

---

## SYMPOSIUM

# The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology

---

PAUL W. BURGESS,<sup>1,2</sup> NICK ALDERMAN,<sup>3</sup> CATRIN FORBES,<sup>1,2</sup> ANGELA COSTELLO,<sup>5</sup>  
LAURE M-A.COATES,<sup>6</sup> DEIRDRE R. DAWSON,<sup>4,7</sup> NICOLE D. ANDERSON,<sup>4,8</sup> SAM J. GILBERT,<sup>1,2</sup>  
IROISE DUMONTHEIL,<sup>1,2</sup> AND SHELLEY CHANNON<sup>2</sup>

<sup>1</sup>Institute of Cognitive Neuroscience, University College London, London, UK

<sup>2</sup>Psychology Department, University College London, London, UK

<sup>3</sup>Kemsley National Brain Injury Rehabilitation Centre, St. Andrew's Hospital, Northampton, UK

<sup>4</sup>Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada.

<sup>5</sup>Neuroscience Unit, King's College Healthcare Trust, London, UK

<sup>6</sup>Institute of Psychiatry, University of London, London, UK

<sup>7</sup>Dept. of Occupational Therapy and Occupational Science; Graduate Dept. of Rehabilitation Sciences, University of Toronto, Canada

<sup>8</sup>Department of Psychology, University of Toronto, Toronto, Ontario, Canada

(RECEIVED May 9, 2005; FINAL REVISION August 18, 2005; ACCEPTED September 19, 2005)

## Abstract

This article considers the scientific process whereby new and better clinical tests of executive function might be developed, and what form they might take. We argue that many of the traditional tests of executive function most commonly in use (e.g., the Wisconsin Card Sorting Test; Stroop) are adaptations of procedures that emerged almost coincidentally from conceptual and experimental frameworks far removed from those currently in favour, and that the prolongation of their use has been encouraged by a sustained period of concentration on “construct-driven” experimentation in neuropsychology. This resulted from the special theoretical demands made by the field of executive function, but was not a necessary consequence, and may not even have been a useful one. Whilst useful, these tests may not therefore be optimal for their purpose. We consider as an alternative approach a function-led development programme which in principle could yield tasks better suited to the concerns of the clinician because of the transparency afforded by increased “representativeness” and “generalisability.” We further argue that the requirement of such a programme to represent the interaction between the individual and situational context might also provide useful constraints for purely experimental investigations. We provide an example of such a programme with reference to the Multiple Errands and Six Element tests. (*JINS*, 2006, *12*, 194–209.)

**Keywords:** Frontal lobes, Multiple Errands Test, Six Element Test, Wisconsin Card Sorting Test, Stroop, Assessment

## INTRODUCTION

This paper addresses the issue of the provenance of clinical tests of executive function, and how best to set about developing the next generation of clinical measures. We will argue that whilst traditional tests of executive function have

been remarkably useful, we are now at the stage in the development of the field where one could create bespoke tests specifically intended for clinical applications rather than adapting procedures emerging from purely experimental investigations, as has been almost exclusively the case until recently. We will further argue that whilst one should continue to consider as potentially useful these procedures which emerge almost coincidentally, a more planned and deliberate programme should consider a “function-led” approach as a particularly fruitful avenue. Such an approach

---

Reprint request to: Dr. Paul W. Burgess, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1E 6BT, UK. E-mail: p.burgess@psychol.ucl.ac.uk

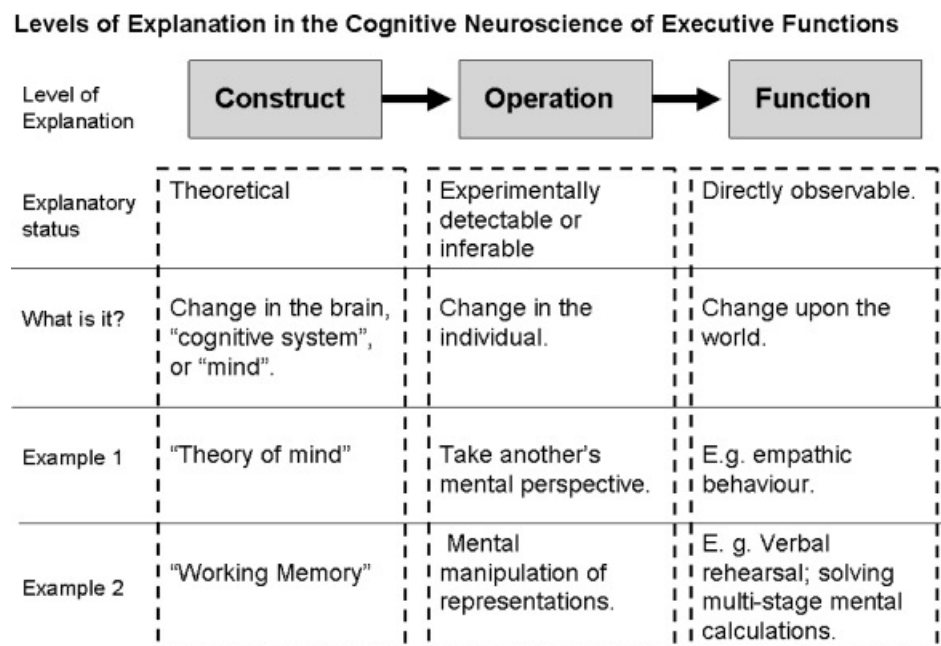
may have many scientific and clinical advantages, and address recent criticisms of existing assessment procedures (e.g., Manchester et al., 2004). But to make these points, it will first be necessary to define our terms:

**Argument 1.** *The distinction between construct, cognitive operation, and behavioural function: A taxonomy for discussion.*

The term *ecological validity*, first coined by Brunswik (1956, pp. 48–50; see Hammond, 1966 for review), was originally used to refer to the degree of relation between a proximal (e.g., retinal) cue and the distal (e.g., object) variable in perceptual experiments (see Hammond, 1998, for discussion). Now, ironically, it is most often used in reference to quite separate concerns of Brunswik's, those of "generalisability" and "representativeness." This misuse (from a historical perspective) is now so prevalent that we will, somewhat reluctantly, adopt the "incorrect" use of the term here. We will adapt the nomenclature of Kvavilashvili and Ellis (2004) for the current application by using "ecological validity" to refer both to the degree of "representativeness" of a task (the extent to which a clinical test corresponds in form and context to a situation encountered outside the laboratory), and the "generalisability" of test results (the degree to which poor performance on the test will be predictive of problems outside the laboratory; for recent examples of this use see Chaytor & Schmitter-Edgecombe, 2003; Clark et al., 2000; Coughlan et al., 2004; Farias et al., 2003; Gioia & Isquith, 2004; Higginson et al., 2000; Kibby et al., 1998; McDonald et al., 2004; Odhuba et al., 2005; Semkovska et al., 2004).

The foregoing argument will also require clear distinctions between the levels of explanation used in the cognitive neuroscience of executive function. So we will start with a very basic description of terms. We will make a distinction between *Constructs*, *Operations*, and *Functions* (see Figure 1). (Perhaps the most commonly used term, *Processes*, will not be used here since it is a term which has been used so flexibly in the literature.) *Construct* refers here (as is conventional) to a hypothetical cognitive resource whose existence is inferred from research findings (e.g., a correlation between two variables). Common examples of constructs are "Working Memory," and "General intelligence." These are ideas which motivate experiments, and are *interpretations* of data sets rather than phenomena that are part of them. In this respect, it is important to note, as Coltheart and Davies (2003) point out, that "The truth of an explanatory theory goes beyond the data and so is never logically guaranteed by the data" (p. 188).

*Operations* in the current context are the individual component steps of cognition which are not directly observable, but are inferred from a combination of task analysis and change in some dependent variable. In a different circumstance this could signify the involvement of some other operation. They may or may not be broken down into sub-components. Many *experimental* observations in the cognitive neuroscience or neuropsychology of executive function would fall into this category. For instance, increases in reaction time or errors, changes in cerebral blood flow and so forth, have all been held to be indicators (i.e., observable correlates of) of cognitive operations. It is important to note here that operations can be understood at the level of the individual rather than the individual's interaction with the environment, and that this is another level at which they



**Fig. 1.** Levels of explanation in the cognitive neuroscience of executive function.

differ from *functions*. The operation of single-word reading can happen in the mind without any outward behavioural sign (e.g., speech), but be accompanied perhaps by changes in blood flow in the reading areas of the brain. It is also important to note that operations are behavioural changes whose *utility* (rather than existence) can be made only in reference to outcome in the world. In one context, the absence of a behaviour can be a useful thing—for example, withholding a behaviour in a motor inhibition task—but in another may be evidence of dysfunction—say, inability to produce a word in the Hayling Sentence Completion Task part 2 (Burgess & Shallice, 1996). *Functions*, by contrast, are the directly observable behavioural output which is the product of a series of operations. They are usually understood in terms of a goal, instruction or intention to act, and it is therefore usually clear what constitutes success or failure. Simple examples might be reading, naming objects, describing a previously heard story; more complex behavioural examples might be preparing a meal, mailing a letter at a previously intended time, and so forth.

An example of the use of this kind of taxonomy appears in the excellent study by Owen et al. (1999). In the abstract to the study, the purpose of the study is described thus: “[this] study used positron emission tomography (PET) to demonstrate that working memory processes within the human mid-dorsolateral and mid-ventrolateral frontal regions are organized according to the type of processing required. . . . Two spatial working memory tasks were used which varied in the extent to which they required different executive processes. . . . During a ‘spatial span’ task that required the subject to hold a sequence of five previously remembered locations in working memory a significant change in blood-flow was observed in the right mid-ventrolateral frontal cortex. . . . By contrast, during a ‘2-back’ task that required the subject to continually update and manipulate an ongoing sequence of locations within working memory, significant blood flow increases were observed in both mid-ventrolateral and mid-dorsolateral frontal regions (Owen et al., 1999, p. 567). It is clear from this description (and the rest of the paper) that the operations of “holding a sequence of numbers” in mind, and continually updating and manipulating an ongoing sequence are considered potentially fractionable, but both fall within the general cognitive domain of “working memory” (a construct).

**Argument 2.** *The study of “central processes” makes special demands: From theory to practice in the cognitive neuroscience of executive function.*

Burgess (1997) argues that the field of executive function (EF) differs from most others in cognitive neuroscience in the differing emphasis on these levels of explanation. Specifically, the use of constructs in the field of executive function is much more prevalent. This is largely because of the “low process-behaviour correspondence” (pp. 85–89) inherent in the study of “central processes” compared with other fields which deal with more “routine processing” or “encap-

sulated resources” (Shallice, 1988; both terms assuming a particular organisation of the cognitive system). The abilities supported by the frontal lobes are thought to operate largely on the products of processing of other systems, and also output through other systems. Thus the operation of the “central executive” (Baddeley, 1996) or “supervisory attentional system” (Shallice, 1988) that the frontal lobes are thought to support is bound up with the operation of other systems to a degree that the nonexecutive components of, say, vision or hearing are not. Anatomical support, for instance, comes from the absence of direct pathways between prefrontal cortex and motor cortex (Leichnetz, 1986; Passingham, 1993, p. 125). And the clearest supporting demonstration is that, unlike with other cerebral regions, destruction of prefrontal cortex is not associated with loss of basic sensory or motor capacities or with obvious impairment of speech (Benton, 1991, p. 3). Linked to this is that at least some of the processing supported by prefrontal structures has been shown to be material-nonspecific (Burgess et al., 2001; 2003; Gilbert et al., 2005). In other words, it operates on input from a variety of sources, with a corresponding lack of dedicated process-output pathways.

This contrasts sharply with conceptions of fields of study dealing with behaviour deriving from the operation of abilities which are supposed to be more dedicated to it. As an example of an “encapsulated resources” theory, Burgess (1997) considers Bruce and Young’s (1986) model of face recognition (where the behaviour of seeing a face and identifying would be a *function*, and the theoretical individual contributory components to this behaviour—such as recognition—are the *operations*). And although more recently, the debate concerning how people identify faces has taken a more complex stance, contrasting these “domain-specific” accounts with “domain-general” ones, where visual recognition proceeds through the operation of resources which are specialised for a particular form of processing that can operate on a wide variety of classes of stimuli (Duchaine et al., 2003, 2004; Kanwisher, 2000), the processes are still assumed to be specialised for the processing of *visual* stimuli. This general principle could be extended to most other areas of enquiry within cognitive neuroscience, such as visual perception, reading, spoken language, motor control, audition, and so forth, where the predominant theories all hold that there are resources dedicated to that domain of cognition, if not the modality (see, e.g., MacSweeney et al., 2004, for an example within the auditory linguistic processing domain) or at least are agnostic as regards this possibility. By contrast, according to most conceptions, executive processes *theoretically* manifest themselves in a range of quite different situations, involving quite different stimuli, the only unifying feature of which might be the involvement of that process.

To illustrate the universality of this viewpoint, we will consider briefly the opinion of six leading theories of executive function. The most obvious example of a theory of the “central processing” kind is Duncan’s (e.g., Duncan, 2005; Duncan et al., 1995) contention that “executive processes”

are synonymous with Spearman's  $g$ , and his recent "adaptive coding" model is a clever theory extending this assertion by addressing the possible neural underpinning of  $g$  (e.g., Duncan & Miller, 2002). However, there are many other theories which share this domain-nonspecific aspect (see, e.g., Burgess & Alderman, 2004; Burgess & Robertson, 2002 for brief review). For instance the theory of Cohen and colleagues (e.g., Cohen & Servan-Schreiber, 1992; Cohen et al., 1990, 1998) holds that their single system may underpin both active memory and behavioural inhibition, determined by the current context, and they specifically state that their own theory is incomplete as regards the representation of many different information types, and this aspect is left as a lacuna in the model.

This aspect of "central processing" (i.e., domain- and material-insensitive) is also reflected in "working memory" theories of executive function. For example, Petrides (1994, 1998; Petrides & Pandya, 2002) maintains that the mid-ventrolateral PFC "subserves the expression within memory of various first-order executive processes, such as (the) active selection, comparison and judgement of stimuli," (p. 107) without specifying further what those stimuli might be. Goldman-Rakic's (1995) working memory account does however propose some basic domain- or material-specificity since the key aspect of it is that the various different frontal lobe regions all perform a similar role in working memory, but that each processes a different type of information. However, the account still holds that different forms of observable behaviour may result from dysfunction in one central processing resource: Working memory is defined as the ability to "hold an item of information 'in mind' for a short period of time and to update information from moment to moment," (Goldman-Rakic, 1998, p. 90) and it is argued that dysfunction of this system can cause a variety of deficits (see Goldman-Rakic & Leung, 2002, for further discussion.).

Even those theories that propose a series of potentially dissociable executive resources still maintain this centrality (e.g., Fuster 1997; 2002), but this characteristic is most obvious in those executive accounts which rest upon the concept of "attention." For example, Stuss et al. (1995) propose seven different attentional functions: sustaining, concentrating, sharing, suppressing, switching, preparing, and setting. But since there is no specification of the specific situations or material which require the differential use of these resources, the reader is invited to assume that they operate regardless of the form or source of the input source. Even the most complex model in this domain, that of Shallice (e.g., Norman & Shallice, 1986; Shallice, 1988, 2002; Shallice & Burgess, 1991a, 1991b, 1993, 1996; Shallice et al., 1989) still utilises the concept of attention, where the possibility of domain- or material-specificity is either denied or left as a lacuna. Most recently, the putative organisation of the SAS has been articulated in more detail (Shallice & Burgess, 1996/1998). In this model, the SAS plays a part in at least eight different processes: working memory, monitoring, rejection of schema, spontaneous schema generation, adoption of processing mode, goal set-

ting, delayed intention marker realisation, and episodic memory retrieval. However it is largely assumed that these central processes are independent of modality (of input or output). (For further evidence see also computational accounts, e.g., Dehaene et al., 1998, and representational accounts, e.g., Grafman, 2002.)

### **Argument 3. "Construct hegemony": Historical hangover?**

The field of executive function has been dominated by these construct-level, "macro" theories for the last 20 years in particular. By contrast there are relatively few accounts of the processing required to deal with specific situations which might tap these constructs. The extent to which this is the case may in fact in time become of interest to historians of science. For instance, one of the earliest and most compelling descriptions of cognitive deficit following frontal lobe damage was given by Penfield and Evans (1935; see also studies by Ackerly & Benton, 1947; Brickner, 1936). This seminal paper described a catastrophic impairment in the higher-order organisation of behaviour, manifesting itself for instance particularly sharply as an inability to coordinate the various activities involved in cooking a meal (see also Fortin et al., 2003). Given this observation, one might imagine that by now we would know quite a lot about the processes involved in cooking a meal, since Penfield and Evan's astute and influential observation in the context of a relative dearth (at that time) of empirical evidence would surely have suggested a "cardinal" situation for tapping the cognitive processes supported by the frontal lobes, and opened up a promising avenue for empirical studies. But this is not the case at all. We know virtually nothing about how the brain allows us to organise simple, everyday activities like cooking, and interacts with environmental factors in doing so (although there are a few honourable exceptions, see, e.g., Kirsh, 1995). Instead, EF research has spent several decades investigating the dynamics of (by comparison) esoteric activities such as performing the Wisconsin Card Sorting Test.

So how and why did EF research develop in this way? This may at least in part be a consequence of the history of research in the field of executive/frontal lobe function. The very existence of cognitive abilities supported by the frontal lobes was doubted for many years (Benton, 1991, p.19). Indeed, whilst early experimental physiologists were making sound progress in the fields of visual perception and language more than a hundred years ago, frontal lobe function was beset with null findings and failures to replicate effects across species. The scepticism this engendered was repeated in the observation of humans with frontal lobe damage: Even the facts of what was to become probably the most famous case of the "frontal lobe syndrome" (Harlow's [1848; 1868] Phineas Gage) were not believed by some at the time of its report. This attitude persisted well into the 1960s: For example, Teuber (1964), one of the leading authorities of the time, referred to the functions of the frontal lobes as a "riddle." In retrospect this lack of development in the area seems odd: The reports of the effects of

frontal lobe damage in humans were actually quite consistent, and there also many successful demonstrations in primates (e.g., Jacobsen, 1931; 1936). Moreover, the effects of damage to frontal lobe and related structures was considered predictable enough for frontal lobe psychosurgery to thrive (e.g., Malmo, 1948) during this period. The scepticism persisted, however, and one might argue that the tendency of investigators to assign ill-defined labels to the behaviours they saw (e.g., “perseveration”), rather than investigate the characteristics of the often myriad situations that provoked them, meant that research became quite conservative in its method (i.e., tending to use the same tasks and paradigms), whilst also—in the absence of constraining empirical evidence—rather ambitious in the scope of its theorising.

However, the past 25 years have seen rapid development in the cognitive neuroscience of executive functions, prompted in particular by technological advance (e.g., new methods of brain imaging, better methods of localising lesions). The repeated and sometimes unexpected findings of prefrontal involvement in even the simplest tasks has overthrown the old view of the frontal lobes as “functionally silent,” and now no leading theoretician of whom we are aware would maintain that the frontal lobes do not support “executive” (Pribram, 1973) abilities that are critical to human cognition.

Such a rapid development has however not been without cost, and it could be argued that the rate of innovation of method in the field of executive function has not matched the rate of technological advance. The majority of the early research papers using the new technologies (e.g., functional brain imaging) understandably leaned towards using the previously existing tools, principally (for example), because of the need to validate new technologies. This may have had the unintended effect of prolonging the use of these older tests and procedures both experimentally and clinically. Moreover, the rapid swing (in historical terms) from one extreme position (i.e., the frontal lobes are functionally silent, ergo “executive functions” don’t exist) to the other (i.e., the frontal lobes support executive processing which is involved in a very wide range of situations [see, e.g., Burgess et al., 2005b, 2006; Grady, 1999 for reviews]) has invited explanation at the construct level. Moving beyond this level is a challenge: The empirical findings in the field of executive functions of course support the newer view at the broadest level, but are still beset with complications. For example, Stuss et al. (2000, 2002) have shown that relatively small changes in task format can have marked differences in terms of demands upon executive abilities (see also Burgess, 1997, for review of methodological difficulties), and the sheer volume of recent findings presents its own conceptual difficulties. For example, in trying to understand the functions of the largest single cytoarchitectonic PFC region (Brodmann’s Area [BA] 10), findings of BA 10 PFC blood flow changes (see, e.g., Gilbert et al., in press; Grady, 1999) in a very wide range of types of tasks in imaging studies provides few constraints for theorising (see Burgess et al., 2005b for discussion).

The combination of these interrelated factors—historical inheritance, rapid conceptual change; numerous but often complex and unpredictable empirical findings—plus the practical and technical difficulties of investigations at the function level has, we would contend, invited accounts of executive function at the *construct* level, and for this reason *basic* scientific investigations (i.e., contrasted with applied or clinical investigations) have been, until very recently, overwhelmingly driven by this perspective.

This construct-level, or construct-to-operation-level theorising, combined with the experimental conservatism engendered by the historical research difficulties has led to research into executive abilities largely concentrating upon these two facets (i.e., examination of particular experimental, and clinical, paradigms on the one hand, and construct-level theorising of the type above on the other) whilst largely ignoring the issue of the translation of the hypothesised resources into the real world. The dynamic interplay between situational factors and the hypothesised resources (i.e., *function-level* analysis) has received very little attention. This has led to gaps in the links between the three explanatory levels which are required in order to link explanations at the basic science level with those required in clinical practice (i.e., my hypothetical resource can be termed Alpha; it supports cognitive operations X, Y and Z; and these resources are tapped by, or used to deal with, situations which have characteristics A, B, and C). Yet it is precisely at this level that clinical tests of executive function should in large part be targeted since in clinical practice the most important issue is usually performance outside the testing situation. How odd that one should not try to mimic (or at least understand) the demands of life outside the clinic when trying to determine the generalisability of test performance within it! Instead a common practice in clinical work and applied research has been to use performance on tests which emerged—often many years ago—from traditional experimental psychological investigations. The standpoint of these original studies has largely been to view the variability in contextual or environmental factors which naturally occur outside the lab as a confound (see Kvavilashvili & Ellis, 2004, for discussion of the long history of similar issues within the memory research domain). But it is exactly at this level (i.e., the interaction between the individual and their context: the “function” level) in which the clinician is interested. This dominance of the construct level investigation in neuroscience executive function research is reflected in (or probably reflects) the historical focus of the related field of cognitive science on the formation and manipulation of internal representations, rather than the ways in which agents interact with their environments in order to achieve goals (see, e.g., Clark, 1997; Hutchins, 1995).

#### **Argument 4. Time to consider a new approach?**

One can add to these criticisms that traditional tests were not developed to address the concerns of clinicians—that is, to measure the most clinically significant deficits and do

so in a way that makes the implications of the test score quite plain.

Consider for example, the history of, and current application of the currently most used and most thoroughly investigated single measure in the field of human frontal lobe function: the Wisconsin Card Sorting Test (WCST). This test was developed (in almost its current form) by Esta Berg in 1948, and slightly modified by Grant and Berg (1948). Both these studies involved only healthy participants, and the procedure was merely a variant of those used much earlier. In fact, the origin of the association between sorting tasks and frontal lobe function can be traced to Gelb and Goldstein (1920), who first developed sorting procedures for assessment of brain damage following the First World War. They adapted for clinical use the methods that had been used by Ach and his colleagues to investigate concept formation in healthy individuals. Following Gelb and Goldstein's discoveries, there were many investigations of sorting impairments in various patients groups, including those with brain-damage, dementia, learning difficulties, and schizophrenia carried out by groups in Germany and Russia (see Weigl, 1927; Goldstein & Scheerer, 1941 for review). The translation into English of a number of these papers during the 1935–45 period was followed by a rapid increase of interest in these methods in the U.S.A. Unfortunately, the reluctance of the American authors of the time to cite the original papers rather than their translations gives the appearance that these later papers were more contemporary than they were.

It is an important historical irony that by the time the WCST was developed, there were already tried and tested (by the standards of the time) sorting procedures available for clinical use. Yet it is the WCST, which was not developed for clinical use, that has now become dominant. This is largely attributable to the effect of a study by Milner (1963) who, more than 40 years since the original demonstration of sorting impairments, administered the Grant & Berg procedure to a group of neurological patients, with virtually the only modification being to give the test twice in succession.

Milner's principal achievement was to show that patients with dorsolateral prefrontal lesions performed worse on the test than those with either orbitofrontal or lesions outside the frontal lobes. However this finding has never to our knowledge been entirely replicated, either by Milner or by any other investigator, although there are many failures to replicate (see, e.g., Anderson et al., 1995; Corcoran & Upton, 1993; van den Broek et al., 1993). Psychometric studies have now shown that the WCST is a multifactorial test (e.g., Greve et al., 1997; Miyake et al., 2000). These results are congruent with those from recent functional imaging studies that show a network of frontal and nonfrontal brain regions to be involved in WCST performance (e.g., Marengo et al., 1993; Mentzel et al., 1998), and similar results from careful human lesion work (Stuss et al., 2000). Both methodologies stress how relatively minor changes in task format can have marked effect upon results.

Now consider in this context the typical uses of the WCST in clinical assessment. It is possible to give a rationale for understanding the three levels of WCST based on current common conceptions thus: WCST is a measure of “working memory” (*construct*), which supports “concept shifting” (*operation*), which enables correct choice of card sorting (*function*). However, the explanation at the level of construct is not very helpful in clinical assessment at the individual level, since such explanations go beyond the data and thus belong in the field of scientific hypothesis-testing rather than patient diagnosis. And the outline of the empirical findings above suggests explanation at the level of operation is not well-established (e.g., there may be many different reasons for failure, especially in different populations). Finally—and this is a particularly critical problem—the situation that performance of the WCST presents is so unlike everyday situations, that knowledge of performance in it is of very little help for assessment since it is of uncertain predictive validity (or “generalisability”): We do not really know what situations in everyday life require the abilities that the WCST measures.

### Further examples of the problem

This situation is largely repeated for most other popular measures used to investigate executive (i.e., “frontal lobe”) function. For example, the Stroop task was invented in 1935 (Stroop, 1935), but not applied in frontal lobe research until 1974 (Perret, 1974). Perret showed that patients with left frontal lobe damage were impaired in the interference condition of the task. Since that time the only partial replication of the result in humans to our knowledge was given recently by Stuss et al. (2001; “partial”, since Stuss et al. showed a left frontal deficit for colour naming not interference!), although the link between test performance and the frontal lobes now also seems secure on the basis of functional imaging results. It again seems likely from the results of the Stuss et al. study, and failures to replicate (see, e.g., Shallice, 1982) that the precise format of the task is an important factor in determining results. A similar picture emerges for the Tower of London test and its variants. This test was invented by Shallice and McCarthy (Shallice, 1982) as an application of AI theory in cognitive neuroscience. It has been generally considered to be a measure of “planning.” However, it is now doubtful that the reasons for task failure in frontal lobe patients could be characterised as a planning impairment (e.g., Goel & Grafman, 1995; Morris et al., 1997; see Burgess et al., 2005a for review). Moreover, the task is multicomponential (Carlin et al., 2000; Hodgson et al., 2000), psychometrically complex (Welsh et al., 1999, 2000) and the frontal lobe contribution to the task may be better characterised as mental imagery (Rowe et al., 2001) and goal conflict resolution (Morris et al., 1997) rather than planning-specific functions. Moreover, what the Tower of London measures may in any case have little relevance to “real-world” planning (Burgess et al., 2005a; Goel et al., 1997). For most of the established tests

of executive function, for example, Trail-Making, Cattell's Culture-Fair test, similar arguments can be made. In short, these procedures share two characteristics: First, they were not originally designed for the purpose to which they are currently put, and second, the dynamics of the tasks have been shown to be complex.

**Argument 5.** *Moving forward: What are the ideal requirements of a suitable task?*

However, it is a third characteristic that makes these points even more significant for this area: The situation that test performance in the clinic presents to the participant is quite unlike most situations which will be encountered outside it (i.e., they show poor "representativeness"). Experimental investigations of basic cognitive systems often use tasks that are simplified "models of the world" to make the problem scientifically tractable. For example, it is easy to define the circumstances where the processes underlying visual motion detection or colour vision operate. Similarly, those working in the field of motor control often use "models of the world" deliberately designed to have characteristics close to those encountered in everyday life (see, e.g., Haggard & Richardson, 1996; Ramnani et al., 2001). However it is not at all clear which "models of the world" are represented by tasks such as the WCST. Certainly current conceptions of them are far removed from their origins. Moreover, it is not at all clear that they are *simplified* models of the world at all. With the WCST, it is widely acknowledged that there is a minimum of three types of failure, which may interact in complex ways. Why has experimentation not favoured tasks that are more obviously like real-world situations?

**Argument 6:** *"Ecologically valid" tests can be psychometrically as sound as experimentally derived ones.*

We have already argued that the history of research in this area and the interconnected factor of construct-led hypothesising has contributed to a dearth of investigation at the function level. But a further contributing problem has probably been the widely held assumption that tasks very close to real world situations will be psychometrically unsound. Recently, this has been shown not to be the case (e.g., Burgess et al., 1998; Knight et al., 2002; Wilson et al., 1998). Indeed, ironically perhaps, tasks that are obviously "models of the world" (e.g., Alderman et al.'s [1996] Zoo-Map Test or Shallice & Burgess's [1991a] Multiple Errands task) generally return psychometric values higher than many traditional neuropsychological measures (e.g., free recall of short stories, or meaningless figures; Coughlan & Hollows, 1985; Knight et al., 2002). And even formalised "real-world" activities such as the Multiple Errands Test (MET) can show surprisingly good psychometric characteristics. The MET is at one extreme of the "representativeness" dimension, being a "real-life" multitasking test carried out in shopping precinct. Participants have to complete a number of tasks whilst following a set of rules (e.g., no shop should be

entered other than to buy something). The tasks vary in terms of complexity (e.g., buy a small brown loaf vs. discover a currency exchange rate), and there are a number of "hidden" problems in the tasks that have to be appreciated and the possible course of action evaluated (e.g., one item asks that participants write and send a postcard, yet they are given no pen). In this way, the task is, like many situations outside the laboratory, quite "open-ended" or "ill-structured" (Goel & Grafman, 2000). There are many possible courses of action, and it is up to the individual to choose one. Knight et al. (2002) report inter-rater reliabilities ranging between .81 to 1.00 for the various measures of the MET, and an internal consistency (Cronbach's alpha) of .77. Moreover, Dawson et al. (2005b) report inter-class correlation coefficients (measuring inter-rater reliability) on a hospital version of the Multiple Errands Test in 30 controls and 30 people with acquired brain injury of  $\geq 0.90$  ( $p \leq 0.01$ ).

Furthermore, and perhaps more important from a clinical standpoint, "real-world model" tasks may have particularly good "generalisability"—that is, that performance in the task situation is predictive of performance in situations outside it (e.g., Fortin et al., 2003). For example, Dawson et al. (2005a) have shown that Multiple Errands performance (using a hospital-based version) not only correlates well in CVA patients with self-report measures of everyday ability (the Sickness Impact Profile [Bergner et al., 1981]; the Mayo-Portland Adaptability Inventory [Malec & Lezak, 2003]), but also with objective assessment of daily living skills (using the Assessment of Motor and Process Skills [AMPS] battery, e.g., Fisher, 2003). Despite the relatively small group size (CVA patients,  $n = 11$ ; healthy controls  $n = 11$ ), Dawson et al. nevertheless showed correlations between MET and the AMPS-M scale which reached significance at  $p < .05$  ( $-.63$ ; the Process scale-MET correlation was  $-.51$ ). In a further study, Dawson et al. (2005b) administered the MET-HV to 30 people with acquired brain injury (and 30 matched controls) and found correlations between the AMPS and the Mayo-Portland scale of between .5 and .8, also suggesting good generalisability.

Critically, however, Knight et al. (2002) also report correlations between carers' ratings of everyday life problems and traditional tests of executive function (Cognitive Estimates [Shallice & Evans, 1978]; Modified Wisconsin Card-Sorting [MWCST; Nelson, 1976]; Tower of London Test [Shallice, 1982]; Verbal Fluency [Miller, 1984]) as well as a version of the Multiple Errands Test adapted for use in a typical hospital environment. Of the traditional executive tasks, only the MWCST perseverative errors and Cognitive Estimates tests showed any significant relation with carers' ratings of dysexecutive symptom clusters (using the DEX; Burgess et al., 1996a). The highest of these correlations was .47 ( $p = .038$ ; between Cognitive Estimates performance and problems with "intentionality," such as distractibility). However the correlation of MET-HV and ratings of intentionality problems was .70 ( $p = .001$ ). Thus these results add to the evidence that performance on more "ecologically valid" tasks can show closer concordance with observed

symptoms in everyday life than performance on traditional experimentally derived tasks (Burgess et al., 1998), and can be adapted to a clinical setting. Finally, performance on tests with high “representativeness” can be more predictive of performance on traditional tasks than are the traditional tasks themselves (Alderman et al., 2003; Burgess et al., 1998; Knight et al., 2002; Wilson et al., 1998).

**Argument 7:** *In clinical work, one cannot ignore the necessity of knowing how the abilities one is measuring translate into behaviour outside the laboratory. In most circumstances this should also be the case in experimental research.*

A final case for the rejection of the old procedures is that they require the introduction of an unnecessary extra stage in brain research. For instance, after 40 years of study of the WCST, and 20 years of Tower of London task study, we do know a little more about the dynamics of these tasks and what they measure. But it is perfectly possible that they measure a form of epiphenomenon which actually has very little significance for understanding the function of the organ under study. Using these tasks for proxies for measurement of real-world disability therefore requires that the operations involved in performing, say, the WCST must be established empirically as a second stage of investigation. This requirement is greatly diminished for measures which are self-evidently formalised models of real-life demands.

**Argument 8.** *The requirements of research paradigms and clinical tools differ.*

The foregoing argument should not be understood as a strong criticism of procedures such as the WCST in basic science research, where they have proved very useful in providing empirical evidence upon which theorising can be based. The problem is that the purpose of tests invented for use in either clinical or experimental contexts differ. Experimental tests usually seek to test a specific hypothesis about brain-behaviour relationships with the aim of elucidating the role of particular brain regions in cognition, or building information processing models of the cognition system that the brain supports. There is usually a conception of what the task measures (and how it is performed). These conceptions routinely do not stand the test of time: This is at the heart of scientific progress. However, the practicing clinician trying to assess and understand the nature of his/her clients' difficulties does not enjoy the same luxury of fallability without consequence. The root of this discrepancy is that scientific hypotheses may go beyond the data, but clinical diagnoses should not.

**Argument 9:** *Where might the next generation of tests come from?*

Considering these factors (and those outlined above), it is not perhaps surprising if current tests derived directly from experimental science are not ideal for clinical purposes.

Moreover, the origin of these tasks is perhaps more opportunistic than might be supposed, taking four main forms: (1) Animal work, where the leap to human clinical neuropsychological applications has, with a few arguable exceptions (e.g., Chorover & Cole, 1966) not been extensive. (2) Human work, where clinical observations of the effects of ablations have been turned into formalised tasks which, virtually unchanged, have then been used clinically (e.g., Wisconsin Card Sorting Test, Tower of London test; Verbal Fluency, etc.). (3) More recently, the observation of frontal lobe haemodynamic changes has then either used these tasks (or similar ones), or else studied a particular operation (e.g., “memory retrieval”) and incidentally found prefrontal involvement. However to our knowledge this research has not (yet) led to the development of a novel clinical neuropsychological test. (4) There has also been more conceptual research into a hypothetical construct (e.g., “working memory”) which is putatively strongly linked to prefrontal function; this work has yielded a few potentially useful clinical tasks (e.g., Baddeley et al., 1997).

Which of these investigative approaches is likely to provide new and better clinical neuropsychological tests of executive function? It is clear that (1) is unlikely to; this field has been in existence longer than any other, and yet yielded few widely used tasks (although the scientific advances it has yielded have been useful in less direct ways). Approach (3) may in time yield useful tasks, but it is unclear at the moment whether this work is necessarily relevant (e.g., there is no necessary reason for correspondence between localisation evidence across human lesion work and functional neuroimaging). This leaves (2) and (4). However, tasks derived from method (4) will need to be validated, as with every psychometric measure, against some criterion. What should this criterion be? Observer ratings? Perhaps. But more objectively, perhaps measurement of ability in the real-world. This will require the development of real-world-like validation tasks. But if we are going that far, why not consider first the use these real-world analogues as the assessment measure rather than their experimentally-derived proxies? A similar argument can be made for (2): If we are to develop tests which are specifically designed for clinical use (rather than adapt experimental paradigms of uncertain validity) surely the most obvious starting point is to consider function-level measurement of abilities expressed as competence in the real world.

**Argument 10:** *Function-led research is a highly promising avenue for the development of new clinical tests.*

Interestingly, Kingstone and colleagues, investigating attention in healthy people, have reached a similar position (e.g., Kingstone et al., 2003, 2005). They ask, “to what extent does the simple, impoverished, and highly artificial experimental task . . . have to do with the many complex, rich, real life experiences that people share?” (p. 344). They go on to argue that laboratory research (in the field of attention) is based on two critical assumptions, both of which



are problematic. The first is that the process of interest is “stable across different situations,” in other words that processes studied in the lab are assumed to be the same as the processes that are expressed in the real world. The second assumption is that one can maximise analytical power by minimizing the variability in a situation save for the factor being manipulated. In regard to the first assumption, Kingstone et al. (2005) suggest that it “eliminates any need or obligation by the scientist to confirm that the factors being manipulated and measured in the lab actually express themselves in the real world” (p. 346). As regards the second assumption, Kingstone et al. point out that “important characteristics of complex nonlinear systems such as the human attention system. . . are only revealed, or emerge, when several variables vary together in highly specific ways (see Ward, 2002; Weinberg, 1975)” (p. 348). (This latter point is especially germane to the following discussion about the abilities supported by rostral PFC processes, since one theory of the role of this region in human cognition suggests that it is involved when two or more other processes have to be coordinated: Ramnani & Owen, 2004.) Overall, Kingstone et al. (2005) conclude that their work is now oriented towards the goal of observing and describing behaviour as it occurs: an approach they term “cognitive ethology,” away from “making causal claims about fundamental processes.” In terms of the framework here, they are moving away from a construct-based approach to their research towards a function-driven one. If our analysis above is correct, their basic science programme is of the sort which would be most likely to yield paradigms which can be well adapted for clinical neuropsychological use.

## WHAT IS FUNCTION-LED RESEARCH? AN EXAMPLE AND CONCLUSION

Thus far, we have explored the position that traditional tests of executive function are unlikely (unless almost by coincidence) to be ideal for clinical purposes. And we have argued that function-led research is the most obvious candidate approach when considering where the new generation of clinical assessment tools might come from. But what does this approach actually entail? One example comes from our own lab, where we started investigations into how the brain supports multitasking (a function) by first considering patients who showed behavioural disorganisation in everyday life (Shallice & Burgess, 1991a; 1991b; see also Eslinger & Damasio, 1985). The first step was to develop a task that was as close as feasible to a formalised version of the “real-world” situations in which the patients had problems: the Multiple Errands Test described above (see Figure 2).

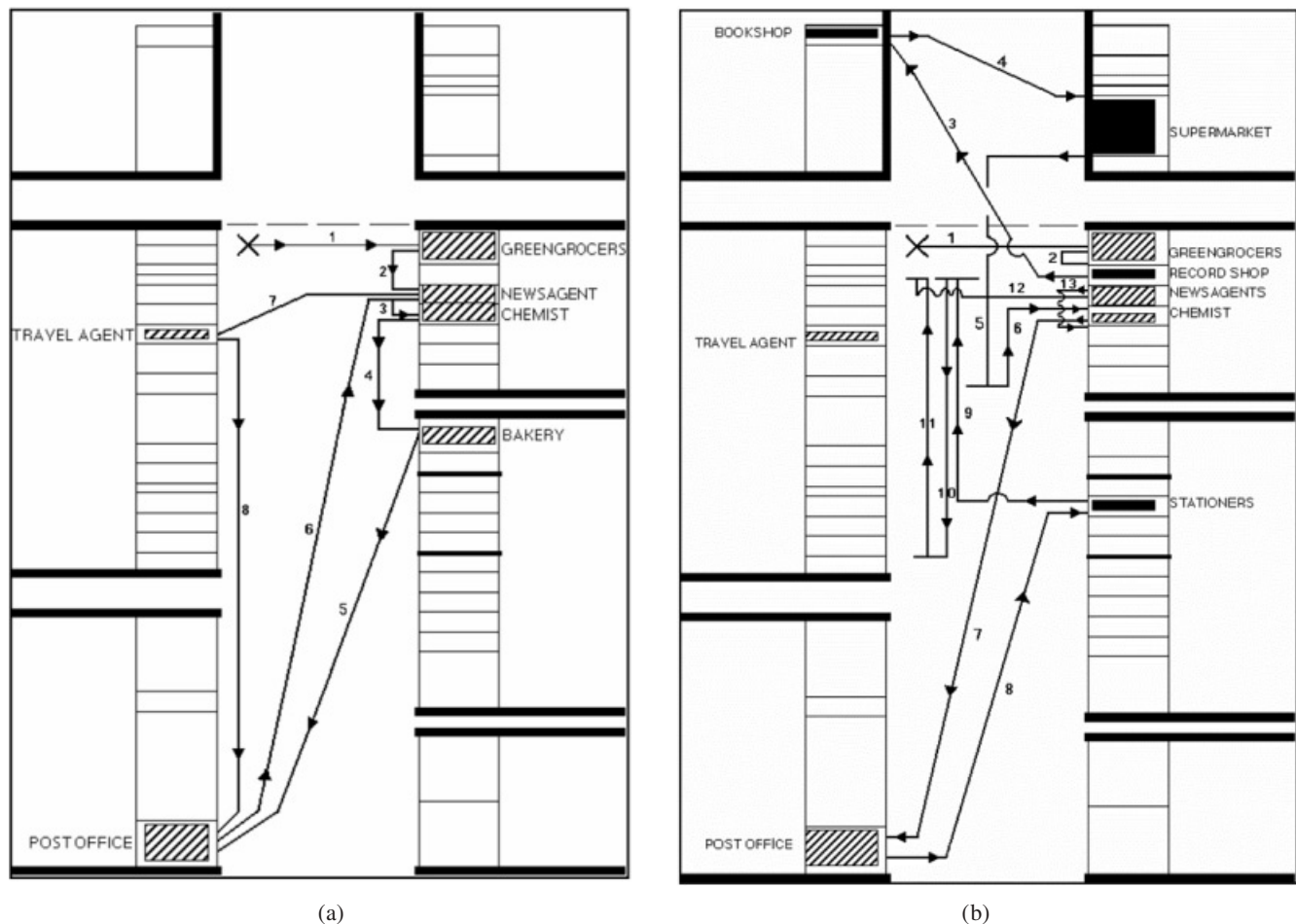
The second was to develop a lab-based “model of the world” which captured a putatively critical component (voluntary multiple delayed task-switching) of the “real-world analogue” task (Burgess, 2000a). Shallice and Burgess (1991a) showed that patients with frontal lobe damage could be markedly impaired on these tasks despite normal performance on traditional clinical tests of executive (and other)

abilities. A third step was to link this basic aspect of multitasking (voluntary delayed task switching) to brain structure. This was carried out by human group lesion studies of lab-based multitasking tests (the SET and the Greenwich Test) which have shown that relatively circumscribed multitasking deficits follow rostral PFC lesions (Burgess et al., 2000; Burgess et al., submitted: described in Burgess et al., 2005b). These findings concurred with those from studies of patients with similar problems (e.g., Bird et al., 2004; Burgess, 2000b; Goldstein et al., 1993; Levine et al., 1998; 1999; see also Figure 3).

Then, working up the levels of explanation from function to construct, the neural underpinnings of a key aspect of this sub-function (the operations involved in the realisation of a delayed intention, i.e., prospective memory) were investigated in a series of functional imaging experiments (Burgess et al., 2001; 2003; Okuda et al., 2004) which confirmed a role for rostral PFC (principally BA 10) in certain forms of prospective memory. Then on the grounds of both the lesion and imaging studies, we hypothesised a construct underlying these various operations: an attentional “gateway” that operates to alter the relative valence of stimulus-oriented or stimulus-independent thought in nonroutine circumstances, and tested the hypothesis of BA 10 involvement in a variety of indicators for such a construct in a series of neuroimaging experiments (Gilbert et al., 2005, 2006; Simons et al., 2005a, 2005b, in press). A further step would be to investigate the generalisability in neurological patients of the indicators for the construct developed through this theoretical stage, thus linking construct-level theorising with function (see Burgess et al., 2005b, 2006, for further details of this research programme).

This argument should not be mistaken as suggesting the sole use of tests which are carried out in the “real world” (i.e., outside the normal lab or examination room setting) such as the MET, or that such tests will *necessarily* prove more clinically useful than others, especially when practical concerns are taken into account. Instead the argument here is that one should consider in the design of *new* tasks the demands that the “real world” may make but that are not presented to the participant performing *existing* tests in a lab setting. By not considering these aspects, we may be missing decrements in many aspects of cognition which are critical to competence in everyday life. Some progress has been made in various areas of neuropsychology (e.g., Robertson et al., 1994; Wilson et al., 2003), but with few exceptions (e.g., Wilson et al., 1996) in the field of executive function, this is not the provenance of the most prevalently used tasks.

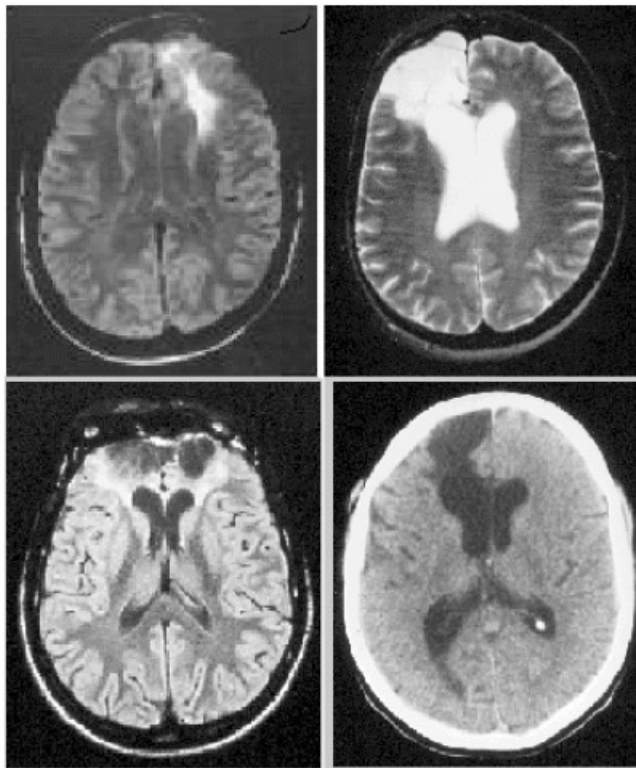
The general scientific approach to the neuroscience of executive function advocated here (i.e., analysing the demands made by real-world situations and then trying to mimic them in the lab) is certainly not unique, with similar programmes occurring recently in related areas such as social cognition (e.g. Channon et al., 2005; Frith & Frith, 2003). If our argument is correct, these kinds of programmes are the most likely sources for the next generation of “execu-



**Fig. 2.** Panel A shows a schematic representation of a typical control performance on Shallice and Burgess's (1991a) Multiple Errands Test. Panel B shows the impaired performance of a patient (DN) with frontal lobe damage who was matched to the control for intellectual functioning, and who also performed well on ten traditional clinical tests of executive function (see Shallice & Burgess 1991a for further details). Hatched shading indicates shops that need to be entered in order to complete the test. Solid block shading indicates shops that do not need to be entered, or are forbidden by the task rules. (Adapted from Burgess & Alderman, 2004.)

tive function" clinical tests. Certainly the one above yielded a clinical test (the Six Element Test) which is widely used, and also a useful task with high "representativeness" (the MET). There are now various versions of both the Multiple Errands (e.g., Knight et al., 2002) and the Six Element tests (e.g., Burgess et al., 1996b; Emslie et al., 2003; Kliegel et al., 2000; Levine et al., 2000; Manly et al., 2002), and they have been used in the investigation of a variety of conditions (e.g., traumatic brain injury, Hoclet et al., 2003; Levine et al., 2000; schizophrenia, Evans et al., 1997; depression, Channon & Green, 1999; ADHD and Oppositional Defiant/Conduct Disorder, Clark et al., 2000; drug abuse, Zakzanis & Young, 2001; sleep deprivation, Nilsson et al., 2005; normal ageing, Garden et al., 2001; "intensive care syndrome," Sukantarat et al., in press), and there is also information about the psychometric properties of the tests, (e.g., Alderman et al., 2003, Jelicic et al., 2001; Kafer & Hunter, 1997; Knight et al., 2002), their "generalisability" (Burgess et al., 1998), and their correlations with other executive function tests (Duncan et al., 1997).

Moreover, the study of differences in patterns of failure on the two tests are revealing patterns of impairment that are both theoretically interesting and clinically relevant (see Burgess et al., 2005b for further details). For instance, Alderman et al. (2003) found a double-dissociation between rule-breaking behaviour and failure to initiate tasks, and these two patterns were associated with different dysexecutive symptoms in everyday life. "Rule-breakers" tend to show problems with control aspects of memory (e.g., show confabulation, temporal sequencing problems, and repetitive behaviour). By contrast, people who fail to complete the set tasks tend to show negative symptoms, such as apathy and shallow affect. From a theoretical perspective, Burgess et al. (2005b) argue that these results may reflect the "real-world" consequences of a time/event dissociation in prospective memory. From a rehabilitation perspective the patterns suggest different treatment applications (Burgess & Robertson, 2002) in a much more direct way than would, say, a dissociation between perseverative behaviour and problems with concept shifting in performance on the WCST.



**Fig. 3.** Brain scans of four neurological patients with rostral prefrontal damage. All patients achieved superior scores on IQ tests, and all achieved excellent scores on traditional executive tasks such as the Wisconsin Card Sorting and Verbal Fluency Tests. However they all showed significant behavioral organization problems in everyday life. Top left: PF (Goel & Grafman, 2000); Top right: GN (Goldstein et al., 1993); Bottom left, right: AP and FS respectively (Shallice & Burgess, 1991a). For ease of understanding, the scans have been transposed so that left on the scan represents the left hemisphere. Adapted from Burgess et al. (2005b).

It is currently too early to tell whether the further theoretical investigations into the constructs underpinning these functions will yield lab-type tasks which are of clinical utility, although this seems possible given the experiences of related fields (e.g., Channon et al., 2005). However it seems likely that going from function to construct presents useful constraints upon theorising which are less apparent in going from construct to function. This approach may therefore be

**Table 1.** Data taken from Alderman et al. (2003), showing a group level dissociation between neurological patients whose failures on the MET-SV were predominately rule-breaks, and those who predominately failed to achieve tasks. The total patient sample was  $n = 50$ , so these two groups together represent 90% of all patients tested.

	Rule Breaks	Task Failures
Rule Breakers ( $n = 19$ )	224	65
Task Failers ( $n = 26$ )	78	171

useful as a starting point for both the development of new and better clinical tests of executive function, and also for basic cognitive neuroscience investigations.

## ACKNOWLEDGMENT

This work was supported by Wellcome Trust grant number 061171 to PWB.

## REFERENCES

- Ackerly, S.S. & Benton, A.L. (1947). Report of a case of bilateral frontal lobe defect. *Research Publications: Association for Research in Nervous and Mental Disease*, 27, 479–504.
- Alderman, N., Burgess, P.W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, 9, 31–44.
- Alderman, N., Evans, J.J., Emslie, H., Wilson, B.W., & Burgess, P.W. (1996). Zoo Map Test. 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.
- Anderson, C.V., Bigler, E.D., & Blatter, D.D. (1995). Frontal lobe lesions, diffuse damage, and neuropsychological functioning in traumatic brain-injured patients. *Journal of Clinical and Experimental Neuropsychology*, 17, 900–908.
- Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A(1), 5–28.
- Baddeley, A., Della Salla, S., Gray, C., Papagno, C., & Spinnler, H. (1997). Testing central executive functioning with a pencil-and-paper test. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 61–80). Hove, U.K.: Psychology 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. 3–34). New York: Oxford University Press.
- Bergner, M., Bobbitt, R.A., Carter, W.B., & Gilson, B.S. (1981). The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care*, 19, 787–805.
- Bird, C.M., Castelli, F., Malik, O., Frith, U., & Husain, M. (2004). The impact of extensive medial frontal lobe damage on “Theory of Mind” and cognition. *Brain*, 127, 914–928.
- Brickner, R.M. (1936). *The intellectual functions of the frontal lobes: A study based upon observation of a man after partial bilateral frontal lobectomy*. New York: Macmillan.
- Bruce, V. & Young, A.W. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305–327.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.
- Burgess, P.W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, U.K.: Taylor and Francis.
- Burgess, P.W. (2000a). Real-world multitasking from a cognitive neuroscience perspective. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 465–472). Cambridge, MA: MIT Press.
- Burgess, P.W. (2000b). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research*, 63, 279–288.
- Burgess, P.W. & Alderman, N. (2004). Executive dysfunction. In

- L.H. Goldstein & J.E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (pp. 185–210). Chichester, U.K.: John Wiley.
- Burgess, P.W., Alderman, N., Emslie, H., Evans, J.J., & Wilson, B.A. (1996a). The dysexecutive questionnaire. In B.A. Wilson, N. Alderman, P.W. Burgess, H. Emslie, & J.J. Evans (Eds.), *Behavioural assessment of the dysexecutive syndrome*. Bury St. Edmunds, U.K.: Thames Valley Test Company.
- Burgess, P.W., Alderman, N., Emslie, H., Evans, J.J., Wilson, B.A., & Shallice, T. (1996b). The simplified six element test. In B.A. Wilson, N. Alderman, P.W. Burgess, H. Emslie, & J.J. Evans (Eds.), *Behavioural assessment of the dysexecutive syndrome*. Bury St. Edmunds, U.K.: Thames Valley Test Company.
- Burgess, P.W., Alderman, N., Evans, J., Emslie, H., & Wilson, B.A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, 4, 547–558.
- Burgess, P.W., Gilbert, S.J., Okuda, J., & Simons, J.S. (2006). Rostral prefrontal brain regions (Area 10): A gateway between inner thought and the external world? In W. Prinz & N. Sebanz (Eds.), *Disorders of volition* (pp. 373–396). Cambridge, MA: MIT Press.
- Burgess, P.W., Quayle, A., & Frith, C.D. (2001). Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia*, 39, 545–555.
- Burgess, P.W. & Robertson, I.H. (2002). Principles of the rehabilitation of frontal lobe function. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 557–572). New York: Oxford University Press.
- Burgess, P.W., Scott, S.K., & Frith, C.D. (2003). The role of the rostral frontal cortex (area 10) in prospective memory: A lateral versus medial dissociation. *Neuropsychologia*, 41, 906–918.
- Burgess, P.W. & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, 34, 263–273.
- Burgess, P.W., Simons, J.S., Coates, L.M.-A., & Channon, S. (2005a). The search for specific planning processes. In G. Ward & R. Morris (Eds.), *The cognitive psychology of planning* (pp. 199–227). Hove, U.K.: Psychology Press.
- Burgess, P.W., Simons, J.S., Dumontheil, I., & Gilbert, S.J. (2005b). The gateway hypothesis of rostral PFC function. In J. Duncan, P. McLeod, & L. Phillips (Eds.), *Measuring the mind: Speed, control, & age* (pp. 215–246). Oxford University Press.
- Burgess, P.W., Veitch, E., Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, 38, 848–863.
- Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia*, 38(5), 655–65.
- Channon, S. & Green, P.S.S. (1999). Executive function in depression: The role of performance strategies in aiding depressed and non-depressed participants. *Journal of Neurology, Neurosurgery and Psychiatry*, 66, 162–171.
- Channon, S., Pellijeff, A., & Rule, A. (2005). Social cognition after head injury: Sarcasm and theory of mind. *Brain and Language*, 93, 123–134.
- Chaytor, N. & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13, 181–197.
- Chorover, S.L. & Cole, M. (1966). Delayed alternation performance in patients with cerebral lesions. *Neuropsychologia*, 4, 1–7.
- Clark, A. (1997). *Being there: putting brain, body and world together again*. Cambridge, MA: MIT Press.
- Clark, C., Prior, M., & Kinsella, G.J. (2000). Do executive function deficits differentiate between adolescents with ADHD and Oppositional Defiant/Conduct Disorder? A neuropsychological study using the Six Elements Test and Hayling Sentence Completion Test. *Journal of Abnormal Child Psychology*, 28, 403–414.
- Cohen, L.D., Braver, T.S., & O'Reilly, R.C. (1998). A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. In A.C. Roberts, T.W. Robbins, & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 195–220). Oxford: Oxford University Press.
- Cohen, J.D., Dunbar, K., & McClelland, J.L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332–61.
- Cohen, J.D. & Servan-Schreiber, D. (1992). Context, cortex and dopamine: A connectionist approach to behaviour and biology in schizophrenia. *Psychological Review*, 99, 45–77.
- Coltheart, M. & Davies, M. (2003). Inference and explanation in cognitive neuropsychology. *Cortex*, 39, 188–91.
- Corcoran, R. & Upton, D. (1993). A role for the hippocampus in card sorting? *Cortex*, 29, 293–304.
- Coughlan, A.K., Crowe, S.F., Mahony, K., & Jackson, M. (2004). Predicting competency in automated machine use in an acquired brain injury population using neuropsychological measures. *Archives of Clinical Neuropsychology*, 19, 673–691.
- Coughlan, A.K. & Hollows, S.K. (1985). *The adult memory and information processing battery (AMIPB)*. St. James's University Hospital, Leeds, U.K.
- Dawson, D.R., Anderson, N., Burgess, P.W., Levine, B., Rewilak, D., Cooper, E.K., Farrow, S., Kosciak, K.M., Krpan, K.M., Lo, A., Peer, M., & Stuss, D.T. (2005a). The ecological validity of the Multiple Errands Test—Hospital Version: Preliminary findings. *Poster presented at meeting of the International Neuropsychological Society, St. Louis, USA, February*.
- Dawson, D.R., Anderson, N., Burgess, P.W., Levine, B., Rewilak, D., Cooper, E.K., Farrow, S., Krpan, K., Peer, M., & Stuss, D.T. (2005b). Naturalistic assessment of executive function: The Multiple Errands Test. *Presentation at the American Congress of Rehabilitation Medicine, Chicago, USA, September (Abstract submitted)*.
- Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Science*, 95, 14,529–14,534.
- Duchaine, B.C., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron*, 43, 469–473.
- Duchaine, B.C., Wendt, T.N., New, J., & Kulomaki, T. (2003). Dissociations of visual recognition in a developmental agnostic: Evidence for separate developmental processes. *Neurocase*, 9, 380–389.
- Duncan, J. (2005). Prefrontal cortex and Spearman's *g*. In J. Duncan, L. Phillips, & P. McLeod (Eds.), *Measuring the mind: Speed, control and age* (pp. 249–274). Oxford: Oxford University Press.

- 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. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology*, *14*, 713–741.
- Duncan, J. & Miller, E.K. (2002). Cognitive focus through adaptive neural coding in the primate prefrontal cortex. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 278–291). Oxford: Oxford University Press.
- Emslie, H., Wilson, F.C., Burden, V., Nimmo-Smith, I., & Wilson, B.A. (2003). *Behavioural assessment of the dysexecutive syndrome in children (BADS-C)*. London, U.K.: Harcourt Assessment.
- Eslinger, P.J. & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient E.V.R. *Neurology*, *35*, 1731–1741.
- Evans, J.J., Chua, S., McKenna, P.J., & Wilson, B.A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine*, *27*, 635–646.
- Farias, S.T., Harrell, E., Neumann, C., & Houtz, A. (2003). The relationship between neuropsychological performance and daily functioning in individuals with Alzheimer's disease: Ecological validity of neuropsychological tests. *Archives of Clinical Neuropsychology*, *18*, 655–672.
- Fisher, A.G. (2003). *Assessment of motor and process skills. Vol. 1: Development, standardization, and administration manual* (5th ed.). Fort Collins, CO: Three Star Press.
- Fortin, S., Godbout, L., & Braun, C.M.J. (2003). Cognitive structure of executive deficits in frontally lesioned head trauma patients performing activities of daily living. *Cortex*, *39*, 273–291.
- Frith, C.D. & Frith, U. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B*, *358*(1431), 459–473.
- Fuster, J.M. (1997). *The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe* (3rd ed.). New York: Lippincott–Raven.
- Fuster, J.M. (2002). Physiology of executive functions: The perception-action cycle. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function*, pp. 96–108. Oxford U.K.: Oxford University Press.
- Garden, S.E., Phillips, L.H., & MacPherson, S.E. (2001). Midlife aging, open-ended planning, and laboratory measures of executive function. *Neuropsychology*, *15*, 472–482.
- Gelb, A. & Goldstein, K. (1920). *Psychologische analysen hirn-pathologischer faelle*. Leipzig: Barth.
- Gilbert, S.J., Frith, C.D., & Burgess, P.W. (2005). Involvement of rostral prefrontal cortex in selection between stimulus-oriented and stimulus-independent thought. *European Journal of Neuroscience*, *21*, 1423–1431.
- Gilbert, S.J., Simons, J.S., Frith, C.D., & Burgess, P.W. (2006). Performance-related activity in medial rostral prefrontal cortex (Area 10) during low-demand tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 45–58.
- Gilbert, S.J., Spengler, S., Simons, J.S., Steele, J.D., Lawrie, S.M., Frith, C.D., & Burgess, P.W. (in press). Functional specialization within rostral prefrontal cortex (area 10): A meta-analysis. *Journal of Cognitive Neuroscience*.
- Gioia, G.A. & Isquith, P.K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology*, *25*, 135–158.
- Goel, V. & Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, *33*, 623–642.
- Goel, V. & Grafman, J. (2000). The role of the right prefrontal cortex in ill-structured problem solving. *Cognitive Neuropsychology*, *17*, 415–436.
- Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lesions in a financial planning task. *Brain*, *120*, 1805–1822.
- Goldman-Rakic, P.S. (1995). Architecture of the prefrontal cortex and the central executive. *Annals of the New York Academy of Science*, *769*, 212–220.
- Goldman-Rakic, P.S. (1998). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. In A.C. Roberts, T.W. Robbins, & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 87–102). Oxford, U.K.: Oxford University Press.
- Goldman-Rakic, P.S. & Leung, H.-C. (2002). Functional architecture of the dorsolateral prefrontal cortex in monkeys and humans. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 85–95). Oxford: Oxford University Press.
- Goldstein, L.H., Bernard, S., Fenwick, P.B.C., Burgess, P.W., & McNeil, J. (1993). Unilateral frontal lobectomy can produce strategy application disorder. *Journal of Neurology, Neurosurgery and Psychiatry*, *56*, 274–276.
- Goldstein, K.H. & Scheerer, M. (1941). Abstract and concrete behaviour: An experimental study with special tests. *Psychological Monographs*, *53* (No. 2) (Whole No. 239).
- Grady, C.L. (1999). Neuroimaging and activation of the frontal lobes. In B.L. Miller & J.L. Cummings (Eds.), *The human frontal lobes: Function and disorders* (pp. 196–230). New York: Guilford Press.
- Grafman, J. (2002). The structured event complex and the human prefrontal cortex. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 292–310). Oxford: Oxford University Press.
- Grant, D.A. & Berg, E.A. (1948). A behavioural analysis of degree or reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology*, *38*, 404–411.
- Greve, K.W., Brooks, J., Crouch, J.A., Williams, M.C., & Rice, W.J. (1997). Factorial structure of the Wisconsin Card Sorting Test. *British Journal of Clinical Psychology*, *36*, 283–285.
- Haggard, P. & Richardson, J. (1996). Spatial patterns in the control of human movement. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 42–62.
- Hammond, K.R. (1966). *The psychology of Egon Brunswik*. New York: Holt, Rinehart, & Winston.
- Hammond, K.R. (1998). Ecological validity: Then and now. Unpublished manuscript available electronically at: <http://brunswick.org/notes/essay2.html>
- Harlow, J.M. (1848). Passage of an iron bar through the head. *Boston Medical and Surgical Journal*, *39*, 389–393.
- Harlow, J.M. (1868). Recovery from the passage of an iron bar through the head. Reprinted in Miller, E. (1993). *History of Psychiatry*, *4*, 271–278.
- Higginson, C.I., Arnett, P.A., & Voss, W.D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology*, *15*, 185–204.
- Hoclet, E., Gronier, D., Joly-Pottuz, B., & Carbonnel, S. (2003). The six element test with twenty-four TBI patients during reha-

- bilitation, and fractionation of the Supervisory System. *Revue de Neuropsychologie*, 13, 263–290.
- Hodgson, T.L., Bajwa, A., Owen, A.M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower-of-London task. *Journal of Cognitive Neuroscience*, 12, 894–907.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Jacobsen, C.F. (1931). A study of cerebral function in learning: The frontal lobes. *Journal of Comparative Neurology*, 52, 271–340.
- Jacobsen, C.F. (1936). Studies of cerebral function in primates: I. The functions of the frontal association areas in monkeys. *Comparative Psychology Monographs*, 13, 3–60.
- Jelicic, M., Henquet, C.E., Derix, M.M., & Jolles, J. (2001). Test-retest stability of the behavioural assessment of the dysexecutive syndrome in a sample of psychiatric patients. *International Journal of Neuroscience*, 110, 73–8.
- Kafer, K.L. & Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society*, 3, 108–19.
- Kanwisher, N. (2000). Domain-specificity in face perception. *Nature Neuroscience*, 3, 759–761.
- Kibby, M.Y., Schmitter-Edgecombe, M., & Long, C.J. (1998). Ecological validity of neuropsychological tests: focus on the California Verbal Learning Test and the Wisconsin Card Sorting Test. *Archives of Clinical Neuropsychology*, 13, 523–534.
- Kingstone, A., Smilek, D., Birmingham, E., Cameron, D., & Bischof, W.F. (2005). Cognitive ethology: Giving real life to attention research. In J. Duncan, L. Phillips, & P. McLeod (Eds.), *Measuring the mind: Speed, control and age* (pp. 341–357). Oxford: Oxford University Press.
- Kingstone, A., Smilek, D., & Ristic, J. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*, 12, 176–180.
- Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73, 31–68.
- Kliegel, M., McDaniel, M.A., & Einstein, G.O. (2000). Plan formation, retention, and execution in prospective memory: A new approach and age-related effects. *Memory and Cognition*, 28, 1041–1049.
- Knight, C., Alderman, N., & Burgess, P.W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation*, 12, 231–255.
- Kvavilashvili, L. & Ellis, J. (2004). Ecological validity and the real-life/laboratory controversy in memory research: A critical and historical review. *History & Philosophy of Psychology*, 6, 59–80.
- Leichnetz, G.R. (1986). Afferent and efferent connections of the dorsolateral precentral areas (area 4, hand/arm region) in the macaque monkey, with comparisons to area 8. *Journal of Comparative Neurology*, 254, 460–492.
- Levine, B., Freedman, M., Dawson, D., Black, S.E., & Stuss, D.T. (1999). Ventral frontal contribution to self-regulation: Convergence of episodic memory and inhibition. *Neurocase*, 5, 263–275. (Special issue on orbitofrontal cortex).
- Levine, B., Robertson, I., Clare, L., Carter, G., Hong, J., Wilson, B.A., Duncan, J., & Stuss, D.T. (2000). Rehabilitation of executive functioning: An experimental–clinical validation of goal management training. *Journal of the International Neuropsychological Society*, 6, 299–312.
- Levine, B., Stuss, D.T., Milberg, W.P., Alexander, M.P., Schwartz, M., & Macdonald, R. (1998). The effects of focal and diffuse brain damage on strategy application: Evidence from focal lesions, traumatic brain injury and normal aging. *Journal of the International Neuropsychological Society*, 4, 247–264.
- MacSweeney, M., Campbell, R., Woll, B., Giampicco, V., David, A.S., McGuire, P.K., Calvert, G., & Brammer, M.J. (2004). Dissociating linguistic and non-linguistic gestural communication in the brain. *NeuroImage*, 22, 1605–1618.
- Malec, J.F. & Lezak, M.D. (2003). *Manual for the Mayo-Portland adaptability inventory*. PM&R–3MB–SMH, Mayo Medical Center, Rochester, MN 55905, USA.
- Malmö, R.B. (1948). Psychological aspects of frontal gyrectomy and frontal lobotomy in mental patients. *Research Publications of the Association for Research into Nervous and Mental Disease*, 27, 537–564.
- Manchester, D., Priestly, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, 18, 1067–1081.
- Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I.H. (2002). Rehabilitation of executive function: Facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, 40, 271–281.
- Marenco, S., Coppola, R., Daniel, D.G., Zigun, J.R., & Weinberger, D.R. (1993). Regional cerebral blood flow during the Wisconsin Card Sorting Test in normal subjects studied by xenon-133 dynamic SPECT: Comparison of absolute values, percent distribution values, and covariance analysis. *Psychiatry Research*, 50, 177–192.
- McDonald, S., Flanagan, S., Martin, I., & Saunders, C. (2004). The ecological validity of TASIT: A test of social perception. *Neuropsychological Rehabilitation*, 14, 285–302.
- Mentzel, H.J., Gaser, C., Volz, H.P., Rzanny, R., Hager, F., Sauer, H., & Kaiser, W.A. (1998). Cognitive stimulation with the Wisconsin Card Sorting Test: Functional MR imaging at 1.5 T. *Radiology*, 207, 399–404.
- Miller, E. (1984). Verbal fluency as a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology*, 23, 53–57.
- Milner, B. (1963). Effects of different brain regions on card sorting. *Archives of Neurology (Chicago)*, 9, 90–100.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Morris, R.G., Miotto, E.C., Feigenbaum, J.D., Bullock, P., & Polkey, C.E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia*, 35, 1147–57.
- Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12, 313–324.
- Nilsson, J.P., Soderstrom, M., Karlsson, A.U., Lekander, M., Akerstedt, T., Lindroth, N.E., & Axelsson, J. (2005). Less effective executive functioning after one night’s sleep deprivation. *Journal of Sleep Research*, 14, 1–6.
- 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: Advances in research*, (Vol. IV, pp.1–18). New York: Plenum. Reprinted in *Cognitive neuroscience: A reader* (M.S. Gazzaniga (Ed.), Malden, Mass: Blackwell, 2000. Initially published as a technical report in 1980.

- Odhuba, R.A., van den Broek, M.D., & Johns, L.C. (2005). Ecological validity of measures of executive functioning. *British Journal of Clinical Psychology, 44*, 269–278.
- Okuda, J., Frith, C.D., Burgess, & P.W. (2004). Organisation of time- and event-based intentions in rostral prefrontal cortex. *NeuroImage, 22*(S1) S53.
- Owen, A.M., Herrod, N.J., Menon, D.K., Clark, J.C., Downey, S.P., Carpenter, T.A., Minhas, P.S., Turkheimer, F.E., Williams, E.J., Robbins, T.W., Sahakian, B.J., Petrides, M., & Pickard, J.D. (1999). Redefining the functional organization of working memory processes within human lateral prefrontal cortex. *European Journal of Neuroscience, 11*, 567–574.
- Passingham, R. (1993). *The frontal lobes and voluntary action*. Oxford, U.K.: Oxford University Press.
- Penfield, W. & Evans, J. (1935). The frontal lobe in man: A clinical study of maximum removals. *Brain, 58*, 115–133.
- Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia, 12*, 323–330.
- Petrides, M. (1994). Frontal lobes and working memory: Evidence from investigations of the effects of cortical excisions in nonhuman primates. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology, Vol. 9* (pp. 59–82). Amsterdam: Elsevier.
- Petrides, M. (1998). Specialized systems for the processing of mnemonic information within the primate frontal cortex. In A.C. Roberts, T.W. Robbins, & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 103–116). Oxford: Oxford University Press.
- Petrides, M. & Pandya, D.N. (2002). Association pathways of the prefrontal cortex and functional observations. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 31–50). Oxford: Oxford University Press.
- Pribram, K.H. (1973). The primate frontal cortex—executive of the brain. In K.H. Pribram & A.R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 293–314). New York: Academic Press.
- Ramnani, N. & Owen, A.M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience, 5*, 184–194.
- Ramnani, N., Toni, I., Passingham, R.E., & Haggard, P. (2001). The cerebellum and parietal cortex play a specific role in coordination: A PET study. *NeuroImage, 14*, 899–911.
- Robertson, I.H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *Test of everyday attention (TEA)*. London: Harcourt Assessment.
- Rowe, J.B., Owen, A.M., Johnsrude, I.S., & Passingham, R.E. (2001). Imaging the mental components of a planning task. *Neuropsychologia, 39*, 315–327.
- Semkovska, M., Bedard, M.A., Godbout, L., Limoge, F., & Stip, E. (2004). Assessment of executive function during activities of daily living in schizophrenia. *Schizophrenia Research, 69*, 289–300.
- 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*. New York: Cambridge University Press.
- Shallice, T. (2002). Fractionation of the supervisory system. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 261–277). New York: Oxford University Press.
- Shallice, T. & Burgess, P.W. (1991a). Deficits in strategy application following frontal lobe damage 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. Benson, (Eds.), *Frontal lobe function and injury* (pp. 125–138). Oxford: 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). Domains of supervisory control and the temporal organisation of behaviour. *Philosophical Transactions of the Royal Society B, 351*, 1405–1412. [Reprinted in A.C. Roberts, T.W. Robbins, & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions*, pp. 22–35. Oxford, U.K.: Oxford University Press, 1998.]
- Shallice, T., Burgess, P.W., Schon, F., & Baxter, D.M. (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. (Reprinted in J. Davidoff (Ed.), *Brain and behaviour: Critical concepts in psychology, Vol. 4*, London: Routledge.)
- Simons, J.S., Gilbert, S.J., Owen, A.M., Fletcher, P.C., & Burgess, P.W. (2005a). Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. *Journal of Neurophysiology, 94*, 813–820.
- Simons, J.S., Owen, A.M., Fletcher, P.C., & Burgess, P.W. (2005b). Anterior prefrontal cortex and the recollection of internally-generated thoughts. *Neuropsychologia, 43*, 1774–1783.
- Simons, J.S., Schölvinc, M., Gilbert, S.J., Frith, C.D., & Burgess, P.W. (In press). Differential components of prospective memory? Evidence from fMRI. *Neuropsychologia*.
- Stroop, J.R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology, 18*, 643–662.
- Stuss, D.T., Binns, M.A., Murphey, K.J., & Alexander, M.P. (2002). Dissociations within the anterior attentional system: Effects of task complexity and irrelevant information on reaction time speed and accuracy. *Neuropsychology, 16*, 500–513.
- Stuss, D.T., Floden, D., Alexander, M.P., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. *Neuropsychologia, 39*, 771–786.
- Stuss, D.T., Levine, B., Alexander, M.P., Hong, J., Palumbo, C., Hamer, L., Murphy, K.J., & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia, 38*, 388–402.
- Stuss, D.T., Shallice, T., Alexander, M.P., & Picton, T.W. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences, 769*, 191–212.
- Sukantarat, K.T., Burgess, P.W., Williamson, R.C.N., & Brett, S.J. (2005). Prolonged cognitive dysfunction in survivors of critical illness. *Anaesthesia, 60*, 847–853.
- Teuber, H.-L. (1964). The riddle of frontal lobe function in man. In J.M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 410–477). New York: McGraw-Hill.
- van den Broek, M.D., Bradshaw, C.M., & Szabadi, E. (1993). Utility of the Modified Wisconsin Card Sorting Test in neuropsychological assessment. *British Journal of Clinical Psychology, 32*, 333–343.
- Ward, L.M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.

- Weinberg, G.M. (1975). *An introduction to general systems thinking*. New York: Wiley.
- Weigl, E. (1927). Zur Psychologie sogenannter Abstraktionsprozesse. *Zeitschrift für Psychologie*, 103, 2–45. [Translated by M. Rioch and reprinted as: On the psychology of so-called processes of abstraction. *Journal of Abnormal and Social Psychology*, 36, 3–33, 1948].
- Welsh, M.C., Revilla, V., Strongin, D., & Kepler, M. (2000). Towers of Hanoi and London: Is the nonshared variance due to differences in task administration? *Perceptual and Motor Skills*, 90, 562–572.
- Welsh, M.C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition*, 41, 231–234.
- Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural assessment of the dysexecutive syndrome*. London: Harcourt Assessment.
- Wilson, B.A., Cockburn, J., & Baddeley, A. (2003). *The Rivermead Behavioural Memory Test (RBMT-II)*. London: Harcourt Assessment.
- Wilson, B.A., Evans, J.J., Emslie, H., Alderman, N., & Burgess, P.W. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation*, 8, 213–228.
- Zakzanis, K.K. & Young, D.A. (2001). Memory impairment in abstinent MDMA (“Ecstasy”) users: A longitudinal investigation. *Neurology*, 56, 966–969.