Fifty Years of Prefrontal Cortex Research: Impact on Assessment

Paul W. Burgess,1 AND Donald T. Stuss2,3,4

1Institute of Cognitive Neuroscience, University College London
2University of Toronto, Toronto, Ontario Canada
3Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada
4Rotman Research Institute of Baycrest, Toronto, Ontario, Canada

(RECEIVED April 13, 2017; FINAL REVISION June 19, 2017; ACCEPTED June 22, 2017)

Abstract

Our knowledge of the functions of the prefrontal cortex, often called executive, supervisory, or control, has been transformed over the past 50 years. After operationally defining terms for clarification, we review the impact of advances in functional, structural, and theoretical levels of understanding upon neuropsychological assessment practice as a means of identifying 11 principles/challenges relating to assessment of executive function. Three of these were already known 50 years ago, and 8 have been confirmed or emerged since. Key themes over this period have been the emergence of ecological validity; discovery of the complexity of the frontal lobe control system; invention of new tests for clinical use; development of key theoretical frameworks that address the role of prefrontal cortex systems in the organization of human cognition; the move toward considering brain systems rather than brain regions; the advent of functional neuroimaging, and its emerging integration into clinical practice. Despite these huge advances, however, practicing neuropsychologists are still desperately in need of new ways of measuring executive function. We discuss pathways by which this might happen, including decoupling the two levels of explanation (information processing; brain structure) and integrating very recent technological advances into the neuropsychologist’s toolbox.

Keywords: Prefrontal cortex, Frontal lobe, Executive function, Neuropsychological tests, Neuropsychology, Functional neuroimaging, Control function

INTRODUCTION

In 1928, Tilney proposed that the entire evolutionary existence of man could be considered as the age of the frontal lobes (Tilney, 1928). The research interest in this brain region was slow to evolve, however, but quickly escalated, particularly with the influential publications of Luria (e.g., 1966), interpreted for neuropsychological assessment by Christensen (1975). In 1986, Stuss and Benson summarized the known published literature in the English language (and some non-English) on the frontal lobes in a little over 300 pages. Comparison of two subsequent edited books (Stuss & Knight, 2002, 2013) revealed that so much research had been done that by 2013 each of the sections could have themselves been books with a particular focus.

Covering all of such a rapid increase in our knowledge in one paper is clearly not possible. Our lens is one key aspect of the professional lives of neuropsychologists: assessment of “executive function.” In particular, we focus on three major transition points that have affected the field significantly. They are not individual papers, but changes in direction or method. The first transition point arose from the development of tests specifically designed to assess particular executive symptoms; the second is related to the evolution of theory, and the relationship of theory development to new neuroanatomical knowledge; the third was the advent of imaging techniques that enabled an unprecedented window into the human brain and functioning. The impact of these transition points will be expressed in the form of 11 principles. Many of these principles actually represent challenges for the future.

The focus is on the functions traditionally associated with the prefrontal cortex (PFC), although the anatomical terms are sometimes also ambiguously used (i.e., frontal, prefrontal). There has been a rather unhelpful tendency in the field to use interchangeably the terms executive dysfunction and frontal lobe dysfunction. A related problem is that the term “executive” is defined in different ways, and often quite broadly. We elect to see executive functions in a narrower manner: a set of cognitive
Despite these limitations, the Initial Assessment Efforts to-late 1960s (Benton, 1991). The attempt to quantify the frontal lobe function already existed in some form by the mid-20th century. Some of the concepts that we currently hold in relation to domain general processes such as attentional control, planning, reasoning, problem solving, and monitoring. In this view, the terms are not synonymous. Executive functions then (see below) are one type of domain general processes. In this review, however, the term is used as originally expressed by a particular researcher. Regardless of one’s bias, taking the perspective of the functions of the frontal lobes and differentiating it from executive functions provides a rationale for a methodological approach and a structure for operational definition.

WHAT DID WE ALREADY KNOW 50 YEARS AGO?

Some of the concepts that we currently hold in relation to frontal lobe function already existed in some form by the mid-to-late 1960s (Benton, 1991). The attempt to quantify the deficits was hampered by the golden age of mental testing with the pioneering work of figures such Cattell, Binet, Galton, and Spearman. This history of test development created two problems for the development of psychometric tests to measure the symptoms of frontal lobe dysfunction. First, the main concern of these early theorists was to measure intelligence. Yet it was apparent from the clinical studies of Phineas Gage and others (Harlow, 1848), and supported by animal research, that the main deficit these patients experienced was not a loss of some generalized “intelligence” resource. What did occur with frontal damage was some alteration in character, which was hard to describe precisely, never mind quantify. Thus the object that was the focus of the psychometric theorists (measurement and identification of a construct called “intelligence) seemed tangential to the characterization of the symptoms of frontal lobe damaged patients.

A related problem was that the intelligence theorists emphasized the psychometric properties of their tests such as internal consistency and test–retest reliability. Yet if the frontal lobes support a system whose purpose it is to deal with novelty and adaptation of behavior, then these metrics may be a poor indicator of the construct validity of the test (e.g., Burgess, 1997). Moreover, Stuss, Murphy, Binns, and Alexander (2003) demonstrated that inconsistency and variability may be a hallmark of frontal lobe dysfunction; a different psychometric approach is required. These problems lead to the first new principle that the last 50 years of research into frontal lobe function in humans has taught us:

New principle 1: You probably can’t measure all cognitive abilities using the same methodological approach.

Initial Assessment Efforts

Despite these limitations, the first 50 years of the 20th century were a rich period for the creation of mental tests, and several tests that we now think of as measures of “executive function” being created then. However, the link to frontal lobe function was usually not a principal aim, nor did notions of “executive function” exist then (the use of the term “executive” in relation to the frontal lobes is often attributed to Pribram, 1973).

One such example is the Verbal Fluency Test, part of Thurstone and Thurstone’s (1938) attempts to discover the factorial structure of intelligence. Another is the Trail Making Test (TMT). Originally conceived as a measure of “distributed attention,” considered as a component of general intelligence, this test was incorporated into the Army Individual Test Battery in 1944, and then later adapted for the Halstead-Reitan Neuropsychological Test Battery. It was Reitan (1958) who raised a further issue for the development of clinical tests of neuropsychological dysfunction. While his study demonstrated the utility of the TMT in distinguishing between patients with brain damage and those without, he also made the point that “patients with brain damage may perform poorly on the TMT as a consequence of several types of impairment associated with variously located brain lesions” (Reitan, 1958, p. 275). Here he highlights understanding of two principles that we have inherited:

Inherited principle 1: Tests of executive function often measure multiple cognitive abilities simultaneously.

Inherited principle 2: Tests of executive function do not just test frontal lobe function. They often may be failed by patients with dysfunction elsewhere as well.

We can add to this a further principle made clear in Reitan (1958): “this situation [i.e., that many different lesions cause many different impairments on executive tests] presents an advantage for the test in its use as a screening instrument but a possible disadvantage in terms of its potential use in differential lateralization or localization of brain lesions” (p. 275).

Inherited principle 3: The characteristics of the best clinical and the best experimental tests of executive function may differ since their aims differ.

These were the beginnings of theorizing about abilities for which the frontal lobes played a special supporting role (e.g., Halstead, 1947). Furthermore, notions that this might be worth considering independently from a construct of “intelligence” were already in gestation. Yet the faith in a specific link between damage to PFC and impairment on the tests appeared to be premature (e.g., Anderson, Bigler, & Blatter, 1995). Reitan and Wolfson (1995) mused, after failing to find frontal specific deficits on the Category Test and Trails B, that “neuropsychologists should adopt a more critical attitude concerning so-called “frontal lobe deficits” (p. 50).

THE MODERN PERSPECTIVE ON THE OLDER FINDINGS

The past 50 years have seen the invention of several executive tests that were specifically designed for use with patients with frontal lobe involvement (e.g., Cognitive Estimation Task, Shallice and Evans, 1978; Tower of London test, Shallice, 1982; Homophone meaning generation test (Warrington, 2000), and the designs of these kinds of studies have been able to examine the relationship with constructs such as general intelligence. So, what have we learned about the three inherited principles over the past 50 years, and what have we added to them?

There is increasing evidence that tests designed to measure executive function often tap multiple cognitive abilities.
Della Sala, Papagno, and Spinnler, 1997; Burgess et al., 2010). The left-hand picture shows the areas of medial area 10 of the prefrontal cortex (rostral PFC) as discovered by Gilbert et al. (2012) have shown how the frontal pole (i.e., the most anterior part of PFC) connects to the occipital lobes (i.e., the most posterior part of the brain) via long association and projection fibers referred to as the inferior fronto-occipital pathway and uncinate fasciculus. These kinds of connections are extremely extensive, and probably explain in large part the consistent associations in activation between certain subregions of the prefrontal cortex and other regions of the brain, as demonstrated for example by Gilbert, Gonen-Yaacovi, Benoit Bolle, & Burgess (2010) (see Figure 2).

Fig. 1. A Brixton Test stimulus page. Participants are asked to predict the position of the filled circle on the next stimulus page (see text for description). Using this task, Burgess and Shallice (1996) showed that patients with frontal lobe lesions can show three different types of failure on this task, thus demonstrating that executive tasks may measure more than one construct simultaneously. Reverberi, Lavaroni, Giglib, and Skrapb (2005) later replicated the frontal deficit on a version of this test, and argued that the pattern of their results suggested functional dissociations between inductive reasoning, monitoring and working memory, with processes relevant to induction supported by left lateral frontal cortex and monitoring and checking by right lateral cortex.

EXECUTIVE TEST PERFORMANCE: BRAIN CIRCUITS VERSUS BRAIN REGIONS

The second principle that was already known 50 years ago, confirmed and extended in the last 50 years, was that tests of executive function do not just test frontal lobe function. The evidence for an association between performance on executive tests and frontal lobe lesions is much more inconsistent than perhaps the early theorists appreciated. In a meta-analysis, Demakis (2004) found that participants with frontal damage performed significantly more poorly than patients with lesions outside the frontal lobes on all components of the Stroop Test and Trails A. However, on two measures widely considered to be sensitive to frontal functions, namely the Category Test and Trails B, they were not significantly different. The conclusion: these “findings are surprising and not consistent with the long-held assumption that these latter measures are sensitive to frontal lobe damage” (p. 446).

Has research over the past 50 years explained why patients with lesions outside the prefrontal cortex fail executive tests, and why multiple regions within PFC can be implicated across different studies? We have identified one definite reason, although the precise details will take many decades to work out thoroughly. As structural neuroimaging and animal studies over the last few years have identified, prefrontal cortex (PFC) has a huge number of connections both within the PFC and outside it. These connections can be, for example, frontal–subcortical, frontal–limbic, frontal–cerebellar, and frontal–cortical. These connections in some cases are very distant from their origins. For instance, Catani et al. (2012) have shown how the frontal pole (i.e., the most anterior part of PFC) connects to the occipital lobes (i.e., the most posterior part of the brain) via long association and projection fibers referred to as the inferior fronto-occipital pathway and uncinate fasciculus. These kinds of connections are extremely extensive, and probably explain in large part the consistent associations in activation between certain subregions of the prefrontal cortex and other regions of the brain, as demonstrated for example by Gilbert, Gonen-Yaacovi, Benoit Bolle, & Burgess (2010) (see Figure 2).

Gratton, Nomura, Perez, and D’Esposito (2012) using graph theory on resting state fMRI data demonstrated how focal brain damage could have a more widespread effect if the damage affects regions “important for the communication between networks.” Regions such as the thalamus have also been considered to have connecting and integrating functions,
<table>
<thead>
<tr>
<th>Study</th>
<th>Condition</th>
<th>Variable</th>
<th>Right medial</th>
<th>Left medial</th>
<th>Right lateral</th>
<th>Left lateral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stuss et al. (1998)</td>
<td>1A. Letter fluency</td>
<td>Errors</td>
<td>6A,8B,9,24</td>
<td>6A,8B,9,24</td>
<td>9/46,46,44,45,47/12</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td>1B. Letter fluency</td>
<td>No. of words</td>
<td>8B,9,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td>1C. Semantic fluency</td>
<td>No. of words</td>
<td>8B,9,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td>2. Shammi &amp; Stuss (1999)</td>
<td>Humor judgment</td>
<td>Two variables</td>
<td>8B,9,10s</td>
<td>8B,9,10s</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td>3. Stuss et al. (2000)</td>
<td>WCST</td>
<td>128 Categories</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perseverations of criterion</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perseverations of response</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Set Loss</td>
<td>10i,11,25,32</td>
<td>10i,11,25,32</td>
<td>9/46,46,44,45,47/12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WCST</td>
<td>64A Categories</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perseverations of criterion</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perseverations of response</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simple RT</td>
<td>Reaction time</td>
<td>6A,8B,9,24</td>
<td>6A,8B,9,24</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Easy task</td>
<td>Simple-Easy difference</td>
<td>6A,8B,9,24</td>
<td>6A,8B,9,24</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex task</td>
<td>Proportional increase</td>
<td>6A,8B,9,24,10s</td>
<td>6A,8B,9,24,10s</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Redundant task</td>
<td>Proportion increase</td>
<td>6A,8B,9,24</td>
<td>6A,8B,9,24</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Second simple task</td>
<td>Omissions</td>
<td>6A,8B,9,24</td>
<td>6A,8B,9,24</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Modified Stroop test</td>
<td>RT slowing</td>
<td>6A,8B,9,24,24s</td>
<td>6A,8B,9,24,24s</td>
<td>9/46,46,44,45,47/12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Modified Stroop test</td>
<td>False positives</td>
<td>6A,8B,9,24,24s</td>
<td>6A,8B,9,24,24s</td>
<td>9/46,46,44,45,47/12</td>
</tr>
</tbody>
</table>
demonstrating the complexity of network organization (Hwang, Bertolero, Liu, & D’Esposito, 2017). It is likely this very significant reciprocal connectivity with so many different non-frontal regions not only resulted in the confusions in operational definitions but also provided the foundation for the prefrontal cortex’s unique contribution to human functioning. It seems likely at this point that researchers who study brain-behaviour relations may move toward analysis at the level of brain circuits rather than isolated brain regions. This enterprise has already started (e.g., Thiebaut de Schotten et al., 2017). Moreover, Koechlin and colleagues (Koechlin, Ody, & Kouneiher, 2003; Koechlin & Summerfield, 2007; see also D’Esposito & Badre, 2012) discussed an anterior–posterior hierarchy within the prefrontal cortex related to higher order processing requirements and selection of action.

### WHAT DO OUR TESTS MEASURE?

#### ON THE PERILS OF MIXING LEVELS OF EXPLANATION

How does one map construct onto brain circuit when one does not know which constructs one is measuring? For instance, prima facie the Category Test (CT) and the WCST are rather similar. In fact the CT and WCST probably share only around 30% common variance: “the CT and WCST should not be regarded as similar measures of one construct such as ‘abstraction’” (Perrine, 1993, p. 461). The problem here perhaps is largely self-made. As mentioned above, the notion of “frontal lobe deficits” mixes levels of explanation (anatomy, information processing) in a dubious way. The tests measure various aspects of executive function, which may or may not be related to one or more regions of the frontal lobes.

This point was perhaps made most clearly by Alan Baddeley (e.g., Baddeley, 1986) as he introduced the term “dysexecutive” to refer to deficits in executive function. The big contribution here was to start to talk about what had previously been called “frontal lobe functions” in the same way that we talk about language or visuo-perceptual skills; we do not refer to these functional abilities by the regions of the brain (e.g., “temporal lobe function”). The inherent weakness of this theoretical innovation was, however, that the term “executive function” is an ill-defined collective for a range of abilities which may or may not relate to each other, and it invites endless re-invention. Nevertheless, it was probably a step in the right direction; let’s call this new principle two:

**New principle 2:** Function and structure are two different levels of explanation; while a goal of neuroscience research is to understand these relationships, it is unrealistic to expect a simple correspondence in discourse between them.

### REASONS FORWEAK PSYCHOMETRIC QUALITIES IN EXECUTIVE TASKS

As noted above, it is not unusual to find low coefficients of reliability or internal consistency in executive tasks. Critical in this respect might be findings from both neuroimaging and other neuropsychological studies that relatively minor changes in task format can have marked effect upon results (e.g., Stuss et al., 2006; see also Table 1). Although this discovery is perhaps not entirely the product of the past 50 years of research into executive functions, this is probably nevertheless one that we might claim for this period with some justification.

**New principle 3:** Seemingly small changes in executive task format can lead to big differences in the results you get.

But why is it that small changes in format can make such a difference, and what does it mean for our understanding both of frontal lobe function and how to assess it? There are several possible explanations. One potential explanation for the apparent instability of many executive function tasks relates to changes in the novelty of the task. This may introduce individual differences into test variance that is hard to predict or control (Burgess, 1997). A second explanation is...
that individual variability within a test and consistency across tests and testing sessions are a necessary and lawful by-product of disturbed control processes (Stuss et al., 2003), that is, variability may itself be a dependent measurement. Third, there is the interface between emotion and cognition, which is often the role of the prefrontal region. Small changes in the phrasing of the task instructions emphasizing either the action that needed to be carried out, or the reward that would be won, led to differential patterns of activation within rostral prefrontal cortex (Gilbert, Gollwitzer, Cohen, Oettingen, & Burgess, 2009; Figure 3). Fourth, differences in instructions are likely to call upon different processes to complete a task, seemingly independent of emotional/cognition interplay (Floden, Vallesi, & Stuss, 2011; Stuss et al., 2000).

Another potential, and perhaps related, explanation is that the frontal lobes are all about the self, and managing the self in terms of others. For instance Benoit, Gilbert, Volle, and Burgess (2010) used fMRI while two factors were crossed: (i) engaging in personality trait or episodic source memory judgments and (ii) the reference person for these judgments, which was either oneself or a friend. The imaging results suggested there may be complex aspects to do with social framing when administering executive tasks that we currently do not understand well, perhaps leading to inconsistencies in test–retest or inter-item statistics, and large variability between test participants.

New principle 4: Variability and inconsistency between and within patients are common. They should be measured and the possible reasons should be investigated during assessment.

**THERE IS NO “PREFRONTAL CORTEX SYNDROME”: THE LIMITS OF ASSOCIATIONS AND DISSOCIATIONS.**

Demonstrations of failures of association between executive symptoms and/or executive test performances are probably more numerous than demonstrations of clear associations. Since there may be more reasons for null findings than significant ones, these findings were probably dismissed over the years, allowing the notion of the existence of a “frontal lobe syndrome” to perpetuate. But it has become clear over the past 50 years that the behavioral and cognitive symptoms of frontal lobe damage can show stark dissociations.

**Dissociations of Cognitive Symptoms of Frontal Lobe Damage**

One of the clearest early examples perhaps was that of Burgess and Shallice (1996b). Using the Hayling Sentence Completion Test (HSCT), they had three separate measures each using the same format of presentation, stimuli and response. Deficits in the three different measures could doubly dissociate from each other but were unlikely to be due to differences in task format or stimuli. Later, Volle et al. (2012) showed why this was the case: each of the three types of error was associated with lesions within different parts of PFC (see Figure 4).

---

*Fig. 3.* Gilbert, Cohen, Gollwitzer, Oettingen, & Burgess (2009) compared activation while participants were performing a computerized prospective memory task under two instruction conditions (shown in the text boxes in Figure 3). Nothing else about what the participants saw, heard or did was different. But the apparently minor changes in instruction were sufficient to change the patterns of activation seen in rostral prefrontal cortex (Brodmann Area 10), as well as change the performance of the participants (the instruction “If the same letter is on both sides then I will press the middle button” produced better performance).
The (a) growth in demonstration of dissociations and (b) the sheer diversity of symptoms began the hunt for the limits of the “unity and diversity” of frontal lobe function (Teuber, 1972). It was perhaps inevitable, therefore, that the theoretical developments thenceforward might steadily get more complex. An example is the supervisory attentional system (SAS) account of Norman and Shallice (1986). It provided a good shorthand explanation for cardinal dysexecutive symptoms such as perseveration, problems with inhibition and dealing with novelty (e.g., the Cognitive Estimation Test). But the weakness of the original account is that it did not look inside the SAS “box” to unpack the structure of the system. Accordingly, the results from a series of lesion studies over the next 25 years were used to delineate the nature of the separate subsystems that putatively together comprise the SAS (Shallice & Burgess, 1996; Shallice & Cooper, 2011).

Other theorists were motivated more by the possibilities suggested by anatomical organization. For instance, Stuss and Benson (1986), guided by the work of Nauta, Pandya and others, emphasized the reciprocal interconnectivity of the frontal lobes with virtually all other brain regions to a degree far beyond other regions. As such, the frontal lobes appeared designed to have a major role in integrating and controlling information from all functional domains. They postulated four hierarchical levels of brain functional organization: (a) more posterior/basal fixed functional systems; (b) drive (medial frontal) and sequencing/setting/integration (lateral frontal) functions; (c) anticipation, goal selection, pre-planning and monitoring as executive control functions (no localization proposed); (d) self-awareness and self-consciousness (more anterior localization, and more integrated functioning), important for self-reflection, self-awareness, and awareness of one’s self with society.

Stuss, Shallice, Alexander, and Picton (1995) approached the multiple processes theory combining the theoretical and anatomical approaches. The theoretical focus was on attention, as defined in the original theory of Norman and Shallice, but starting with the assumption that there are likely components of this attentional system. Closely allied with the first basic assumption is that a task measuring an attentional construct, such as sustained attention, is not equivalent to a fundamental process. An extensive literature review led to a grouping of tasks. All the tasks were analyzed, on the basis of which the authors postulated five frontal component processes related to the anterior attentional system and the interaction with more fixed posterior processes: energize a schema; inhibit a schema; adjust contention scheduling; monitor goal fulfillment, and if-then logic to adjust the system. Combinations of these processes were hypothetically able to explain performance on all of the different attentional tasks.

To test these ideas, a new battery of tests to investigate attention was developed: ROBBIA, the Rotman Baycrest Battery to Investigate Attention (see Stuss & Alexander, 2007, for a summary of ROBBIA and sub-components). To isolate processes, the common assumption that the frontal lobes were needed primarily for novel and complex tasks was ignored, since complex or novel tasks would by definition require multiple integrated processes. The simplest single test was developed, which was then used as scaffolding upon which more complex demands were made. Well-known clinical tests, such as the Stroop, TMT, WCST, and Verbal Fluency were also included as an addendum to examine if similar processes could be revealed in both the clinical and experimental measures.

The results were consistent and replicated across different tests, different task modalities (i.e., memory, reaction time, language), and different patient groups (Alexander, Stuss, & Fansabedian, 2003; Alexander, Stuss, Picton, Shallice, & Gillingham, 2007; Alexander, Stuss, Shallice, Picton, & Gillingham, 2005; Picton, Stuss, Shallice, Alexander, & Gillingham, 2006; Picton et al., 2007; Shallice, Stuss, Alexander, Picton, & Derksen, 2008; Shallice, Stuss, Picton, Alexander, & Gillingham, 2008; Stuss et al., 2005; Stuss, Binns, Murphy, & Alexander, 2002; see Stuss & Alexander, 2007, for an overview). No matter which task was used, including the clinical tasks (Stuss et al., 1998, 2001a, 2000; Stuss, Floden, Alexander, Levine, & Katz, 2001b), task performance could be explained by three cognitively and anatomically distinct frontal processes. They were labeled: “energization,” process of initiation or sustaining responses (clinically, think slowness, apathy); “task setting,” ability to establish a stimulus-response relationship to respond to a target with specific attributes (clinically, think planning, organizing, learning how to do a new task); “monitoring,” process of checking a task over time, and for quality control (clinically, think checking, staying on task).

Each was related to a different frontal brain region: energization, superior medial; task setting, left lateral; monitoring, right lateral. The term “executive” was reserved for the left and right lateral frontal functions. This was done because of the operational definition of the term executive as well as the influence of the dual origin of brain development (see below), and earlier experience with focal lesion patients such as in the leucotomy studies.

The results were a modification from the original hypothesized model. Energization and monitoring remained, but the hypothesized processes of if-then logic and adjusting contention scheduling seemed to be subsumed by the left lateral task setting/planning process. The focus on parsimony of processes through the development of scaffolded tasks also suggested that any of the supposed “inhibitory tasks,” such as the Stroop, Suppression tasks, etc., all could be explained by one or more of the other processes such as energization or task setting (Alexander et al., 2007; Floden et al., 2011; Stuss & Alexander, 2007; Stuss et al., 2001b).

These studies exemplify how the “newer” tests might be characterized psychometrically. The sub-group stratification results in less variance, and more consistent results. The major method is to illustrate that the same results can be demonstrated across different patient groups, across the same tests repeated, and across different modalities (i.e., the same processes with the same anatomical localization in attentional, language, and memory tasks).
Anatomy also suggested looking beyond attention to understand frontal lobe functions and emphasize that the terms dysexecutive and frontal lobe syndromes cannot be used interchangeably. Pandya and colleagues had reintroduced Sanides’ concept of a dual origin of cerebral cortex development (Pandya & Barnes, 1987; Pandya & Yeterian, 1996). Two primordial regions, archicortical and paleocortical, representing hippocampal and olfactory areas, evolve, respectively, into a more dorsal system subserving sensory and cognitive processes, and a medial ventral/orbital system related to emotional functions. This representational division includes the frontal lobes, and suggests the importance of including emotional/behavioral regulation (paleocortical) in addition to attentional (archicortical) processes.

In the mid-1970s, Stuss and colleagues were provided the opportunity of studying the long-term effects of frontal leucotomies. Of particular fortune was that different groups of patients, all with similar diagnoses from the same institution by the same groups of physicians, could be compared on either having undergone or not a frontal leucotomy (see- mingly the only long-term study with such a control) or on the degree of behavioral recovery after surgery.

Several important lessons were learned which paved the way for future research, as illustrated by the “good recovery” leucotomy patients. First, even though the patients were described as “frontal” patients, lesion location was an important factor. The leucotomy patients, with pathology primarily in orbitofrontal regions, were not significantly different from the matched normal control group on standard tests of “executive functions”: WCST, TMT, Stroop, and several clinically used attentional measures such as Serial Sevens (Stuss et al., 1981, for overview). As was demonstrated by many labs later, these functions are subserved primarily by lateral frontal regions. There was a second corollary conclusion. If those tests are considered “executive,” then not all regions of the prefrontal cortex should be considered as executive.

Third, to assess social behavior and emotional responsivity, the standard clinical personality tests were not valuable, and new tests had to be developed (Stuss & Benson, 1983). In so doing, it was essential to dissociate the actual deficit underlying an impoverished score. In this instance, this demonstrated that the frontal leucotomy patients had full knowledge of appropriate social responsiveness, but did not apply this knowledge in a consistent manner. Fourth, the importance of context in assessment was emphasized.

Observationally, for example, the patients appeared to have a severe attentional disorder, but performed excellently in test situations. What was observed was the patient behaving in an unstructured situation; what was tested was their performance when the examiner “became the frontal lobes” of the patient in a regimented test situation. This led to the notion that it might be impossible to assess the executive function in an office setting for certain patients. However, it may be that it was the demands of the tests themselves that were the limiting factor, not the setting in which they were conducted (Burgess, Alderman, Volle, Benoit, & Gilbert, 2009).

One of the advances of these early observations, therefore, was the demonstration of a potential separation between emotional and what had been considered “cognitive” dysexecutive symptoms. A key study in this respect was the report of E.V.R. by Eslinger and Damasio (1985). The unfortunate man suffered bilateral ablation of orbital and ventromedial frontal cortices, and experienced “personality” and motivational changes that altered the course of his life. A critical aspect of this presentation was that his performance on IQ measures was still superior. Soon afterward, Shallice and Burgess (1991) also reported three cases who had suffered frontal lobe damage, in these cases, principally involving rostral PFC (area 10). All of these cases also had superior IQs, and did well on a range of neuropsychological tests.

However, like E.V.R., they had found it impossible to return to work. However, in their cases their deficits perhaps seemed slightly different in some respects from E.V.R.’s, seemingly most apparent in situations requiring multitasking and prospective memory. Deficits in these functions were then shown to be associated with damage to rostral (polar) damage in a group lesion study by Burgess et al. (2000), and a version of the Six Element multitasking test that was originally devised by Shallice and Burgess (1991) became part of the Behavioural Assessment of the Dysexecutive Syndrome (BADS) assessment battery.

Thus Shallice and Burgess pursued mainly the cognitive features of their patients’ deficits. However, patient E.V.R. provoked a theory that dealt with the emotional aspects of patients’ problems, called the somatic marker hypothesis (Damasio, 1996). This line of enquiry in turn led to the development of the Iowa Gambling Task (Bechara, Damasio, Tranel, & Damasio, 2005), an attempt to create an objective neuropsychological measure of the emotional consequences of frontal lobe dysfunction. This spurred further initiatives in test development and interpretation of measures of social and emotional behavior (Clark, Cools, & Robbins, 2004; Fellows & Farah, 2005; Floden, Alexander, Kubu, Katz, & Stuss, 2008; Szczepanski & Knight, 2014).

Starting from lobotomy studies and clinical cases, particularly E.V.R., evidence was consistent that damage to the orbitofrontal/ventral medial frontal regions resulted in a pattern of performance different from damage to the lateral frontal regions. These patients had notable emotional control and behavioral self-regulation problems, despite normal performance on the classic frontal lobes tests such as the WCST and the Stroop. A broader observation is that emotional and cognitive deficits after PFC damage could be dissociated from deficits in IQ and from each other with differences in anatomical localization. Newer concepts in network anatomical connectivity also supported the multiple frontal processes data.

Alexander, DeLong, and Strick (1986) reported five separate frontal cortical-subcortical circuits. Three of these circuits map onto the separate processes reported above, and
Cummings (1995), although considering all as “executive,” did differentiate them as drive, superior medial; emotional regulation, ventral medial; lateral-executive. Two of them, executive and emotional regulation, are consistent with the dual origin data. Finally, consistent with the original concept proposed by Stuss and Benson (1986) on self-awareness and self-consciousness, now supported by Pandya’s description (personal communication) of the evolutionary importance of polar Brodmann area 10, research revealed a role of area 10 in metacognition and self-awareness (Craik et al., 1999; Shammi & Stuss, 1999; Stuss, Gallup, & Alexander, 2001; see also Wheeler, Stuss, & Tulving, 1997).

Consideration of theory of mind and related process has been a useful segue for the field in considering social behavior changes in patients, and there have been related attempts looking at other aspects of social behavior (Channon, 2004).

Based on these anatomical and behavioral data, Stuss proposed there are four primary categories of frontal functions (Stuss, 2011a), each related to different brain regions: energization, task setting/planning, monitoring/checking, and metacognition. It is uncertain if more categories will be unveiled; however, it is highly likely that there are subdivisions within the categories, perhaps based on more precise localization or a hierarchical organization of representation, such as an abstract-concrete distinction (e.g., Badre, 2008).

For example, there are different types of monitoring (Petrides, 2013); there appear to be hierarchies of “task setting” (D’Esposito & Badre, 2012); and some of the functions of area 10 can be divided according to notions of stimulus-oriented versus stimulus-independent attending.

More specifically, Burgess’s “gateway hypothesis” (e.g., Burgess, Dunomtheil, & Gilbert, 2007) holds that one of the roles of rostral PFC (frontal lobe area 10) is to act as a control over attending either to stimulus-independent information (e.g., one’s inner mental chatter) or to stimulus-oriented attending (e.g., attending to the outside world) with lateral and medial aspects of area 10 playing different roles within these attending modes (for review of the evidence see Burgess and Wu, 2013). So the emphasis here is on a medial–lateral organization rather than a rostral–caudal one. Moreover, it appears that the medial PFC region that supports mentalizing is immediately causal to the ones supporting this attentional gateway (Gilbert et al., 2006), so we assume this is a functionally different portion of PFC. Overall, whichever view is taken on these functional subdivisions, nowadays most theorists would feel comfortable arguing that the “executive” system consists of multiple interacting subsystems, perhaps not all deserving the label “executive.” The question instead, is how many, and how do they interact with each other, and other systems within the brain.

New principle 5: What has been historically considered the executive system consists of multiple subsystems, with anatomical and behavioral separation. These separate systems will likely require several tests or measures to examine clinically.

Other accounts pursue the “unity” rather than the “diversity.” Most notably, John Duncan and his colleagues in Cambridge, United Kingdom, propose an account of frontal lobe function that relates to Charles Spearman’s notions of “g” or general intelligence (for review, see Duncan and Miller, 2013). On this “adaptive coding” account, PFC neurons do not have fixed functions. Instead, they have extensive and random inputs from a variety information sources, either from outside the individual (e.g., sensory or motor signals) or within it (e.g., memories and other stimulus-independent thoughts). As a result, there are large populations of neurons within PFC that can be involved in many apparently different functions, with the role that they play being determined by the behavioral context.

Duncan’s elegant experiments have provided ample evidence to suggest that at least some subregions of PFC seem to support behavior across a wide range of situations rather than being specific to one (e.g., Duncan & Owen, 2000). This wider application of functions is expressed as several domain general processes in the process specific model working together in more complex tasks (Stuss, 2011a). Another view of this more global influence type is that of the role of medial PFC in a “default mode network” (e.g., Raichle, 2015). From these works, we can propose that some parts of PFC seem to contribute to cognition that is used in many different situations.

New principle 6: Some executive processes are used in a wide variety of situations. So assessing them could be useful in predicting competence across a range of situations in everyday life.

**REPRESENTATIVENESS AS A POTENTIAL SOLUTION TO THE QUESTION “WHAT IS IT THAT THIS TEST MEASURES?”**

One expectation is that these levels of explanation (function, structure) will grow in ways that are related but independent from each other. This is already happening. For instance, prospective memory (the ability to carry out an intended action after a delay period filled with a distracting activity) is considered an executive function (although its origins were not in neuropsychology) but has grown into a mature research area which does not require reference to executive function for engagement, and also uses neuroimaging and experimental psychology techniques (see Burgess, Gonen-Yaacovi, & Volle, 2011; Gonen-Yaacovi & Burgess, 2012, for review of neuroimaging and experimental psychology findings, respectively).

The move toward using neuropsychological tests that have greater “ecological validity” (or perhaps more correctly, representativeness) was substantially driven by this closely related clinical problem: if a patient fails a test of “frontal lobe dysfunction” what will it mean in terms of their disability in everyday life, or the treatment they should receive? Partly, this issue was addressed by the development of questionnaires aimed at characterizing the range of a patient’s dysexecutive problems (e.g., the Dysexecutive Questionnaire (DEX) from the BADS (Wilson, Evans, Emslie, Alderman, & Burgess, 1998); the Behavior Rating Inventory of
Executive Function (BRIEF) (Gioia, Isquith, Guy, & Kenworthy, 2000); the Frontal Systems Behavior Scale (FrSBe) (Grace & Malloy, 2001), as well as screening measures such as the Frontal Assessment Battery (FAB) (Dubois, Slachevsky, Litvan, & Pillon, 2000).

But another trajectory has been to develop psychometric tests more obviously related to performance in everyday life, such as the BADS battery Wilson et al., 1998), the various tests more obviously related to performance in everyday life, the Frontal Assessment Battery (FAB) (Dubois, Slachevsky, Malloy, 2001), as well as screening measures such as the 2000); the Frontal Systems Behavior Scale (FrSBe) (Grace & Malec, Stuss, & Whyte, 2006; Stuss, 2011b). In research, if guided by a single construct theory, one might perhaps use more complex tests. A multiple process approach demands the construction of paradigms to measure separate processes, and the conditions that might affect the implementation of these processes. This approach is successful as well with fMRI studies (D’Esposito & Badre, 2012; Floden et al., 2011; Vallesi, McIntosh, Alexander, & Stuss, 2009; Vallesi, McIntosh, Crescentini & Stuss, 2012; Vallesi, McIntosh, Slailice, & Stuss, 2009).

The problem we face, however, is that in many cases the excellent theoretical advances that have been made in our understanding of the prefrontal cortex and the executive system over the past 50 years has not always been translated into new and better psychometric devices. Given the relatively primitive nature of the traditional tests we inherited they actually do a remarkable job in many situations. However, it has also become a principle of executive system assessment to say that whilst a deficit detected on an executive test might signify something noteworthy, the absence of such a finding does not mean that a deficit is not present. Very likely, this means that we just do not have the appropriate tools yet to detect that person’s problems (see Burgess et al., 2009).

New principle 8: If your patient does not show a deficit on your executive tests, it does not mean that they don’t have a problem. You just might not have tested the domain where their problem lies.

Our progress over the past 50 years is notable, but there is a long way to go. Partly this is because in the first couple of decades since the inception of functional neuroimaging, the streams of research and neuroimaging tended to develop quite independently. In some cases, this has probably led to errors in theorizing (see Burgess, Gonen-Yaacov, & Volle, 2012), and in many cases, paradigms that might have been usefully added to the neuropsychologist’s arsenal have not made the leap from the brain scanning suite to the clinic. There are many new areas of cognition that are now being studied by neuroimaging techniques that are highly relevant for those of us who assess executive function (e.g., future thinking), and methodological cross-talk with related fields (e.g., developmental psychology/ psychiatry, e.g., Spitzer, White, Mandy, & Burgess, 2016; Stuss, Gallup, and Alexander, 2001) is now fully under way and promises new perspectives and procedures.

Helfrich and Knight (2016) proposed the study of prefrontal cortex using oscillatory dynamics to understand its role in orchestrating networks. Furthermore, wireless functional near-infrared spectroscopy (fNIRS) has been used for the first time to measure activity in prefrontal cortex while people are carrying out executive activities (in this case, prospective memory) walking around a street environment (Pinti et al., 2015). The prospect of being able to gain real-time measurements of PFC activity while patients are engaged in the kinds of naturalistic tasks where they experience problems in everyday life opens up huge possibilities for cross-talk between neuroimaging and neuropsychology, bringing structure and function far closer together than is

IMPACT OF THEORY ON PRACTICE

There are two basic guiding approaches to measuring executive deficits: (a) replicate the situation in testing where the person has a problem; (b) adopt a perspective based on a theoretical understanding of frontal lobe functioning. Most clinicians will adopt a mixture of the two of these, filtered through the availability of the psychometric tests available to them, and the appropriateness for that client group. This also impacts one’s approach to rehabilitation. In terms of using theory as a guide, for instance, dysfunction in the single construct g would be reflected in goal neglect (Duncan et al., 2008). Goal Management Therapy was designed to treat goal neglect (Levine et al., 2000). The multiple processes approach demands use of tests that are simple and focused, or can be deconstructed (i.e., the Boston process approach) (Stuss, 2007, 2011a, 2011b), so that the specific impairment might be revealed, and treatment directed to that specific impairment (e.g., self-awareness, task setting . . .), either in isolation or in sequence. Even when treating multiple processes, the approach would require awareness of the potentially differential impact of each process.

Understanding and treating the effects of traumatic brain injury has been viewed through this lens (Cicerone, Levin, Malec, Stuss, & Whyte, 2006; Stuss, 2011b).
currently available. Thus we may be on the cusp of a new dawn for assessment of executive function.

ACKNOWLEDGMENTS

There are no conflicts of interest in relation to this manuscript. This study is a review of past literature, and there are no sources of financial support.

REFERENCES


The organisation of mind

Shallice, T., & Evans, M.E. (1978). The involvement of the frontal


