and WAIS group as a between-subjects factor. All comparisons among means were carried out using Tukey's HSD test. There was, as expected, a significant overall effect of WAIS group on fluency performance, $F(1,30)=10.16$, $p<.01$, with the higher WAIS group producing more items in the fluency tasks. There was no overall effect of fluency type, $F(1,30)=3.07$, n.s., but random generation did result in poorer performance, $F(1,30)=22.91$, $p<.001$. Although the effect of random generation was very highly significant, the magnitude of the effect did not suggest a large deficit in performance with the imposition of RG (mean control performance=15.6 items, mean RG performance=12.8 items). There was an interaction between fluency type and task condition, $F(1, 30)= 7.61$, $p<.07$, such that RG affected design, but not letter, fluency.

The three-way interaction between WAIS, fluency type, and task condition was also significant, $F(1, 30)=5.65$, $p<.05$, and is plotted in Fig. 9.2. The magnitude of the RG effect on letter fluency was approximately equivalent in both low and high WAIS groups, and non-significant in both cases. However, the figural fluency results were somewhat different: the low WAIS participants were not significantly affected by having to produce random digits concurrently, while for the high WAIS group, there was a highly significant and large drop in performance when random generation was imposed.

Discussion

The interpretation of fluency tests as measures of executive processing was supported here by the finding of significant interference between random generation and fluency, especially in the case of figural fluency. The non-significant effect of random generation on letter fluency does not support the use of letter fluency as an executive test. In contrast, the significant interference between figural fluency and random generation does support interpretation of this test in terms of executive function. Some caution should be exercised in interpretation of the interaction between random generation and fluency type, because different modalities of random generation were used and oral random generation involves the extra demand of generating verbal codes. However, if the

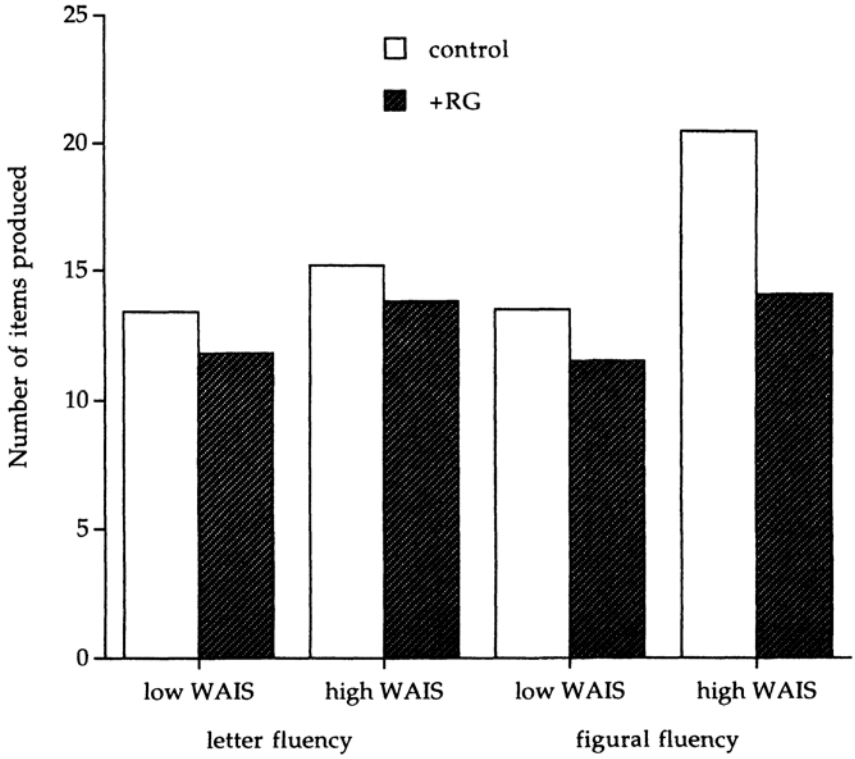


FIG. 9.2. Interaction between WAIS score group, fluency type (letter/figural), and fluency condition (control/random generation).

interference with figural fluency was stronger simply because oral random generation was more difficult this should have affected lower WAIS participants more strongly, which was not the case.

Both low and high WAIS groups showed a non-significant drop in letter fluency performance with random generation. This suggests that in both groups letter fluency was not a strong test of online effortful processing. Instead, findings of correlations between intelligence and fluency strategy use might have been attributable to the higher-intelligence participants using more efficient metastrategies with which to attempt the task. Letter fluency tasks do not seem to be as novel and effortful as is widely assumed. Indeed, they may be susceptible to strategies generated in response to previous encounters with tasks such as crossword puzzles.

In contrast, there was a dramatic effect of concurrent random generation on figural fluency in the high WAIS group, resulting in a drop in performance from about 21 to 14 items produced in a minute. However, there was not a significant effect of random generation on figural fluency in the low WAIS group. This

interaction is contrary to the common finding that lower-intelligence individuals are more affected by increases of task complexity, including the imposition of dual tasks. It therefore seems unlikely that the relationship between intelligence and fluency is due to differences in low-level resource characteristics. The results instead suggest that the relationship between intelligence and figural fluency strategy use reported earlier may be due to differences in the approach to the fluency task in terms of the effortful generation of strategies. High WAIS scorers seem to have adopted an effortful, strategic approach to the control figural fluency task, producing a large number of figures. However, they were also hard hit by concurrent random generation, which resulted in a substantial drop in figural fluency performance. (Interestingly, the high WAIS group performed at almost exactly the same level on the fluency task in the dual-task condition as the low WAIS group did in the control condition.) Low WAIS participants seem to have taken a more passive, non-executive approach to the task, attaining a poor score in the control condition, but being little affected by concurrent random generation. This evidence supports the idea that the relationship between intelligence and figural fluency is attributable to differences in executive function, with high-intelligence individuals adopting a more strategic approach to the task. Figural fluency may therefore be useful as an executive task in normal populations, especially if fluency scores are used in conjunction with assessments of production strategy use.

PROBLEMS IN THE MEASUREMENT OF EXECUTIVE FUNCTION

This final section outlines some of the problems which are inherent in the assessment of executive functioning. These problems include the requirement for novelty, the unpleasantness of executive tasks, and questions of validity and reliability. Such factors should be considered when evaluating executive test scores.

The Need for Novelty

Executive tests must be novel (Denckla, 1994; Duncan, 1986; Shallice & Burgess, 1991b). Any task which is well practised is likely to require automatic rather than effortful processing. Novelty was also emphasised in the construction of fluid intelligence tests, prompting a search for “culture-free” intelligence tests which were not affected by knowledge or experience (Cattell & Cattell, 1960). A typical example of such a test is Raven’s progressive matrices (RPM), designed to be free of the influence of culturally acquired knowledge owing to its novel, non-verbal format (Raven, 1960). RPM seems to have most of the qualities essential for an executive test: it is hypothetically novel, effortful to carry out, and involves working memory (Carpenter et al., 1990). Yet RPM is rarely used

as an executive test within neuropsychology, because intelligence tests do not obey the criterion used to define executive tests, namely sensitivity to frontal lobe lesions.

In fact, questions can be raised about the novelty of RPM: most people are likely to have encountered it, or similar matrix reasoning problems, during the course of educational or occupational assessment. The entire male population of many European countries have sat the RPM (Flynn, 1987). For most people, although the actual *content* of individual items of the RPM may be novel, the *format* of the items is not. Denckla (1994) stresses that in executive tests both content and form must be novel. RPM may therefore not be an appropriate executive test, because many individuals might have an appropriate template, in terms of goal structures, with which to tackle the test (although successful implementation and monitoring would still be required to attempt individual items).

This may also apply to the frontal lobe tests used to measure executive function. For example, letter fluency is supposed to be novel because, in most everyday situations, searches through memory are not carried out using phonemic criteria. However, there are a number of situations in which just such a search might be necessary, for example in word games or crossword puzzles. Crossword experts might find the format (if not the exact content) of letter fluency tasks rather familiar, and be able to use acquired strategies to attempt them. For any particular executive test there may therefore be individual differences in how novel any particular person finds the test format depending on their previous experiences. This raises the possiblity that a frontal lobe patient may perform well on a "frontal" test because that test is not really tapping executive processes in that particular individual.

The requirement for novelty may therefore be problematic, because it is difficult to ensure that a task is equally novel for all individuals. We should be wary of interpreting *good* performance on a frontal test as evidence of intact general executive functioning, without knowing whether the test format and content is novel for that individual.

The Unpleasantness of Executive Tasks

When attempting a cognitive task, participants generally try to formulate an effective strategy to maximise performance at minimal effort (even in tasks as simple as digit span; Belmont, Freeseman, & Mitchell, 1988). They therefore strive to make tasks less effortful and more automatic. However, the whole purpose of executive tests is to prevent this lapse into automaticity and provoke novel and effortful processing. Because of this, executive tasks tend to be rather unpleasant to attempt: "Tasks which are well automatised or pleasantly comfortable are not executive function test battery candidates" (Denckla, 1994, p. 123).

When task requirements are relatively novel (e.g. figural fluency tests) participants can have difficulties in comprehending the task instructions, and this may result in poor performance and distress because they have misunderstood what is required of them. Frontal tests can seem very bizarre to those sitting them. For example, normal individuals find the "cognitive estimates" test (Shallice & Evans, 1978), which requires them to make ballpark estimates of quantities such as the height of a double-decker bus, completely odd, and often find it difficult to take the task seriously. Patients sometimes find the Wisconsin Card-Sort Test so unpleasant that they refuse to comply with test administration (Parker & Crawford, 1992). Bhutani et al. (1992) report floor effects on the Wisconsin Card-Sort Test in the normal elderly because participants reportedly found the task so difficult to understand. While some degree of unpleasantness may be an inevitable feature of tasks designed to be effortful and challenging, distress and demotivation might confound measurement of task performance.

Psychometric Problems with Reliability and Validity

In order for a test to be useful for assessing cognitive function, it must be reliable and valid. Unreliable tests are likely to result in noisy data and poor predictive power. Unfortunately, little is known about the reliability of many frontal lobe tests (Denckla, 1994). The whole idea of assessing test—retest reliability in executive function is problematic because a task cannot be novel the second time around. Even conventional parallel forms of the same test do not overcome this problem: the content may be new, but the format is not. Also, how could the spontaneous generation of, say, a particular task strategy be measured twice?

Executive tests have mainly been validated to a *criterion,* that is, their sensitivity to frontal lobe lesions. This is problematic, given the large variation between frontal lobe patients as to which tasks they perform poorly on. Besides, it is by no means accepted by everyone that executive function resides in the frontal lobes (Reitan & Wolfson, 1994). There is a circularity of argument at work here: frontal lobe patients have been found to perform poorly on tests thought to measure executive function, therefore executive function is located in the frontal lobes, therefore any test on which frontal patients perform poorly is an executive test. This is reminiscent of the circularity that has plagued validation of intelligence tests, whereby a test can only be confirmed as an accurate measure of intelligence if performance correlates highly with established measures of intelligence.

More important is *construct* validity, i.e. whether a test actually measures what it claims to. It is worrying that "frontal" tests have been so widely accepted as measures of executive function when there is such a paucity of empirical evidence regarding the cognitive processes involved in these tasks. However, validation of executive tests is particularly problematic because it is impossible

for a single test score to estimate "pure" executive function without involvement of other aspects of cognition. This is because executive function manifests itself by operating on other cognitive processes. Executive tests may therefore involve linguistic or visuospatial processing, short- or long-term memory components, and motor or verbal responses. If any of these functions fail, there is likely to be a deleterious effect on test performance. This means that there are always many possible reasons for poor performance on an executive test. A low score on an executive test should therefore not be used alone as evidence of dysexecutive functioning.

It is unfortunate that widely used executive tests are of unknown validity. A century of research into intelligence has been plagued with both disagreements about the nature of intelligence, and circular methods of validating test scores. Fortunately, there is reasonable consensus about the cognitive processes thought to be involved in executive function, which bodes well for assessment. Up until now, though, executive tests have mainly been validated against an unreliable criterion, rather than through experimental investigation of the processes involved. There is a pressing need for detailed experimental evaluation of what the tests actually assess.

CONCLUSIONS

Executive functioning involves the planning of goal structures to attempt a task, and monitoring and control of task performance. There is an urgent need for tests which assess executive function accurately, because executive processes are central to recent theories of memory, attention, and intelligence. Also, deficits in executive function have been proposed to explain the cognitive changes that occur in normal ageing, in Alzheimer's disease, and various other neuropsychological disorders. Currently, so-called "frontal lobe" or "executive tests" are widely used in clinical and normal populations. Poor performance upon such tests is attributed in psychological terms to dysfunction of the control processes of cognition, and in neuroanatomical terms to frontal lobe damage. However, the evidence for brain localisation of executive task deficits is rather weak, and there has been little study of the underlying cognitive deficits involved in poor performance.

Evidence relating to fluency, one of the most widely accepted measures of executive function, was examined here in order to provide some preliminary assessment as to whether deficits in performance on letter and figural fluency tasks might be attributable to differences in executive aspects of cognition. Little has been published offering task analysis of fluency performance, or concrete evidence for the involvement of executive processes. The evidence here supported the use of figural fluency as an executive test, because both age and intelligence related to the use of production strategies on the task, and performance was severely disrupted by the imposition of random generation.

However, the evidence was less supportive of the use of letter fluency as an executive task: intelligence (but not age) predicted the use of retrieval strategies, and there was little interference from concurrent random generation.

There is a great need for more information about the executive tasks currently in use through the use of detailed task analysis and experimental manipulations. There are many problems associated with the construction of executive tasks: the demand for novelty, the unpleasantness of the tests, the inability to measure executive function in isolation, and the difficulty of producing reliable tests. It does not yet seem inevitable that assessment of executive function is impossible (as some have argued; see Fodor, 1983; Reitan & Wolfson, 1994); however, scores from neuropsychological tests currently used as measures of executive function should be treated with caution.

ACKNOWLEDGEMENT

The experimental data (fluency and random generation) described in this chapter were collected by Alison Flett and Pat Chalmers, Psychology Department, Aberdeen University.

REFERENCES

Anderson, M. (1992). *Intelligence and development.* Oxford: Blackwell.

Anderson, S.W., Damasio, H., Jones, R.D., & Tranel, D. (1991). Wisconsin Card Sort Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13,* 909–922.

Axelrod, B.N. & Henry, R.R. (1992). Age-related performance on the Wisconsin Card Sorting, Similarities and Controlled Oral Word Association tests. *The Clinical Neuropsychologist, 6,* 16–26.

Baddeley, A.D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A.D. (1990). *Human memory: Theory and practice.* Hove: Lawrence Erlbaum Associates Ltd.

Baddeley, A.D. & Hitch, G. (1974). Working memory. In G.A.Bower (Ed.), *Recent advances in learning and motivation* (pp. 647–667). New York: Academic Press.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long term memory. *Journal of Experimental Psychology: General, 113,* 518–540.

Baddeley, A.D. & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7,* 212–230.

Belmont, J.M., Freeseman, L.J., & Mitchell, D.W. (1988). Memory as problem solving: The cases of young and elderly adults. In M.M.Gruneberg, P.E.Morris, & R.N.Sykes (Eds.), *Practical aspects of memory: Current research and issues.* Chichester: John Wiley.

Benton, A.L. (1968). Differential behavioural effects of frontal lobe disease. *Neuropsychohgia, 6,* 53–60.

Bhutani, G.E., Montaldi, D., Brooks, D.N., & McCulloch, J. (1992). A neuropsychological investigation into frontal lobe involvement in Dementia of the Alzheimer Type. *Neuropsychology, 6,* 211–224.

Binetti, G., Magni, E., Padovani, A., Cappa, S.F., Bianchetti, A., & Trabucchi, M. (1993). Neuropsychological heterogeneity in mild Alzheimer's disease. *Dementia, 4,* 321–326.

Borkowski, J.G., Benton, A.L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia, 5,* 135–140.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of processing in the Raven Progressive Matrices task. *Psychological Review, 97,* 404–431.

Cattell, R.B. & Cattell, A.K.S. (1960). Handbook for the individual or group Culture Fair intelligence test. Champaign IL: IPAT.

Chan, A.S., Butters, N., Paulsen, J.S., Salmon, D.P., Swenson, M.R., & Maloney, L.T. (1993). An assessment of the semantic network in patients with Alzheimer's Disease. *Journal of Cognitive Neuroscience, 5,* 254–261.

Cohn, S.J., Carlson, J.S., & Jensen, A.R. (1985). Speed of information processing in academically gifted youths. *Personality and Individual Differences, 6,* 621–629.

Craik, F.I.M., Morris, R.G., & Gick, M.L. (1990). Adult age differences in working memory. In G.Vallar & T.Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 247–267). New York: Cambridge University Press.

Crowe, S.F. (1992). Dissociation of two frontal lobe syndromes by a test of verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 14,* 327–339.

Daigneault, S., Braun, C.M.J., & Whitaker, H.A. (1992). Early effects of normal aging on perseverative and non-perseverative prefrontal measures. *Developmental Neuropsychology, 8,* 99–114.

Damasio, A.R. (1994). *Descartes' error: Emotion, reason and the human brain.* New York: Grosset/Putnam.

Della Sala, S., Logie, R.H., & Spinnler, H. (1992). Is primary memory deficit of Alzheimer patients due to a "central executive" impairment? *Journal of Neurolinguistics, 7,* 325–346.

Denckla, M. (1994). Measurement of executive function. In G.Reid Lyon (Ed.), *Frames of reference for the assessment of learning disabilities, New views on measurement issues* (pp. 117–142). Baltimore, MD: Paul H. Brookes.

Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuropsychology, 3,* 271–290.

Duncan, J. (1994). Attention, intelligence and the frontal lobes. In M.S.Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Engle, R.W. & Rosen, V. (1994). Working memory capacity and retrieval. International Conference on Working Memory, July, Cambridge, UK.

Estes, W.K. (1974). Learning, theory and intelligence. *American Psychologist, 29,* 740–749.

Eysenck, H.J. (Ed.) (1982). *A model for intelligence.* Berlin: Springer-Verlag.

Eysenck, H.J. (1986). The theory of intelligence and the psychophysiology of cognition. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 1–34). Hillside, NJ: Lawrence Erlbaum Associates Inc.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Fodor, J.M. (1983). *The modularity of mind.* Cambridge, MA.: MIT Press.

Gilhooly, K.J., Logie, R.H., Wetherick, N.E., & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory and Cognition, 21,* 115–124.

Hunt, E.B. (1974). Quote the Raven? Nevermore! In L.W.Gregg (Ed.), *Knowledge and cognition.* Potomac: Lawrence Erlbaum Associates Inc.

Jensen, A.R. (1982). Reaction time and psychometric g. In H.J.Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.

Lezak, M.D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Light, L.L. (1992). The organisation of memory in old age. In F.I.M.Craik & T.A.Salthouse (Eds.), *The handbook of aging and cognition* (pp. 111–165). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Logie, R.H., Gilhooly, K.J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition, 22,* 395–410.

Martin, A. & Fedio, P. (1983). Word production and comprehension in Alzheimer's Disease: The breakdown of semantic knowledge. *Brain and Language, 19,* 124–141.

Martin, A., Wiggs, C.L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks. *Neuropsychologia, 32,* 1487–1494.

McCrae, R.R., Arenberg, D., & Costa, P.T. (1987). Declines in divergent thinking with age: Cross-sectional, longitudinal and cross-sequential analyses. *Psychology and Aging, 2,* 130– 137.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology, 23,* 53–57.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J.M.Warren & K.Akert (Eds.), *The frontal granular cortex and behaviour* (pp. 313–334). New York: McGraw-Hill.

Mittenberg, W., Seidenberg, M., O'Leary, D.S., & DiGiulio, D.V. (1989). Changes in cerebral functioning associated with normal aging. *Journal of Clinical and Experimental Neuropsychology, 11,* 918–932.

Monsch, A.U., Bondi, M.W., Butters, N., Paulsen, J.S., Salmon, D.P., Brugger, P., & Swenson, M.R. (1994). A comparison of category and letter fluency in Alzheimer's disease and Huntingdon's disease. *Neuropsychology, 8,* 25–30.

Norman, D.A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R.J.Davidson, G.E.Schwartz, & D.Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). New York: Plenum Press.

Parker, D.M. & Crawford, J.R. (1992). Assessment of frontal lobe dysfunction. In J.R. Crawford, D.M.Parker, & W.M.McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 267–291). Hove: Lawrence Erlbaum Associates Ltd.

Parkin, A.J. & Walter, B.M. (1992). Recollective experience, normal aging and frontal dysfunction. *Psychology and Aging, 7,* 290–298.

Parks, R.W., Lowenstein, D.A., Dodrill, K.L., Barker, W.W., Yoshii, F., Chang, J.Y., Emran, A., Apicella, A., Sheramata, W.A., & Duara, R. (1988). Cerebral metabolic effects of a verbal fluency test: A PET scan study. *Journal of Clinical and Experimental Neuropsychology, 10,* 565–575.

Pashler, H. (1994). Dual task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Pearson, D.G., Logie, R.H., & Green, C. (1996). Mental manipulation, visual working memory, and executive processes. *Psychologische Beitrage, 38,* 324–342.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal category behaviour. *Neuropsychologia, 12,* 323–330.

Phillips, L.H. (in press). Age and individual differences in letter fluency. *Developmental Neuropsychology.*

Randolph, C., Braun, A.R., Goldberg, T.E., & Chase, T.N. (1993). Semantic fluency in Alzheimer's, Parkinson's and Huntingdon's disease: Dissociation of storage and retrieval failures. *Neuropsychology, 7,* 82–88.

Raven, J.C. (1960). *Guide to the Standard Progressive Matrices.* London: H.K.Lewis.

Reitan, R.M. & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4,* 161–198.

Riccio, C.A., Hall, J., Morgan, A., Hynd, G.W., Gonzalez, J.J., & Marshall, R.M. (1994). Executive function and the Wisconsin Card Sorting Test: Relationship with behavioural ratings and cognitive ability. *Developmental Neuropsychology, 10,* 215–229.

Roberts, R.J., Hager, L.D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393.

Ruff, R.M. (1989). *Ruff Figural Fluency Test administration manual.* San Diego: Neuropsychological Resources.

Ruff, R.M., Light, R.H., & Evans, R.W. (1987). The Ruff Figural Fluency Test: A normative study. *Developmental Neuropsychology, 3,* 37–51.

Salthouse, T.A. (1988). Effects of aging on verbal abilities: Examination of the psychometric literature. In L.L.Light & D.M.Burke (Eds.), *Language, memory and aging* (pp. 17– 35). Cambridge: Cambridge University Press.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, 298,* 199–209.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shallice, T. & Burgess, P. (1991a). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shallice, T. & Burgess, P. (1991b). Higher order cognitive impairments and frontal lobe lesions in man. In H.S.Levin & A.L.Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125– 138). New York: Oxford University Press.

Shallice, T., & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 4,* 294–303.

Sincoff, J.B. & Sternberg, R.J. (1987). Two faces of verbal ability. *Intelligence, 11,* 263–276.

Stankov, L. (1983). Attention and intelligence. *Journal of Education Psychology, 75,* 471–490.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge: Cambridge University Press.

Sternberg, R.J. & Detterman, D.K. (Eds.) (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Stuss, D.T. & Benson, D.F. (1986). *The frontal lobes.* New York: Raven.

Vik, P. & Ruff, R.M. (1988). Children's figural fluency performance: Development of strategy use. *Developmental Neuropsychology, 4,* 63–74.

Whelihan, W.M. & Lesher, E.L. (1985). Neuropsychological changes in frontal functions with aging. *Developmental Neuropsychology, 1,* 371–380.

CHAPTER TEN
A Neural Systems Approach to the Cognitive Psychology of Ageing Using the CANTAB Battery

T.W.Robbins, M.James, A.M.Owen, B.J.Sahakian, L.McInnes, Patrick Rabbitt*

INTRODUCTION

Considerable data are now available which bear on cognitive theories of ageing (Rabbitt, 1993; Salthouse, 1985, 1991). However, these data and theories are not well integrated with accumulating neurobiological information about the ageing brain. While ageing often leads to generalised atrophy of the brain (Petit, 1982), specific regions and systems have from time to time been implicated (Flood & Coleman, 1988). For example, there is now accumulating evidence of major effects in structures such as the hippocampus (e.g. Nagahara, Nicolle, & Gallagher, 1993), and regions of the association cortex (e.g. parietal, prefrontal, certain regions of the temporal cortex; Brody, 1994; Mann, Yates & Marcyniuk, 1984; Parashos & Coffey, 1994) as well as in certain of the chemically defined systems of the isodendritic reticular core of the brain (Whitehouse, 1994), such as the noradrenergic locus coeruleus (Vijayshankar & Brody, 1979), the dopamine-containing cells of the substantia nigra, pars compacta (Hirai, 1968; Marshall, Drew, & Neve, 1983), and the cholinergic cells of the basal forebrain (Fis cher, Chen, Gage, & Bjorklund, 1991; Smith, Deadwyler, & Booze, 1993; Stroesner-Johnson, Rapp, & Amaral, 1992). Evidence of disproportionate atrophy of the frontal cortex in ageing (Petit, 1982) and selective reductions in regional cerebral blood flow in this region (Shaw et al., 1984) have encouraged some to focus on neuropsychological changes in "frontal" function in ageing

* T.W.Robbins, Department of Experimental Psychology, University of Cambridge, Cambridge, U.K.; M.James, Department of Psychology, National Hospital, London, U.K.; A.M.Owen and B.J.Sahakian, Department of Psychiatry, University of Cambridge, Cambridge, U.K.; L.McInnes, North-East Age Research Panel, Department of Psychology, Newcastle University, U.K.; Patrick Rabbitt, Age and Cognitive Performance Research Centre, Manchester University, U.K.

(Whelihan & Lesher, 1985; see West, 1996, for a recent review). This chapter considers the use of neuropsychological tests in the healthy ageing population with proven utility in the assessment of brain-damaged patients, where it is possible to infer the relationship between performance deficits and the anatomical localisation of the brain damage or dysfunction. Such an approach may even enable the development of novel hypotheses at the purely cognitive level. We will focus particularly on tests sensitive to frontal lobe dysfunction that are suggested to tap aspects of "executive" function, those control processes by which an individual optimises his or her cognitive performance.

This neuropsychological approach is facilitated by several developments: first the growing use of structural imaging techniques (such as MRI) for gaining precise information about the locus of brain damage; second, the use of functional imaging (using PET or MRI) to provide information from healthy controls about the likely neural substrates of complex cognitive operations; third, growing knowledge about the interconnectivity and organisation of neural structures into discrete systems; and finally the use of experimentally discrete lesions and manipulations to make inferences concerning localisation of function in experiments with animals.

As much of our knowledge about ageing in the CNS depends on the interpretation of neuroanatomical and neurochemical evidence from experiments with animals, parallel information concerning the functional correlates of these changes is important when attempting to relate the findings to humans. Consequently, this has required the development of sophisticated neuropsychological tests for animals in several cognitive domains. In some cases, such test development has been based on understanding how cognitive capacities in animals have been shaped by environmental and evolutionary constraints, for example, in foraging. In other cases, the test development depended upon advances in animal learning theory. Applying this information to the problem of human ageing, however, has been limited by the degree to which the experimental paradigms can be extrapolated to humans. This important issue has been addressed so far by only a few other investigators (e.g. Bartus, Dean, & Beer, 1980; Flicker, Bartus, Crook, & Ferris 1984; Solomon, Beal, & Pendlebury, 1988), mainly stimulated by research into Alzheimer's disease.

This connection between work on experimental animals and human neurological disease has formed a major focus of our own research effort in the design of the Cambridge Neuropsychological Test Automated Battery (CANTAB), the theoretical impetus for which was to adapt paradigms developed for testing animal models of dementia, in order to relate the findings to man. This battery thus comprises a set of computerised neuropsychological tests, some of which are based on neuropsychological tests for animals, and some of which are versions of established clinical neuropsychological tests, modified so that they can also be used in experiments with animals.

The battery, which operates on a standard IBM-type PC, capitalises on novel technological developments which include the use of a touch-sensitive screen, so that immediate feedback for responding to the test stimuli can be provided most effectively and the participant is thus not faced with the problems of divided attention provided by the conjoint requirement to attend to a video display unit but respond via a keyboard. The tests are graduated in level of difficulty, and so boost motivation in the elderly participant. The use of such computerised tests, after initial screening and training to overcome possible problems posed by the unfamiliar test setting has the advantage of avoiding the distress sometimes associated with pencil and paper tests and formal interviews. The stimuli and contingencies of the tests are presented in a standard way, and objective and accurate measures of responding, particularly in terms of latency, are ensured. Finally, the computerised collection of data should simplify the collation and analysis of results from large-scale longitudinal and cross-sectional studies of ageing.

CANTAB BATTERIES

Currently, the CANTAB tests are divided into three main batteries, "visual memory and learning," "working memory and planning," and "attention." In each battery, the principle is that there are one or more complex tests (i.e. having several cognitive components) that can be fractionated into their constituent parts. Thus, the visual memory battery has separate tests of visual pattern recognition and spatial recognition memory, as well as a difficult visual delayed matching to sample task. There is also a visuospatial paired associates task which requires the participant to remember and learn the location of up to eight different visual patterns. Thus, the pattern and spatial recognition tasks are likely to tap components of the paired associate learning task (as confirmed by factor analysis, see below, Table 10.2; see Robbins et al., 1994 for full test descriptions).

The working memory and planning battery has three main tests (see Owen, Downes, Sahakian, Polkey, & Robbins, 1990): the Tower of London test of single contingency planning, devised by Shallice and McCarthy (Shallice, 1982); a self-ordered spatial working memory task; and a test of spatial span, modelled after that of Corsi (Milner, 1971). The reasoning behind this set of tests is that the two tests of spatial memory provide controls for some aspects of performance on the test of planning, for example, the ability to retain and execute a sequence of five moves, the longest number of moves required in the planning test. The form of the Tower of London test employed requires the participant to consider two arrangements of a set of three differently coloured balls resting in three suspended "socks" or "stockings" of different lengths. The top arrangement is the goal position which the participant has to reach by manipulating the balls in the initial position, below. Merely touching the appropriate ball and then the position to which it is to be moved is all that is required. The participants are

instructed on the number of moves in which each problem can most efficiently be completed. The accuracy of solving the problem (in terms of both the number of "excess moves" and "perfect" solutions, which use the minimum number of moves) provides measures of accuracy of problem solving, whereas the latencies to think either prior to the initiation of the problem solution ("initial thinking time"), or during it ("subsequent thinking time") are also measured. These latencies are corrected by subtracting "movement time" measures; the problem solutions are played back to the participant move-by-move on the upper display, and the times taken to copy the computer-generated moves on the lower display are subtracted from the total latencies. It is important to realise that this test was designed to measure "look ahead", planning function, rather than the trial and error sequencing that mainly characterises the Tower of Hanoi variant, which has recently been criticised as a test sensitive to frontal lobe dysfunction (Goel & Grafman, 1995). Furthermore, certain measures in the test (e.g. thinking times for perfect solutions) provide measures of planning that are *not* contaminated by move retracing elements that are claimed to confound pure measures of planning (see Goel & Grafman, 1995).

Finally, the "attentional" battery consists not only of easy one-choice and five-choice reaction time tasks, similar to those employed by Leonard, and in experimental analogues in rats (see Robbins, Muir, Killcross, & Pretsell, 1993), but also a visual search paradigm, and a set of visual discrimination tasks which decompose the Wisconsin Card-Sorting Test (WCST). The latter is often used as a clinical test for impaired frontal lobe function. The visual search procedure requires the participants to match complex visual stimuli which are similar to those used in the delayed matching to sample test of the visual memory battery. They must match a centrally displayed stimulus to a peripheral stimulus placed in one of eight locations around the screen by releasing a keypad (reaction time) and touching the appropriate stimulus on the screen (movement time). Sets of two, four or eight stimuli are displayed, so that the participant has to conduct a serial search, according to a Sternberg-like procedure (see Downes et al., 1989).

The WCST analogue is based on a set of visual discriminations, initially using simple stimuli (e.g. shapes or lines) and then stimuli compounded from these two perceptual dimensions. The participant has to respond selectively to an exemplar from one dimension during discrimination learning and reversal procedures, on the basis of computer feedback, and then receives transfer tests of intra- or extradimensional shifts, with novel exemplars to assess the capacity to maintain selective attention to the entrained dimension, or to shift to a previously irrelevant one. The extradimensional shift, as defined in animal and human learning paradigms, is in fact the core requirement of the WCST (see Downes et al., 1989 for full description).

The battery thus contains at least three tests that might be expected to be sensitive to frontal lobe dysfunction, on the basis of clinical and experimental

TABLE 10.1 Summary of Neuropsychological Findings Using the CANTAB Battery

	Temporal/ Hippocampal Lesions	*Frontal Lesions*	*DAT*	*Parkinson's Disease**
Tests of Visual Memory				
Pattern Recognition	✖	✓	✖	✓
Spatial Recognition	✓	✖	✖	✖
Delayed matching to sample	✖	✓	✖	✖
Tests Sensitive to Frontal Lobe Dysfunction				
Spatial working memory				
Between error score	✖	✖	✖	✖
Strategy score	✓	✖	✓	✓
Tower of London				
Minimum moves	✓	✖	✖	✓
Initial thinking time	✓	✓	—	✖
Subsequent thinking time	✓	✖	—	✓
Attentional shift task	✓	✖	✓ (early) ✖ (late)	✖

NOTE: ✖=impairment; ✓=no impairment; DAT=dementia of the Alzheimer type.
* The exact effects of Parkinson's disease depend on disease severity.
SOURCE: Taken from Downes et al., 1989; Owen et al., 1990, 1991, 1992, 1993a, 1995; Sahakian et al., 1988, 1991; Sahgal et al., 1991, 1992.

evidence (see Table 10.1): the Tower of London test of planning, the spatial working memory task, and the WCST analogue (see Fig. 10.1). However, it is an important methodological feature of the battery that each of these tests is not only graded in difficulty, but also generally decomposed into a number of cognitive components. It is an unusual feature of our approach in assessing cognitive performance during ageing that we consider several different aspects of executive and non-executive function in tests that have been carefully studied using patients with neurosurgical excisions of different regions of the cortex, including the prefrontal cortex.

NEURAL VALIDATION

The CANTAB tests have been quite widely used, on a variety of clinical populations, as well as in functional imaging studies and for the testing of effects of psychoactive drugs, in normal volunteers. In the clinical application of CANTAB, adequate comparisons have to be made with the performance of normal volunteers who are matched with the patients on the basis of age, gender, and IQ measures. This requirement to standardise the CANTAB tests in large populations of elderly volunteers has provided some of the impetus for the data reported in this chapter. It is clear that the various tests provide dissociations of deficits across different clinical populations, sometimes on a basis that can be related to the dysfunctioning of quite specific brain regions. For example, the

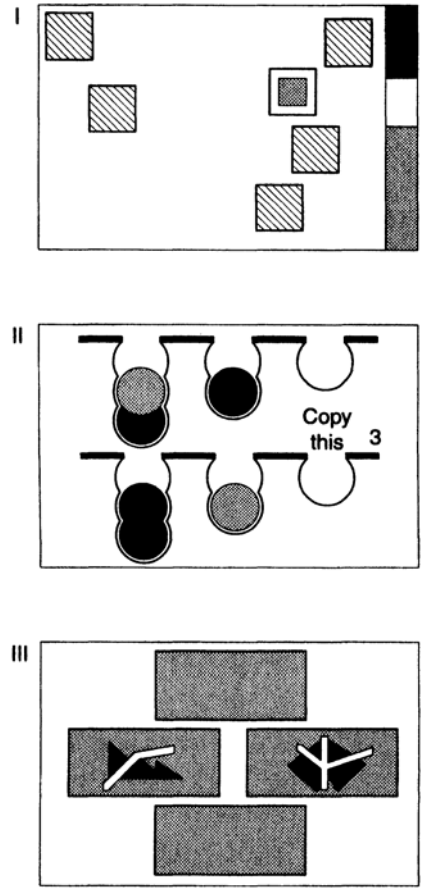


FIG. 10.1. Schematic drawings of three tests sensitive to frontal lobe dysfunction: I. spatial working memory; II. the Tower of London test of planning; and III. the attentional set-shifting task (intra/extradimensional shifting). Source: Robbins et al. (1992). Reproduced by kind permission of Oxford University Press. See Downes et al., 1989 and Owen et al., 1990 for further explanation.

tests of pattern recognition memory and delayed matching to sample are indeed sensitive to neurosurgical damage of the temporal lobes, including the amygdala and hippocampus (Owen, Sahakian, Semple, Polkey, & Robbins, 1995; see Table 10.1), as would have been expected on the basis of certain animal studies (Mishkin, 1982). On the other hand, performance on such tests is relatively insensitive to neurosurgical excisions of the frontal lobe (Owen et al., 1995; Table 10.1). In contrast, general performance on tests sensitive to frontal lobe dysfunction, such as the extradimensional shift paradigm and the Tower of London, is not impaired by temporal lobe damage (Owen et al., 1990, 1991,

1995). Furthermore, on certain tests, for example, spatial working memory, there are impairments in some aspects of test performance following frontal lobe damage that are not present following damage to the temporal lobe, and vice versa (see Owen et al., 1995, 1996c). Thus, the use of an effective strategy for conducting this self-ordered spatial working memory task is impaired by frontal lesions in humans, whereas this is not the case for lesions of the temporal lobe, which nevertheless still impair performance at the more difficult levels. The visuospatial paired associate learning test is another that is affected by damage to both regions, although probably for different reasons (Owen et al., 1995).

There is less evidence available to differentiate the possible role of the parietal cortex in performance of these tests, but recent evidence using PET (Baker et al., 1996; Owen et al., 1996a, 1996b) makes it clear that the parietal cortex is heavily implicated, probably bilaterally, in the Tower of London test. In the latter functional imaging study, the left dorsolateral prefrontal cortex, as well as the left caudate nucleus also exhibited increases in regional cerebral blood flow (rCBF). Thus, it is important to realise that the Tower of London task activates a network of neural structures with related functions.

Activation using PET has also illuminated the neural substrates of three tests of different aspects of spatial memory. For example, tasks similar to the CANTAB spatial recognition memory test are known to activate the right dorsolateral prefrontal cortex in man, in keeping with its similarity to the delayed response task that is prototypically associated with this region in work with monkeys (Jonides et al., 1993; Owen et al., 1996a). Moreover, a test similar to spatial span activates the right midventrolateral regions, whereas the spatial working memory task activates both of these zones (Owen et al., 1996a). These data are also consistent with the rationale of the CANTAB working memory and planning battery: the separate tests of spatial function help to decompose the components of the planning task.

Experiments with monkeys performing the extradimensional shift test have recently shown that lesions of the dorsolateral prefrontal cortex selectively impair the extradimensional shifting component while leaving reversal learning unaffected, whereas orbitofrontal lesions have the opposite patterns of effects (Dias et al., 1996). These data are consistent with the view that the prefrontal cortex is heterogeneous with respect to function in primates (Petrides, 1994), and suggest that "executive functions" may themselves comprise a loose collection of control processes, at least some of which are mediated by different regions of the prefrontal cortex.

Results from patients with Alzheimer's and Parkinson's diseases also help to define neural and neurochemical substrates of some of the CANTAB tests (see Table 10.1). For example, the anticholinesterase tacrine has been shown to improve the accuracy and speed of performance of patients with Alzheimer's disease (Sahakian et al., 1993), data consistent with results from animal studies

suggesting that accuracy of detection of visual signals in a similar Continuous Performance test depends on the integrity of the cholinergic projections from the basal forebrain to the frontal cortex (Muir, Everitt, & Robbins, 1995). Moreover, initial thinking time is slowed in Parkinson's disease patients, but remediated partly through treatment with L-Dopa (Lange et al., 1992), suggesting that the slowing of movement caused by nigro-striatal dopamine degeneration is paralleled by a slowing of thinking ("bradyphrenia") which may similarly depend on dopamine loss from the striatum. This critical structure is quite possibly the caudate nucleus, on the basis of the functional imaging studies of normal volunteers described above. Both serotoninergic and noradrenergic influences on cognitive function are suggested by recent studies of the effects of low tryptophan drinks (Park et al., 1994) or the alpha-2 receptor agents clonidine and idazoxan (Coull, Middleton, Robbins, & Sahakian, 1995 a, b), on different aspects of performance on the CANTAB tests. Some of these data are consistent with the hypothesis that alpha-2 receptors in the frontal cortex are implicated in spatial working memory performance, especially in aged monkeys (Arnsten & Goldman-Rakic, 1984).

Overall, it should be realised that it is naive to assume that there is necessarily a one-to-one relationship between a deficit on a given neuropsychological test and brain damage to an anatomically defined region. Many brain regions are interconnected in neural networks which mean that performance on a particular test may be impaired by lesions to widely distributed brain structures. This is particularly true of the prefrontal cortex which has major reciprocal projections to posterior cortical areas, as well as to subcortical areas such as the hypothalamus, the striatum, and the chemically defined systems of the reticular core of the brain (see Goldman-Rakic, 1987). The challenge posed by this anatomical circuitry is to understand which distinct types of information processing are subserved by each component of the neural system or network. However, it is clear that the way in which the prefrontal cortex is anatomically interconnected with different brain regions makes it an ideal candidate for coordinating processing between many different regions and thus mediating many aspects of executive function.

STUDIES WITH LARGE POPULATIONS OF NORMAL ELDERLY VOLUNTEERS

It has been possible to gain insights into the utility of the neuropsychological approach espoused here through a standardisation of the CANTAB tests on a large number of elderly volunteers (up to about 800) in collaboration with the North-East Age Research Panel under the direction of Professor P.Rabbitt. This study has made possible a detailed analysis of how cognitive functions measured by the CANTAB tests, which presumably reflect the altered functioning of some

of the component neural systems, discussed above, decline with age. As the study was conducted on a panel that has been engaged in a detailed longitudinal investigation of the cognitive effects of ageing, it has also been possible to utilise information from other psychometric tests intrinsic to the longitudinal study, as well as to make useful comparisons with other neuropsychological batteries that have been used on this population. Although this study is necessarily cross-sectional at present, with all the problems of interpretation that this entails in ageing studies, it is planned to complement it by retesting the panel at a later date.

Most of the data to be described were obtained from a population aged between 55 and 79 years. A more limited data set is available from volunteers younger than 50 years. Thus, it is possible to examine age-related declines in performance over this age range, and in that sense identify which neuropsychological tests, and putatively which brain systems, are most vulnerable to the effects of ageing. This analysis, however, is limited only to this particular age range and this has certain problems of interpretation. For example, it appears that there are very large differences in performance of the extradimensional shift test between small groups of control participants relatively young (i.e about 30 years old) and old (70 years plus) (Owen et al., 1991). This result has recently been confirmed using an independent sample of young and elderly participants (Fig. 10.2). But it appears that most of this difference must occur before 55 years, as there is little further decline in performance specifically in extradimensional shifting after that age, although the ability to form and maintain an attentional set shows a small degree of impairment (see Fig. 10.2). The problems of interpretation are thus analogous to those of identifying tests sensitive not only to cognitive decline in early Alzheimer's disease, but also to its subsequent course.

Another problem in tracking performance over wide ranges of age and intellectual ability is that of avoiding tests contaminated by strong ceiling and floor effects. This problem potentially affects interpretation of results from the delayed matching to sample paradigm in the present study (Fig. 10.3). Memory retention declines mildly over delays in the youngest group (<50 years), but is significantly inferior to performance in the perceptual control condition of simultaneous matching to sample condition, demonstrating some independence from ceiling effects at superior levels of performance. Performance of the 55–59 age group is reduced in parallel over the different delays compared with the youngest group, suggesting no specific effect of ageing on this form of recognition memory in these middle-aged groups, but a general, delay-independent decline in recognition memory accuracy that might be attributable to a perceptual factor. However, some of the subsequent declines in performance appear to be delay-dependent, as shown by the significant reductions in accuracy observed in individuals from the 55–59 year age group compared with those from

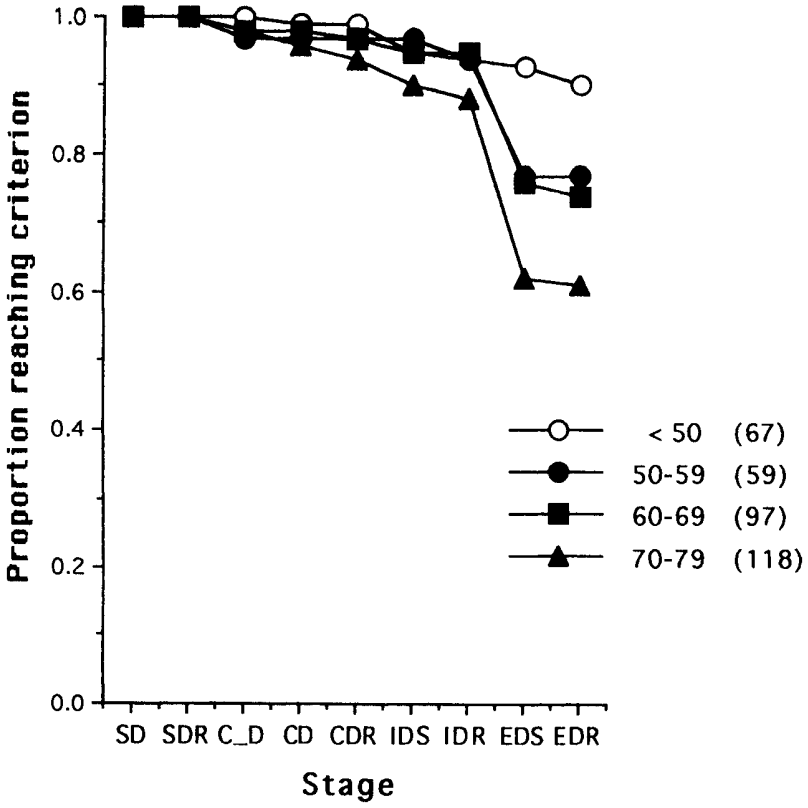


FIG. 10.2. Cumulative proportion of participants successfully reaching criterion at each stage of the attentional set-shifting paradigm. SD=simple discrimination; SR=simple reversal; C—D=compound discrimination, separated elements; CD=compound discrimination, superimposed elements; CDR=compound reversal; intradimensional shift; IDR=intradimensional reversal; ED=extradimensional shift; EDR=extradimensional reversal. *Ns* refer to subjects in each age band, as defined.

the 60–64 year age group, evident at the longest delay (12s). These data suggest that the neural system in control of visual recognition memory, which presumably incorporates several components of the temporal neocortex (Murray, 1992), is susceptible to the effects of ageing. There are further, age-dependent declines in accuracy of performance for older groups, but these are limited to relatively short (4s) delays. At the longer delays, performance accuracy appears to reach floor levels in this population of about 70% correct, probably constrained by verbal encoding strategies based on colour. It is significant that patients with Alzheimer's disease at similar ages may exhibit more pronounced delay-dependent deficits in percentage correct scores that reflect the failure of encoding strategies (see Sahakian et al., 1988; Sahgal et al., 1991). The

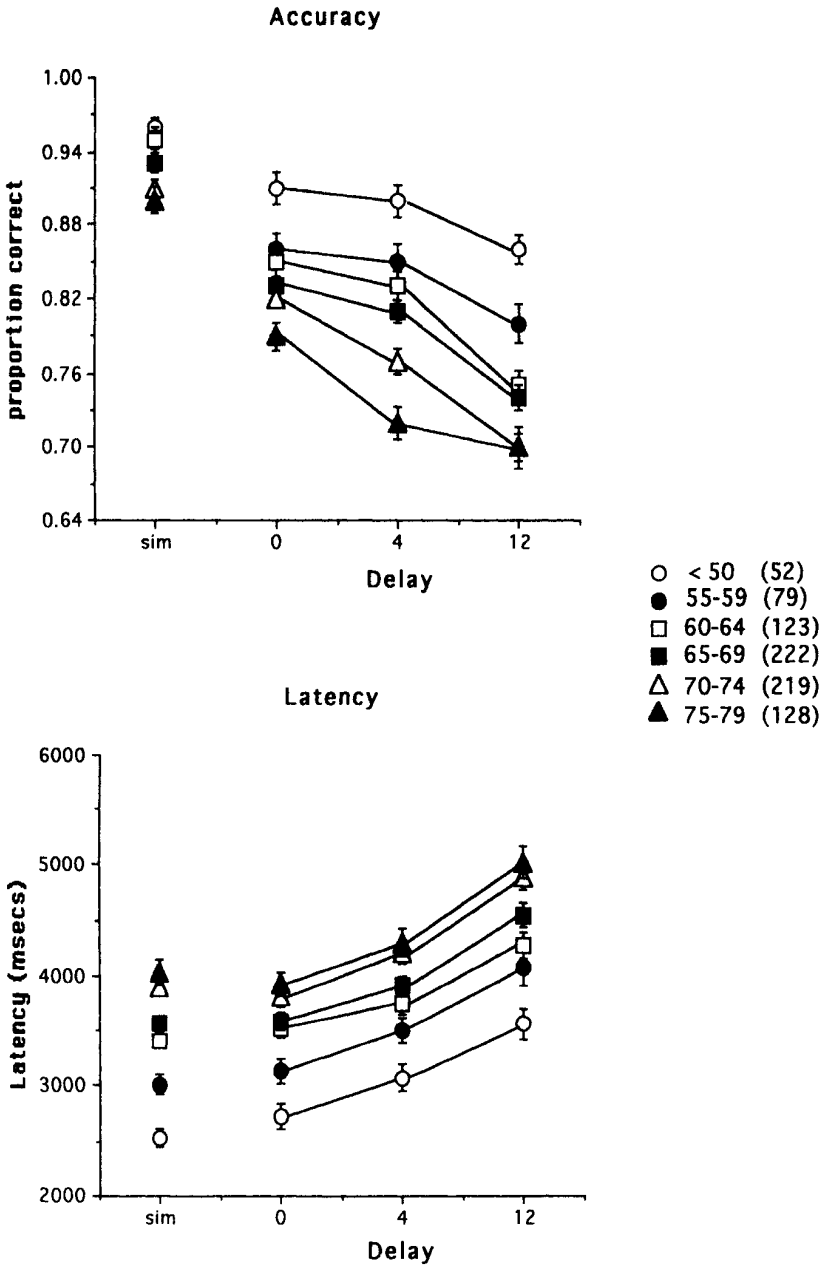


FIG. 10.3. Delayed matching to sample as a function of age in the large elderly sample studied by Robbins et al., 1994; with the addition of another group of similar NART IQ, <50 years. Mean+SEM of measures of both accuracy and latency of performance are depicted.

complexity of the effects is shown by consideration of the latency for correct responses measure, which shows a rather different pattern, in which there appears to be a general slowing over age, not always in concert with the percentage correct measure.

These results are complemented by performance on the CANTAB visual pattern recognition test, which is related to delayed matching to sample in its dependence on intact temporal lobe function (Owen et al., 1995), but is perceptually less complex. This may account for the similar performances in the young (< 50) and 55–59 year age groups. There is then a smooth, though gentle, decline in performance, with no statistical difference between the three eldest quincades, which is consistent with mild temporal lobe dysfunction (see Fig. 10.4), and with the apparent relative preservation of the inferior temporal gyrus in ageing (Brody, 1994). This smooth decline is to be contrasted with the discontinuities evident in spatial recognition memory performance in the same group. This form of recognition memory (unlike pattern recognition) has been shown to be impaired in patients with frontal lobe excisions (Owen et al., 1995). When different visual and spatial requirements are combined, as in the case of the visuospatial paired associate task, major differences are shown for the youngest group versus the rest, especially on the errors-to-criterion index (Fig. 10.5), suggesting that ageing differences arise when there is a need to coordinate different stimulus attributes (encoding in this case, "what" and "where"). Discontinuities in performance decline are again exhibited, as in the case for spatial recognition memory.

Spatial working memory performance is also especially sensitive to differences between the young (< 50 years) and the rest, but also to decline in the most elderly quincades (Fig. 10.6). It is difficult to read too much significance into the data for these latter two tests with respect to the neural loci of the performance decrements, as they have been shown to depend on distinct neocortical regions. Both are affected by damage to the temporal and frontal lobes, although probably for different reasons in each case. For example, use of a common, effective strategy for performing the self-ordered spatial working memory task is impaired by frontal, but not temporal lobe lesions (Owen et al., 1995). This has led to the hypothesis that a major part of the frontal deficit on this task is strategic, and therefore executive in nature, whereas the temporal lobe deficits are a consequence of impaired spatial working memory capacity. The use of strategy is greater in younger (<50 years) healthy volunteers, but remains rather constant over subsequent quincades, in contrast to the precipitous decline in spatial working memory performance shown in Fig. 10.6. Further evidence for the multiple component nature of the spatial working memory task can be gleaned from the psychometric analysis of task performance described below, which is consistent with the involvement of distinct cognitive components which depend on distinct neural systems.

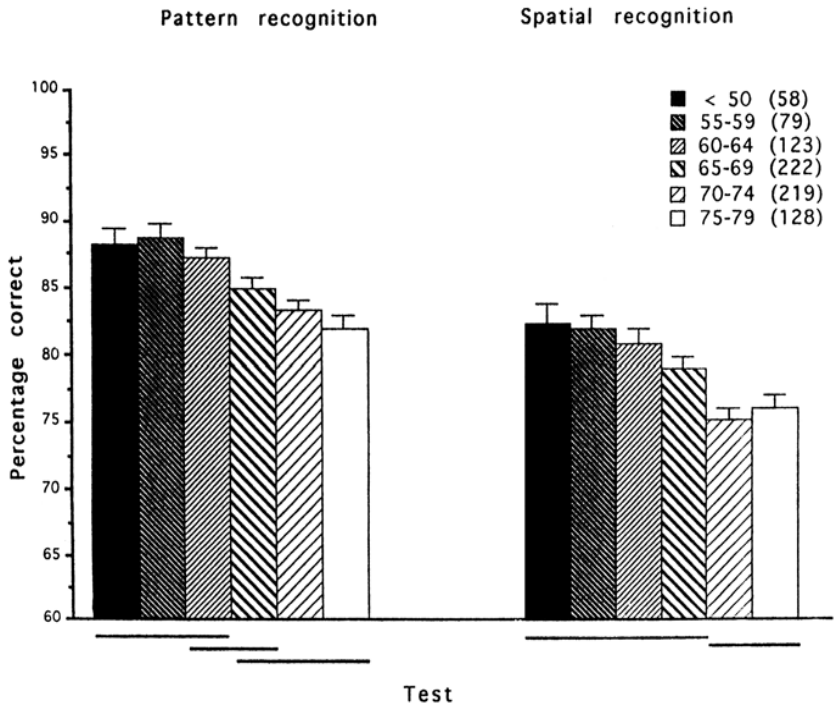


FIG. 10.4. Pattern and spatial recognition performance as a function of age (see Robbins et al., 1994). Non-overlapping horizontal lines indicate significant differences.

PRINCIPAL COMPONENT ANALYSES

This multivariate approach may help to avoid many of the pitfalls inherent in identifying single tests with specific brain regions, and also enables relationships between various tests to be identified or checked against existing evidence from brain-damaged patients or functional imaging studies. The initial analysis utilised data from nearly 800 participants collected on tests from the visual memory battery, the test of spatial working memory, and the visual search, matching to sample task. The results were quite clear-cut. Four factors were identified after varimax rotation of the original solution. These are depicted in Table 10.2, together with the loadings for particular variables from the tests. The solution specified independent factors that could be associated with visual learning or memory, speed of responding, the spatial working memory task (perhaps including its executive features), and perceptual function. The same solution was obtained after repeated ahalyses of random subsets of the larger data set, and, more importantly, for each five-year age group. The latter result suggests that the structure of cognitive abilities represented by this subset of CANTAB tests remains constant over age. As this statistical structure presumably reflects the coordinated functioning of distinct regions of the cerebral cortex, as well as its

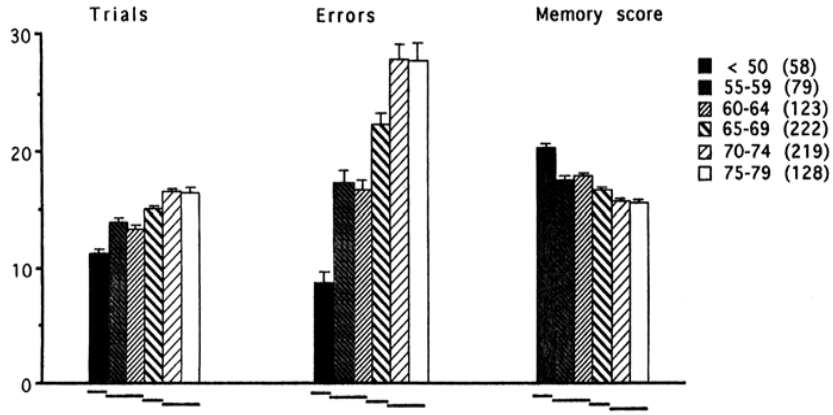


FIG. 10.5. Visuospatial paired associate learning according to three measures of number of (1) trials to criterion; (2) errors to criterion; (3) memory score on the first trial (see Robbins et al., 1994 for further details and caption to Fig. 10.4).

interactions with subcortical structures such as the striatum and its modulation by the ascending, chemically defined systems of the reticular core of the brain, the implication is that ageing degrades these systems in parallel. If any single cerebral region had been particularly susceptible to ageing effects, then the entire structure of interrelationships among the tests, whether probing functions mediated by the deficient structure, or alternatively not engaging it at all, would have been thrown out of balance. It is, however, possible that the diminishing influence of a single neurotransmitter system which innervates the entire cerebral cortex (e.g. the coeruleo-cortical noradrenergic system, or the basal forebrain cholinergic system) could account for such a pattern of results.

The analysis was also able to relate defining features of the population such as IQ and age to the critical cognitive factors. IQ was computed as a combined measure of five separate IQ tests (Nelson Adult Reading Test, Nelson, 1982; Mill Hill I and II, AH41 and AH42, Heim, 1968), and found to load most strongly on factor 1, visual learning and memory. (In fact, separate analyses of NART and AH4–1 and AH4–2 also show the same preferential loading on factor 1, unpublished findings.) By contrast, age loaded only on the cognitive speed factor. Such data would appear to violate hypotheses postulating that age results in progressive loss of IQ because of generalised reductions in cognitive processing speed (e.g. Salthouse, 1991). As those aspects of visual memory which load on factor 1 are particularly susceptible to deficits in early Alzheimer's disease (Sahakian et al., 1988; Sahgal, et al., 1991), the loading of age on factor 2 rather than factor 1 would also support the hypothesis that Alzheimer's disease is not simply an exaggeration of normal ageing processes.

Some caveats are necessary concerning these conclusions, however. The factor analysis pertains in the main to tests from the CANTAB battery; it would be more convincing to obtain converging evidence from an independent set of

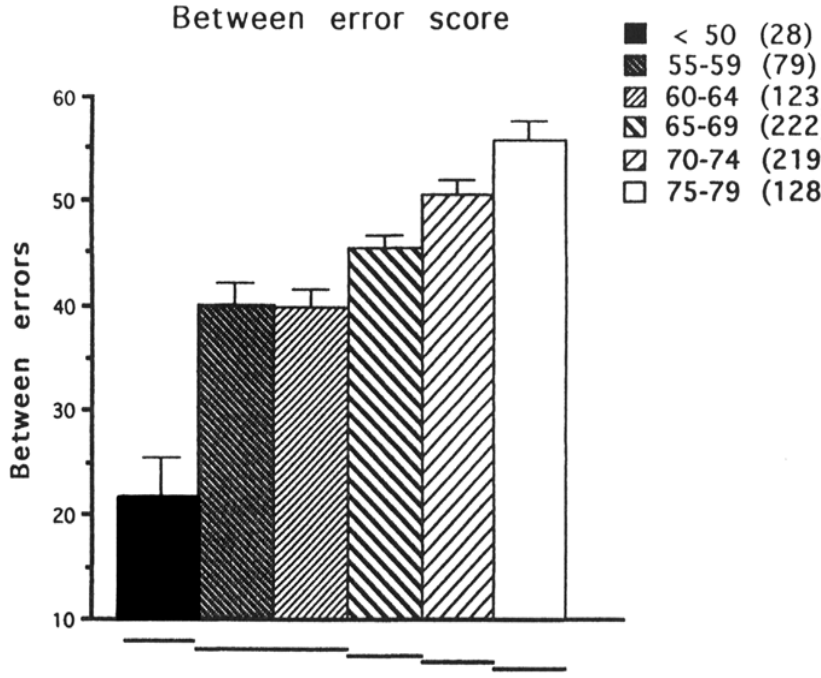


FIG. 10.6. Spatial working memory performance (mean+SEM between search errors), as a function of age (see Robbins et al., 1994 and also caption to Fig. 10.4).

neuropsychological tests. When such analyses have been feasible in the clinic in well-studied single cases, such converging evidence has been forthcoming (e.g Robbins et al., 1995).

It could also be argued that the CANTAB battery has not spanned a sufficient range of tests of cognitive function to be able fully to reflect integrated cortical functions. It is true, for example, that there are no verbal components to the CANTAB tests; however, such functions were reflected in some of the IQ tests administered. Another possible area of cognition covered insufficiently in the factor analysis was that tapped by tests of executive function, as performance on the spatial working memory test (the only contributor to loadings on factor 3 of Table 10.2), cannot simply be assumed to reflect its executive features. Converging evidence from other tests is required, which, it is presumed, would also be sensitive to frontal lobe dysfunction. Consequently, the final data to be described examine the structure of the CANTAB tests sensitive to frontal lobe dysfunction, the Tower of London test of planning, spatial working memory, and the extradimensional shift test (Fig. 10.2), together with their respective control components (such as spatial span in the case of the Tower task). In addition, we also provide an analysis of data in a subset of subjects given both

TABLE 10.2 Summary of Loadings for CANTAB Tests on Factors 1–4 Following Factor Analysis (N=787)

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*
Paired Associate Learning Trials	.82			
Memory score	.77			
Pattern Recognition	.73			
Spatial Recognition	.54			
DMTS (sim) Accuracy				.71
DMTS (del) Accuracy	.63			
DMTS Latency		.76		
MTS (Visual Search) Accuracy				.80
Latency		.72		
Spatial Working Memory (Between-search) errors			.88	
Spatial Working Memory (Within-search) errors			.77	

SOURCE: Robbins et al, 1994. Reproduced by kind permission of Karger Basel.

the CANTAB tests and a number of pencil and paper tests of executive function devised by Paul Burgess and Tim Shallice.

Table 10.3 shows the factor analysis of the tests of executive function described above, together with the participants' performance on the visual memory tasks, with a total *N*=201. There was a seven-factor solution, with eigenvalues>1.00, which accounted for about 67% of the variance. Factor 1, as previously, was associated with tests of visual memory and learning. Factor 4 had loadings of variables related to latency measures on different tests, similar to factor 2 of the original analysis, but now incorporating the initial thinking time measure from the Tower of London test of planning (i.e. the time taken to initiate solutions). This factor presumably relates to speed of information processing. Factor 3 of the current analysis corresponds to factor 3 from the original analysis (Table 10.2, with the larger *n*), with the addition of loading for a measure of strategy on this self-ordered test of memory. This index has been shown to correlate significantly with good performance on the task, and, as mentioned above, is sensitive to frontal lobe, but not temporal lobe, lesions. As this index clearly relates to executive functioning in so far as it measures how well subjects utilise a strategy that can demonstrably enhance their performance, it would be reasonable to assume that factor 3 does indeed represent aspects of executive function. Factors 6 and 7 apparently reflect further aspects of executive function; the former factor has loadings from the two variables derived from Tower of London performance that are the most sensitive to frontal lobe excisions in the study of Owen et al. (1990), minimum moves, and subsequent thinking time. The

former measure reflects the efficiency of problem solving, whereas the latter highlights the need for frontal patients to monitor carefully their solutions, even when these are accurate (see Owen et al., 1990). Each of the measures under factor 2 is also affected adversely by frontal lesions, including initial movement time, which forms a component of the yoked control test for the Tower of London and reflects continuous performance type functions. Overall, these functions may reflect working memory in the sense of maintaining information "online" for brief periods before responding. The loading for strategy score presumably results because the computation for this measure depends on the number of times a participant retraces a "route" at the beginning of successive searches; this then requires that the participant remember the last route employed. Overall, it is apparent that the tests of "executive function" do not all load on a single factor, as one might perhaps have anticipated. The allocation of certain tests such as the attentional shift task to a separate factor (factor 7) is also consistent with the heterogeneity of executive function, although this interpretation may be limited by the correspondence of this factor with performanc e on a single test. However, further inspection of the factor loadings for all of the variables entered revealed that errors on the matching to sample, visual search test from the Attentional battery also had a tendency to load on this factor (.39, thus below the threshold for being reported). This test also loaded in a limited way (.34) to factor 5, thus replicating the relationship between these variables shown in Table 10.2 that was identified as a possible visual perceptual factor. Introducing the data from the attentional set-shifting paradigm clearly had the effect of splitting the factor loadings between these two sources of covariance.

The third and final analysis is presented in Table 10.4. A large number of participants received some of the CANTAB tests, including the spatial working memory battery, and a battery of tests of frontal lobe function assembled by Burgess and Shallice. This analysis, conducted along the same lines as the previous ones, shows several features of interest. The first is that the Brixton test of spatial anticipation from their battery (Burgess & Shallice, 1996) loads together with performance from two of the tests of spatial memory from CANTAB that are sensitive to frontal lobe dysfunction. Therefore, this factor reflects performance from tests from different batteries that are sensitive to frontal lobe lesions, serving to cross-validate both, and again confounding the general belief that intercorrelations between tests of frontal or executive function are necessarily low. However, other well-known tests of frontal lobe function including verbal fluency and cognitive estimates clearly load separately from these. In fact, Table 10.4 shows that they tend to load with measures of IQ, including relatively verbal (NART, similarities and AH4−1) and non-verbal (AH 4−2 and digit symbol) components. While this reinforces in part the suspicion of Duncan (1995) that frontal lobe tests often generally reflect IQ, particularly of the "fluid" variety, the lack of loading of our main tests of executive function

TABLE 10.3 Summary of Main Loadings for CANTAB Tests on Factors 1–7 Following Factor Analysis (N=201)

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*	*Factor 7*
Paired Associate Learning							
Trials	−.83						
Memory score	.79						
Pattern Recognition	.73						
DMTS (del) Accuracy	.65						
Tower of London							
Initial movement time		.68					
Spatial Recognition		−.63					
Tower of London							
Subsequent movement time		.52					
Spatial Working Memory							
Strategy score		.45	.43		.44		
Spatial Working Memory							
Within-search errors			.84				
Between-search errors			.74				
Tower of London							
Initial thinking time				.79			
DMTS Latency				.69			
MTS (Visual Search) Latency				.62			
Spatial Span							
DMTS (sim) accuracy					.65		
Tower of London							
Subsequent thinking time						−.82	
Minimum moves						.74	
Attentional Shift							
Stage reached							−.89
Errors							.67

TABLE 10.4 Summary of Loadings on Factors 1–6 Following Factor Analysis (*N*=277)

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*
AH4-1	.82					
AH4-2	.74					
NART	.78					
Digit Symbol	.63					
Similarities	.57					
Verbal Fluency (F)	.71					
Cognitive Estimates	.53					
Age		.95				
Spatial Working Memory			.60			
Spatial Recognition			.76			
Brixton			.57			
MTS Latency (8)				.90		
Pattern Recognition					.79	
Paired Associate Learning Errors					.77	
Hayling Errors						.88

with these IQ measures suggests that not all tests of executive function that are sensitive to frontal lobe damage are subject to this interpretation.

EFFECTS OF AGEING ON COGNITIVE FUNCTION

The analyses presented in Tables 10.2 and 10.4 do not strongly support the hypothesis that functions sensitive to frontal lobe damage are especially sensitive to ageing effects. The extradimensional shifting test is sensitive early in the course of ageing, but not thereafter (see Fig. 10.2). However, even if a stronger relationship did exist, it would have to be determined whether the mode of failure really resembled the effects of frontal lobe damage (which are lack of control over perseverative responding) or, alternatively, of basal ganglia damage (which have more diffuse effects) (Owen et al., 1993b). The spatial working memory test does show a clear age-related deficit (Fig. 10.6), but the strategy measure does not. The Tower of London test measures (data not shown) tend to highlight initial thinking time, rather than measures more clearly related to frontal deficits, such as subsequent thinking time or minimum move solutions. In fact, if one considers the qualitative pattern of results shown in Table 10.1 across various patient groups, it is striking that it is basal ganglia disorders, as exemplified especially by Parkinson's disease, that most strongly reflect the cluster of variables most sensitive to ageing effects, rather than damage to temporal or frontal lobes, or dementia of the Alzheimer type. This interestingly mirrors the comparison that can be made between the deficits in Parkinson's

disease and following frontal lobe damage; generally the same types of test are sensitive or insensitive in both cases, but the deficits are qualitatively distinct. This is certainly the case for each of the three tests of frontal lobe dysfunction considered here (see Owen et al. 1992, for further discussion). This may be consistent with the anatomical relationships that exist between the frontal cortex and the striatum, in the form of "cortico-striatal functional loops" (Alexander, Delong, & Strick, 1988). The frontal cortex and striatum may thus contribute in related, but distinct, ways to performance of presumed "executive" tasks, which may serve to refine our notions about these complex processes.

While the possible analogy of ageing with Parkinson's disease is consistent with the preserved factor structure found for the factor analysis of the large sample (Table 10.2) as a function of age, it would be naive to suppose that the analogy is more than that. To begin with, different tiers of the substantia nigra are susceptible to ageing and Parkinson's disease (Scherman et al., 1989). Furthermore, Parkinson's disease is not associated with delay-dependent deficits in DMTS performance which clearly occurred in some of the aged quincades tested here (Fig. 10.3) and implicate impaired temporal lobe functions, which may be superimposed upon the more general picture we have described. We maintain that the development of yet more specific and sensitive tests of functioning of different neural systems will help disentangle these various forms of the cognitive sequelae of ageing.

ACKNOWLEDGEMENTS

We thank Dr J.Semple for his collaboration in these studies, and Professor Tim Shallice and Dr Paul Burgess for their willingness to allow us to analyse their data (Table 10.4). This research was supported by a Wellcome Trust programme grant to T.W.R., B.J.S., B.J.Everitt and A.C.Roberts and by a grant from the Nuffield Foundation to P.R., T.W.R. and B.J.S. The CANTAB tests are available from Cambridge Cognition, Flint Lane, Ely Road, Waterbeach, Cambridge CB5 96Z, U.K., fax (0)1223 441017.

REFERENCES

Alexander, G.E., Delong, M.R., & Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357– 381.

Arnsten A. & Goldman-Rakic, P.R. (1984). Catecholamines and cognitive decline in aged non-human primates. *Annals New York Academy of Sciences, 444,* 218–234.

Baker, S.C., Rogers, R.D., Owen, A.M., Frith, C.D., Dolan, R.J., Frackowiak, R.S.J., & Robbins, T.W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia, 34,* 515–526.

Bartus, R.T., Dean, R.L., & Beer, B. (1980). Memory deficits in aged cebus monkeys and facilitation with central cholinomimetics. *Neurobiology of Aging,* 1, 145–152.

Brody, H. (1994). Structural changes in the ageing brain. In J.R.M.Copeland, M.T.AbouSaleh, & D.Blazer (Eds.), *Principles and practice of geriatric psychiatry* (pp. 30–33). Chichester: Wiley.

Burgess, P.W. & Shallice, T. (1996). Bizarre responses, rule detection and frontal lobe lesions. *Cortex, 32,* 241–259.

Coull, J.T., Middleton, H.C., Robbins, T.W., & Sahakian, B.J. (1995a). Contrasting effects of clonidine and diazepam on tests of working memory and planning. *Psychopharmacology,* 120, 311–321.

Coull, J.T., Middleton, H.C., Robbins, T.W., & Sahakian, B.J. (1995b). Clonidine and diazepam have differential effects on tests of attention and learning. *Psychopharmacology, 120,* 322–332.

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

Downes, J.J., Roberts, A.C., Sahakian, B.J., Evenden, J.L., Morris, R.G., & Robbins, T.W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: Evidence for a specific attentional dysfunction. *Neuropsychologia, 27,* 1329–1343.

Duncan, J. (1995). Attention, intelligence and the frontal lobes. In M.Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Fischer, W., Chen, K.S., Gage, F.H., & Bjorklund, A. (1991). Progressive decline in spatial learning and integrity of forebrain cholinergic neurons during aging. *Neurobiology of Ageing, 13,* 9–23.

Flicker, C., Bartus, R.T., Crook, T., & Ferris, S.H. (1984). Effect of ageing and dementia upon recent visuospatial memory. *Neurobiology of Aging, 5,* 275–283.

Flood, D.G. & Coleman, P.D. (1988). Neuron numbers and sizes in aging brain: Comparisons of human, monkey, and rodent data. *Neurobiology of Ageing, 9,* 453–463.

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. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F.Plum (Ed.), *Handbook of physiology, the nervous system, higher functions of the brain, Vol 5, Part 1* (pp. 373–417). Bethesda, MD: American Physiological Society.

Heim, A. (1968). *The AH-4 IQ test.* Slough, UK: NFER-Nelson.

Hirai, S. (1968). Ageing of the substantia nigra. *Advances in Neurological Science, 12,* 845–849.

Jonides, J., Smith, E.E., Koeppe, R.A., Awh, E., Minoshima, S., & Mintun, M.A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363,* 623–625.

Lange, K.W., Robbins, T.W., Marsden, C.D., James, M., Owen, A.M., & Paul, G.M. (1992). L-Dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology, 107,* 394–404.

Mann, D.M.A., Yates, P.D., & Marcyniuk, B. (1984). Alzheimer's presenile dementia, senile dementia of the Alzheimer type and Down's syndrome in middle age form an age-related continuum of pathological changes. *Neuropathology and Applied Neurobiology, 10, 185–207.*

Marshall, J.F., Drew, M.C., & Neve, K.A. (1983). Recovery of function after mesotelencephalic dopaminergic injury in senescence. *Brain Research, 259,* 249–260.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27,* 272–277.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London, 298,* 85–95.

Muir, J.L., Everitt, B.J., & Robbins, T.W. (1995). Reversal of visual attentional dysfunction following lesions of the cholinergic basal forebrain by physostigmine and nicotine but not by the 5HT-3 antagonist, ondansetron. *Psychopharmacology, 118,* 82–92.

Murray, E. (1992). Medial temporal lobe structures contributing to recognition memory: The amygdaloid complex versus the rhinal cortex. In J.Aggleton (Ed.), *The amygdala* (pp. 453– 470) New York: Wiley-Liss.

Nagahara, A.H., Nicolle, M.M., & Gallagher, M. (1993). Alterations in [3-H]-kainate receptor binding in the hippocampal formation of aged Long Evans rats. *Hippocampus, 3,* 269–277.

Nelson, H.E. (1982). *National Adult Reading Test (NART): Test manual,* Windsor, UK: NFER-Nelson.

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). Extradimensional versus intradimensional 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., Summers, B.A., Marsden, C.D., Quinn, N.P., Lange, K., & Robbins, T.W. (1992). Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain, 115,* 1727–1751.

Owen, A.M., Beksinska, M. James, M., Leigh, P.N., Summers, B.A., Quinn, N.P., Sahakian, B.J., & Robbins, T.W. (1993a). Visuospatial memory deficits at different stages of Parkinson's disease. *Neuropsychologia, 31,* 627–644.

Owen, A.M., Roberts, A.C., Hodges, J.R., Summers, B.A., Polkey, C.E., & Robbins, T.W. (1993b). 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). Visuospatial short term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia, 33,* 1–24.

Owen, A.M., Evans, A.C., & Petrides, M. (1996a). 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.

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

Owen, A.M., Morris, R.G., Sahakian, B.J., Polkey, C.E., & Robbins, T.W. (1996c). Double dissociations of memory and executive functions in working memory task

following frontal lobe excision, temporal lobe excisions or amygdalo-hippocampectomy in man. *Brain, 119,* 1597–1615.

Parashos, I.A., & Coffey, C.E. (1994). Anatomy of the ageing brain. In J.R.M.Copeland, M.T.Abou-Saleh, & D.Blazer (Eds.), *Principles and practice of geriatric psychiatry* (pp. 36– 50). Chichester: Wiley.

Park, S.B., Coull, J.T., McShane, R.H., Young, A.H., Sahakian, B.J., Robbins, T.W., & Cowen, P.J. (1994). Tryptophan depletion in normal volunteers produces selective impairments in learning and memory. *Neuropharmacology, 33,* 575–588.

Petit, T.L. (1982). Neuroanatomical and clinical neuropsychological changes in aging and dementia. In F.I.M.Craik & S.Trehub (Eds.), *Aging and cognitive processes* (pp. 1–21). New York: Plenum Press.

Petrides, M. (1994). Frontal lobes and working memory: Evidence from investigations of the effects of cortical excisions in non-human primates. In F.Boller & J.Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 9, pp. 59–82). Amsterdam: Elsevier.

Rabbitt, P. (1993). Does it all go together when it goes? The nineteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 46A,* 385–434.

Robbins, T.W., Muir, J.L., Killcross, A.S., & Pretsell, D.S. (1993). Methods for assessing attention and stimulus control in the rat. In A.Sahgal (Ed.), *Behavioural neuroscience: A practical approach, Vol. 1* (pp. 13–47). Oxford: IRL Press at Oxford University Press.

Robbins, T.W., James, M., Lange, K.L., Owen, A.M., Quinn, N.P. & Marsden, C.D. (1992). Cognitive performance in multiple system atrophy. *Brain, 115,* 271–291.

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., Shallice, T., Burgess, P.W., James, M., Rogers, R.D., Warburton, E.C., & Wise, R.S.J. (1995). Selective impairments in self-ordered memory in a patient with a unilateral striatal lesion. *Neurocase, 1,* 217–230.

Sahakian, B.J., Morris, R.G., Evenden, J.L., Heald, A., Levy, R., Philpot, M., & Robbins, T.W. (1988). A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain, 111,* 695–718.

Sahakian, B.J., Downes, J.J., Roberts, A.C., Philpot, M., Levy, R., & Robbins, T.W. (1991). Preserved attentional function and impaired mnemonic function in dementia of the Alzheimer type. *Neuropsychologia, 28,* 1197–1213.

Sahakian, B.J., Owen, A.M., Morant, N.J., Eagger, S.A., Boddington, S., Crayton, L., Crockford, H.A., Crooks, M., Hill, K., & Levy, R. (1993). Further analysis of the assessment of attentional and mnemonic function using CANTAB. *Psychopharmacology, 110,* 395–401.

Sahgal, A., Sahakian, B.J., Robbins, T.W., Wray, C., Lloyd, S., Cook, J.H., McKeith, I.G., Disley, J., Eagger, S., Boddington, S., & Edwardson, J.A. (1991). Detection of visual memory and learning deficits in Alzheimer's disease using the Cambridge Neuropsychological Test Automated Battery (CANTAB), *Dementia, 2,* 150–158.

Sahgal, A., Lloyd, S., Wray, C., Galloway, P., Robbins, T.W., Sahakian, B.J., McKeith, I.G., Cook, J.H., Disley, J.C.A., & Edwardson, J.A. (1992). Does visuospatial memory in Alzheimer's disease depend on the severity of the disorder? *International Journal of Geriatric Psychiatry, 7,* 427–436.

Salthouse, T.A. (1985). *A theory of cognitive aging.* Amsterdam: North Holland.

Salthouse, T.A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Scherman, D., Desnos, C., Darden, F., Pollak, P., Javoy-Agid, F., & Agid, Y. (1989). Striatal dopamine deficiency in Parkinson's disease: Role of aging. *Annals of Neurology, 26,* 551– 557.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B., 298,* 199–209.

Shaw, T.G., Mortel, K.F., Sterling Meyer, J., Rogers, R.C., Hardenberg, J., & Cutaia, M.M. (1984). Cerebral blood flow changes in benign aging and cardiovascular disease. *Neurology, 34,* 855–862.

Smith, M.L., Deadwyler, S.A., & Booze, R.M. (1993). 3-D reconstruction of the cholinergic basal forebrain system in young and aged rats. *Neurobiology of Aging, 14,* 389–392.

Solomon, P.R., Beal, M.F., & Pendlebury, W.W. (1988). Age-related disruption of classical conditioning: A model systems approach to age-related memory disorder. *Neurobiology of Aging, 9,* 935–946.

Stroesner-Johnson, H.M., Rapp, P.R., & Amaral, D.G. (1992). Cholinergic cell loss and hypertrophy in the medial septal nucleus of the behaviourally changed aging rhesus monkey. *Journal of Neuroscience, 12,* 1936–1944.

Vijayshankar, N. & Brody, H. (1979). A quantitative study of the pigmented neurons in the nuclei locus coeruleus and subcoeruleus in man as related to ageing. *Journal of Neuropathology and Experimental Neurology, 38,* 480–497.

West, R.L. (1996). An application of prefrontal cortex function theory to cognitive ageing. *Psychological Bulletin, 120,* 272–292.

Whelihan, W.M. & Lesher, E.L. (1985). Neuropsychological changes in frontal lobe functioning with aging. *Developmental Neuropsychology, 1,* 371–380.

Whitehouse, P.J. (1994). Neurochemistry of the ageing brain. InJ.R.M.Copeland, M.T. Abou-Saleh, & D.Blazer (Eds.), *Principles and practice of geriatric psychiatry* (pp. 61–63). Chichester: Wiley.

CHAPTER ELEVEN
Behavioural Assessment of the Dysexecutive Syndrome

Barbara A.Wilson, Jonathan J.Evans, Nick Alderman, Paul W. Burgess, Hazel Emslie*

PURPOSES OF ASSESSMENT

Assessments are carried out in order to answer questions. Consequently, the nature of the question determines the assessment procedure. So we would use different tests and procedures for testing out a theoretical model than we would for trying to predict the likelihood of successful return to work for a brain-injured patient. An example of the former is the work of Baddeley, Logie, Bressi, Della Sala, and Spinnler (1986), who were trying to support or refute their hypothesis that patients with Alzheimer's disease had a deficit in the central executive component of the working memory model (Baddeley & Hitch, 1974). An example of the latter is Shallice and Burgess's (1991) six elements test which requires participants to plan and organise their activity over a 15-minute period while following certain rules. Mayes (1986) discusses the different concerns of researchers and clinicians with regard to memory assessments but his views are equally relevant to assessments of executive functioning. When thinking about frontal lobe or executive functioning clinicians are likely to be concerned with answering such questions as:

1. Does this person have cognitive impairments that are typically associated with the Dysexecutive syndrome (DES)?
2. What specific kinds of tasks are particularly affected?

---

* Barbara A.Wilson, MRC-APU Rehabilitation Research Group, Addenbrookes Hospital, Cambridge, U.K.; Jonathan J.Evans, Oliver Zangwill Centre, Princess of Wales Hospital, Ely, Cambs, U.K.; Nick Alderman, The Kemsley Unit, St. Andrews Hospital, Northampton, U.K.; Paul W.Burgess, Department of Psychology, University College London, London, U.K.; Hazel Emslie, MRC-APU Rehabilitation Research Group, Addenbrookes Hospital, Cambridge, U.K.

3. How do the problems affect functioning in everyday life?
4. To what extent are the executive deficits related to other cognitive problems such as language, memory, and perception?
5. How does this person compare with others of the same age or with the same diagnosis?
7. How does the patient and the family cope?
8. Which difficulties cause the most distress?
9. Can the patient go back to work/college/school?
10. What kind of treatment should we offer?

In order to answer these questions we need a number of "tools," i.e. tests or assessment procedures. Standardised, objective tests will answer some questions well. They can tell us how this person compares with others, providing the norms are satisfactory. They can tell us whether the patient has particular difficulty with certain tasks, such as following rules and reduced verbal fluency.

However, standardised tests are not good at answering some of the questions listed above which are more concerned with the impact of deficits on everyday functioning. These questions require a more behavioural or functional approach such as direct observation, questionnaires (to relatives and staff—not the patient, who may have poor insight), and analogue or role-playing situations.

Standardised tests are poor at predicting real-life performance because they do not take into account such factors as motivation, personality, individual styles, family support, and previous reinforcement history. One way to enhance their relevance to everyday life is to combine information from these tests with information from the observational, self-report, and functional measures (see Hart & Hayden, 1986; Wilson, 1991; Wilson, Baddeley, & Cockburn, 1989, for a further discussion of these issues). In addition, one can enhance the ecological validity of standardised tests by making them more like real-life tasks or, in other words, we can devise tests that map directly on to everyday behaviours. The Functional Communication for Aphasic Adults (Holland, 1980), the Rivermead Behavioural Memory Test (Wilson, Cockburn, & Baddeley, 1985) and the Cognitive Competency Test (Wang & Ennis, 1986) are examples of such tests.

The remainder of this chapter will describe a new ecologically valid test designed to predict real-life problems arising from the DES.

The Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) is a test aimed at pre dicting everyday problems arising from the DES. The DES closely resembles what was once called the "frontal lobe syndrome," a term that "is used to refer to an amorphous, varied group of deficits resulting from diverse aetiologies, different locations, and variable extents of abnormalities (Stuss & Benson, 1984, p. 3). Baddeley and Wilson (1988) argued that specification of a syndrome in terms of localisation is unfortunate and potentially misleading. We do not classify memory, language, reading, or perceptual deficits in this way as it would be

inadequate and limit our understanding of the observed cognitive phenomena. Similarly, a functional definition seems more appropriate to the deficits arising from frontal lobe damage.

Although there is great variability in the extent and degree of impairment in patients with frontal lobe damage, certain features are highly characteristic. Rylander (1939) described them as "disturbed attention, increased distractibility, a difficulty in grasping the whole of a complicated state of affairs…well able to work along routine lines [but] cannot learn to master new types of tasks" (p. 22).

Shallice (1982) believes this pattern of deficits can be described as an impairment in attentional control. He explains this in terms of a model of attention whereby a supervisory system exerts an executive function. Baddeley (1986) believes Shallice's "supervisory system" is analogous to the central executive component of working memory (Baddeley & Hitch, 1974) and suggests the term dysexecutive syndrome as a functional characterisation of patients with this particular pattern of deficits. It is possible that emerging evidence for the dissociability of various executive functions will lead to the conclusion that the grouping of deficits together as a syndrome is inappropriate. Indeed, when Baddeley coined the term in 1986 he suggested that it would only be a "stop-gap" term. However, for the time being, the term offers a useful functional characterisation of a range of deficits which certainly frequently co-occur in brain-injured patients.

PROBLEMS WITH EXISTING TESTS

Patients with the DES are likely to be impulsive, distractible, have problems utilising feedback and behave inappropriately in social situations (Alderman & Ward, 1991). They are not always the easiest of patients to assess. Although the individual component skills of executive functioning, i.e. the building blocks, may be intact, what is impaired is the "ability to initiate their use, monitor their performance and use this information to adjust their behaviour" (Burgess & Alderman, 1990, p. 183). Most tests, however, concentrate on the building blocks or component skills.

Shallice and Burgess (1991) make the point that in most neuropsychological tests "the patient typically has a single explicit problem to tackle at any one time, the trials tend to be very short…task initiation is strongly prompted by the examiner and what constitutes successful trial completion is clearly characterised" (pp. 727–728).

Rarely, in neuropsychological tests, are patients required to organise or plan their behaviour over longer time periods or to set priorities in the face of two or more competing tasks. Yet it is these sorts of executive abilities that are a large component of many everyday activities.

Some of the most widely used tests of executive functioning such as the Wisconsin Card-Sort Test (Berg, 1948), the Stroop test (Stroop, 1935), and the Reitan Trail Making Test (Reitan, 1958) can be performed normally by some patients with obvious impairments of executive functioning (Eslinger & Damasio, 1985). Shallice and Burgess (1991) described three such people in some detail. The patients performed satisfactorily on tests thought to be sensitive to frontal lobe lesions yet were unable to act effectively on their own initiative. Shallice and Burgess developed two tasks which they believed would require adequate functioning of the supervisory system or, in Baddeley and Hitch's terminology, the central executive. The tests were (1) the six elements test, which requires participants to organise their activities in order to carry out six tasks in a limited time period and without breaking certain rules, and (2) the multiple errands test, which is carried out in a shopping centre and involves participants buying certain objects, finding out certain information, being in a particular place at a particular time, and following certain rules while doing these things.

Shallice and Burgess demonstrated that these tests were sensitive to the everyday dysexecutive problems experienced by three patients.

In the multiple errands test, for example, patients with the DES made more errors than controls and showed qualitatively different behaviours (such as walking out of a shop without paying). The multiple errands test has high ecological validity and captures the non-routine, problem solving, planning, organisation and initiative required for everyday functioning. However, it also requires considerable time, planning, and organisation on the part of the tester and, consequently, is not always practicable in clinical situations.

The six elements test seemed better suited to standardisation in clinical situations so we decided to modify the test in order to make it simpler for the more severely impaired and less intellectually able patients often seen by neuropsychologists. We also designed a number of other tests similar to real-life activities that would cause difficulty for some patients with the DES. This evolved into the BADS.

DESCRIPTION OF THE TESTS

Modified Six Elements Test. This task is a simplified version of the original Shallice and Burgess (1991) task and involves the patient being given instructions to do three tasks (dictation, simple arithmetic and picture naming) each of which has two sections. The patient has 10 minutes in which to do at least some of all six sections, but they are not allowed to do sections A and B of the same task one after the other. Once it is clear that the patients have understood the test instructions, they are left to do the test. Scores used for analysis are the number of tasks attempted, and the number of rule breaks made.

Zoo Map. Patients are given a map of a zoo and a set of instructions relating to places they have to visit and rules they must stick to (such as not using some paths twice). In the first part of the test, the order in which the visits should take place is not specified and thus patients are required to plan a route which enables them to visit all the places, without breaking any rules. In the second part, another identical map is provided, but the order in which places should be visited is specified. Thus the main difference between the tasks is the considerably greater planning element of the first task. Scoring is based on the number of errors made and the number of places visited in the correct sequence, and the time taken to plan and execute the plan is recorded.

Action Program. This is a novel problem-solving task based on a description of a task by Klosowska (1976). The patient is presented with a set of materials including a large round container, half full of water, with a removable lid which has a small hole in the top; a tall thin container with a cork loose in the bottom of the tube, a piece of shaped stiff wire which is not long enough to reach the cork, and a small cylindrical tube to which a top can be attached. The patients are required to remove the cork from the tall tube, making use of the props provided and without touching the main assembly with their hands (the wire is used to remove the lid from the large container, the top is attached to the small cylindrical tube, and this is used to take water from the large container and pour it into the tall tube in order to float the cork to the top). A score based on the number of steps (out of five) completed without prompting is recorded.

Key Search. This is an adaptation of a test from the Stanford Binet test. The patients are provided with a piece of paper with a large square drawn on it. They are asked to imagine the square is a large field and they must draw the path they would take to search this field in order to find some lost keys. A score is awarded according to a scoring criterion for rating the functionality of the search pattern adopted.

Rule Shift Cards Test. This test uses 20 cards selected from an ordinary pack of playing cards and examines the patients' ability to respond correctly to a rule and to shift from one rule to another. In the first part of the test they are asked to say "Yes" to a red card and "No" to a black card. This rule, typed on a card, is left in full view throughout to reduce memory constraints. Cards are laid one at a time. Time taken and the number of errors are recorded. Almost all patients find this part of the test very easy and very few people fail; nevertheless this part of the test is important for two reasons. First, it may boost morale for those people who struggle with other subtests. Second, and even more importantly, it establishes a "set" against which perseverative errors are more likely to occur in part 2. In part 2, the rule is changed. Patients are asked to respond "Yes" if the card laid is the same colour as the previously laid card and "No" if it is a different colour. Once again, the typed rule is left in full view and the measures are time and number of errors. Some patients revert to the previous rule when doing this task, so it is a

simple measure of ability to shift from one rule to another and to keep track of the current rule.

Temporal Judgment. This test examines the ability of patients to make judgments regarding time. Responses to four questions are recorded and scored. Examples include "How long do most dogs live for?" and "How long does it take a dentist to do a routine dental check-up?" These four questions were found in the pilot study to discriminate reliably between people with and without brain damage, and to predict executive functioning.

Dysexecutive (DEX) Questionnaire. This is a 20-item questionnaire, constructed in order to sample a range of problems commonly associated with the dysexecutive syndrome. Items include statements such as "I act without thinking, doing the first thing that comes to mind," "I have difficulty thinking ahead and planning for the future," and "I find it difficult to keep my mind on something and am easily distracted." Each item is scored on a five point (0–4) Likert scale (ranging from "never" to "very often"). One version is designed to be completed by the patient and another by a relative or carer. An overall impairment score (maximum impairment score is 80) is derived from totalling the 20 individual item scores. An "insight" score can also be derived for each person based upon the difference between the rating of the patient and their relative/carer.

PRELIMINARY RESULTS

A pilot study carried out in 1993 (Alderman, Evans, Burgess, & Wilson, 1993) with 33 controls and 36 brain-injured individuals suggested that (1) the BADS is a better predictor of everyday problems (as measured by the relatives'/carers' ratings on the questionnaire) than the modified WCST (Nelson, 1976); (2) the modified six elements test and the zoo map test are the most sensitive individual subtests in the battery; (3) using scores from all subtests is best for predicting impaired everyday functioning; and (4) double dissociations are not uncommon. For example, one of our DES patients had no difficulty with the action plan (practical problem-solving task) but experienced considerable difficulty with the plan of search (paper and pencil problem solving). Another patient showed the reverse of this pattern.

More recently, a larger scale validation study (Wilson et al., 1996) has been carried out with the BADS following the success of the initial pilot study. In this study a method of scoring the battery was devised based on the performance of 216 controls. In addition, 90 brain-injured people also completed the battery. Results confirmed that the BADS has good interrater and test—retest reliability and avoids many of the problems described by Shallice and Burgess (1991). The performance of the brain-injured group on all six subtests was significantly poorer than that of the normal controls.

The validation study also confirmed the BADS is a reliable predictor of everyday executive functioning. The overall "profile" score derived from the battery is highly correlated with relatives'/carers' ratings of executive functioning in the brain-injured group using the DEX questionnaire. A preliminary functional analysis of the questionnaire suggests it is possible to identify three factors from it, one relating to behaviour, one to cognition, and another to emotion. The overall BADS score is predictive of all three factors. Certain subtests are better predictors of different factors. For example, it would appear that the action program, rule shift cards and temporal judgment tests are good predictors of the behaviour factor, whereas the zoo map test is associated with the factor relating to cognition. The modified six elements test is associated with both these factors. Further, more sophisticated factor analysis work is underway at present.

The BADS has also been considered alongside several other variables to determine what the best predictors are of everyday functioning problems. These included two well-known frontal lobe tests, the modified Wisconsin Card-Sorting (Nelson, 1976) and cognitive estimates (Shallice & Evans, 1978) tests. Other variables considered were those of age, NART—R (Nelson & Willison, 1991) and WAIS—R (Wechsler, 1981) FSIQs. Stepwise regression was used to determine the best predictors of executive problems among the brain-injured group (as rated by relatives/carers on the DEX). The only variable to enter the equation was the BADS. Similarly, the best and only predictor of the behaviour, cognition, and emotion factors to emerge from three subsequent analyses was that of the BADS.

These results confirm the implications from the earlier pilot study that the BADS is a valid test of everyday executive skills. As a battery, it appears to avoid many of the problems described by Shallice and Burgess (1991) and consequently is a better predictor of these skills in comparison to a number of other measures.

Another group for whom assessment of executive functioning is relevant is people with a diagnosis of schizophrenia and, in a recent study, the performance of schizophrenic patients on the BADS was compared with that of a brain-injured group and a healthy control group (Evans, Chua, McKenna, & Wilson, 1997). In recent years there has been a growing interest in the neuropsychology of schizophrenia. Brain imaging studies (e.g. Berman, Zec, & Weinberger, 1986; Rubin et al., 1991; Weinberger, Berman, & Zec, 1986) have suggested there may be a specific frontal lobe deficit in schizophrenia and a number of studies have found schizophrenic patients to be impaired on tests such as the Wisconsin Card-Sort test (Beatty et al., 1993, 1994; Braff et al., 1991; Kolb & Whishaw, 1983) and verbal fluency (Kolb & Whishaw, 1983), tests traditionally considered to be sensitive to frontal lobe executive impairment. Frith (1992) proposed that three principal cognitive abnormalities could account for all the major signs and

symptoms of schizophrenia: a disorder of willed action, a disorder of self-monitoring, and a disorder in monitoring the intentions of others. He argues that a disorder of willed action can account for symptoms such as poverty of action, perseveration, and utilisation behaviour shown by schizophrenic patients and also by patients with frontal lobe lesions and can be equated with a failure at the level of the supervisory attentional system in the Norman and Shallice (1986) model of the control of action.

We were therefore interested to see how a group of people diagnosed as schizophrenic would perform on the BADS battery. The battery was administered to 31 schizophrenic patients, 35 brain-injured patients and 26 healthy controls. All three groups were matched for age and estimated premorbid IQ. Results showed that the performance of the schizophrenic group was very similar to that of the brain-injured group, with significant differences between both patient groups and the control group on all of the BADS tasks, apart from the zoo map task and the self-ratings on the dysexecutive questionnaire (i.e. many relatives of brain-injured and schizophrenic patients rate their relatives to have significant problems while the patients do not recognise the severity of the problems).

Many, but not all, people with schizophrenia suffer a general cognitive impairment, but if an executive impairment is fundamental to schizophrenic symptomatology (or at least some of it), then the executive impairment should still be present in many of the patients who do not show a more general cognitive decline. In this study, therefore, a subsample of the total group were selected on the basis that they showed limited discrepancy between an estimate of premorbid IQ and current IQ. When this group was compared with a matched subgroup from the healthy volunteer group, two tests (the key search and rule shift cards) no longer differentiated the groups, but three other tests (six elements, action plan, and the temporal judgment test), as well as the overall BADS summary profile score and the relatives' ratings on the dysexecutive questionnaire still showed significant differences between the groups, suggesting that at least some people with schizophrenia do experience a specific executive impairment which is over and above any generalised cognitive decline.

A further question examined in this study was the relationship between memory impairment and performance on the BADS. It is well documented now that many people with schizophrenia have a severe memory impairment and we were concerned to establish that this was not the explanation of the apparent executive impairment. We specifically looked at the six elements test (since it might be argued to have the greatest potential memory load of the BADS tasks) and found that there were no significant correlations between performance on the Rivermead Behavioural Memory Test and the six elements, indicating that the presence of a memory impairment is not sufficient to explain an impaired performance on the executive tests.

Finally, since we are ultimately interested in the ability of neuropsychological tests to predict problems in everyday living, we looked at the relationship between performance on the BADS tests and the ratings of severity of problem given by relatives on the dysexecutive questionnaire. For the schizophrenic group, the predictors of impaired everyday executive functioning among the BADS tests were one of the measures from the modified six elements test (number of tasks attempted) and the zoo map test, but, in contrast to the brain-injured group, the overall BADS profile score did not correlate significantly with carers' ratings (and Evans et al. discuss various potential explanations for this conflicting finding). Nevertheless, it would appear that the BADS battery can provide useful information about impairment in executive functioning in people with a diagnosis of sehizophrenia. Such information might be particularly useful in assessing and preparing patients for moves from institutionalised care into more independent living situations. Institutions, which have a lot of routine and formal structure, may impose very little demand for executive functioning on those who live in them. Problems with such executive functioning may therefore not be identified prior to moving into more independent living situations, where patients may be expected to take greater responsibility for planning and organising their time, including scheduling of activities of daily living and making more day-to-day decisions. The BADS provides a tool for picking up problems of this sort, and this may be particularly necessary for those patients who appear to be cognitively well preserved and therefore may function relatively well within the institutionalised environment, and may even perform well on IQ tests such as the WAIS—R, but nevertheless have a specific executive impairment.

CONCLUSION

A range of assessments is needed to provide a clear picture of a patient's neuropsychological functioning. We need to combine information from standardised tests with information from behavioural observations, interviews, self-report measures, and other assessment procedures. We describe the BADS, an ecologically valid test, which bridges laboratory-based and observational measures to provide an assessment of non-routine everyday problems arising from impairments in executive functioning. We hope this test will lead to improvements in the assessment of the DES and indirectly to better treatment and management of executive problems.

REFERENCES

Alderman, N. & Ward, A. (1991). Behavioural treatment of the dysexecutive syndrome: Reduction of repetitive speech using response cost and cognitive overlearning. *Neuropsychological Rehabilitation, 1,* 65–80.

Alderman, N., Evans, J., Burgess, P., & Wilson, B.A. (1993). Behavioural assessment of the Dysexecutive Syndrome. *Journal of Clinical and Experimental Neuropsychology, 15,* 69–70 (abstract).

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

Baddeley, A.D. & Hitch, G. (1974). Working memory. In G.H.Bower (Ed.), *The psychology of learning and motivation, Vol. 8* (pp. 47–89). New York: Academic Press.

Baddeley, A.D., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology, 38A,* 603–618.

Baddeley, A.D. & Wilson, B.A. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7,* 212–230.

Beatty, W.W., Jocic, Z., Monson, N., & Staton, D. (1993). Memory and frontal lobe dysfunction in schizophrenia and schizoaffective disorder. *The Journal of Nervous and Mental Disease, 181,* 448–453.

Beatty, W.W., Jocic, Z., Monson, N., & Katzung, V.M. (1994). Problem solving by schizophrenic and schizoaffective patients on the Wisconsin and California card sorting tests. *Neuropsychology, 8,* 49–54.

Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology, 39,* 15–22.

Berman, K.F., Zec, R.F., & Weinberger, D.R. (1986). Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: Role of neuroleptic treatment, attention and mental effort. *Archives of General Psychiatry, 43,* 126–135.

Braff, D.L., Heaton, R., Kuck, J., Cullum, M., Moranville, J., Grant, I., & Zisook, S. (1991). The generalised pattern of neuropsychological deficits in outpatients with chronic schizophrenia with heterogeneous Wisconsin Card Sorting Test results. *Archives of General Psychiatry, 48,* 891–898.

Burgess, P.W. & Alderman, N. (1990). Rehabilitation of dyscontrol syndromes following frontal lobe damage: A cognitive neuropsychological approach. In R.L1.Wood & I.Fussey (Eds.), *Cognitive rehabilitation in perspective* (pp. 183–203). Basingstoke: Taylor & Francis Ltd.

Eslinger, P.J. & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, Cleveland, 35,* 1731–1741.

Evans, J.J, Chua, S.E., McKenna, P.J., & Wilson, B.A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine, 27,* 635–646.

Frith, C.D. (1992). *The cognitive neuropsychology of schizophrenia.* Hove: Lawrence Erlbaum Associates Ltd.

Hart, T. & Hayden, M.E. (1986). The ecological validity of neuropsychological assessment and remediation. In B.Uzzell & Y.Gross (Eds.), *Clinical neuropsychology of intervention* (pp. 21–50). Boston: Martinus Nijhoff.

Holland, A.L. (1980). *Communicative abilities in daily living.* Baltimore, MD: University Park Press.

Klosowska, D. (1976). Relation between ability to program actions and location of brain damage. *Polish Psychological Bulletin, 7,* 245–255.

Kolb, B. & Whishaw, I.Q. (1983). Performance of schizophrenic patients on tests sensitive to left or right frontal, temporal or parietal function in neurological patients. *The Journal of Nervous and Mental Disease, 171,* 435–443.

Mayes, A.R. (1986). Learning and memory disorders and their assessment. *Neuropsychologia, 24,* 25–39.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313–324.

Nelson, H.E. & Willison, J. (1991). *National Adult Reading Test, Second Edition.* Windsor: NFER-Nelson.

Norman, D.A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R.J.Davidson, G.E.Schwartz, & D.Shapiro (Eds.), *Consciousness and self-regulation. Advances in Research and Theory, Vol.4* (pp. 1–18). New York: Plenum Press.

Reitan, R.M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8,* 271–276.

Rubin, P., Holm, S., Friberg, L., Videbech, P., Andersen, H.S., Bendsen, B.B., Stromso, N., Larsen, J.K., Lassen, N.A., & Hemmingsen, R. (1991). Altered modulation of prefrontal and subcortical brain activity in newly diagnosed schizophrenia and schizophreniform disorder. A regional cerebral blood flow study. *Archives of General Psychiatry, 48,* 987–995.

Rylander, G. (1939). Personality changes after operation on the frontal lobes. *Acta Psychiatrica Neurologica,* Supplement No. 30.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298,* 199–209.

Shallice, T. & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 14,* 294–303.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Stuss, D.T. & Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95,* 3–28.

Wang, P.L. & Ennis, K.E. (1986). Competency assessment in clinical populations: An introduction to the cognitive competency test. In B.P.Uzzell & Y.Gross (Eds.), *Clinical neuropsychology of intervention.* Boston: Martinus Nijhoff.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* New York: The Psychological Corporation.

Weinberger, D.R., Berman, K.F., & Zec, R.F. (1986). Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: Regional blood flow evidence. *Archives of General Psychiatry, 43,* 114–124.

Wilson, B.A. (1991). Theory, assessment and treatment in neuropsychological rehabilitation. *Neuropsychology, 5,* 281–291.

Wilson, B.A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. (1996). *Behavioural assessment of the dysexecutive syndrome.* Bury St Edmunds: Thames Valley Test Co.

Wilson, B.A., Baddeley, A.D., & Cockburn, J.M. (1989). How do old dogs learn new tricks: Teaching a technological skill to brain injured people. *Cortex, 25,* 115–119.

Wilson, B.A., Cockburn, J., & Baddeley, A.D. (1985). *The Rivermead Behavioural Memory Test.* Bury St Edmunds, Suffolk: Thames Valley Test Company.

Author Index

Subject Index