




Distinct performance profiles on the Brixton test in frontotemporal dementia

Sulakshana P. Rao¹, Ranita Nandi¹, Aparna Dutt¹, Narinder Kapur²,
Jennifer M. Harris^{3,4}, Jennifer C. Thompson^{3,4} and
Julie S. Snowden^{3,4*} 

¹Neuropsychology and Clinical Psychology Unit, Duttanagar Mental Health Centre, Kolkata, India

²Research Department of Clinical, Educational and Health Psychology, University College London, UK

³Cerebral Function Unit, Manchester Centre for Clinical Neurosciences, Salford Royal NHS Foundation Trust, UK

⁴Division of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, UK

The Brixton Spatial Anticipation Test is a well-established test of executive function that evaluates the capacity to abstract, follow, and switch rules. There has been remarkably little systematic analysis of Brixton test performance in the prototypical neurodegenerative disorder of the frontal lobes: behavioural variant frontotemporal dementia (bvFTD) or evaluation of the test's ability to distinguish frontal from temporal lobe degenerative disease. We carried out a quantitative and qualitative analysis of Brixton performance in 76 patients with bvFTD and 34 with semantic dementia (SD) associated with temporal lobe degeneration. The groups were matched for demographic variables and illness duration. The bvFTD group performed significantly more poorly ($U = 348$, $p < .0001$, $r = .58$), 53% of patients scoring in the poor–impaired range compared with 6% of SD patients. Whereas bvFTD patients showed problems in rule acquisition and switching, SD patients did not, despite their impaired conceptual knowledge. Error analysis revealed more frequent perseverative errors in bvFTD, particularly responses unconnected to the stimulus, as well as random responses. Stimulus-bound errors were rare. Within the bvFTD group, there was variation in performance profile, which could not be explained by demographic, neurological, or genetic factors. The findings demonstrate sensitivity and specificity of the Brixton test in identifying frontal lobe degenerative disease and highlight the clinical value of qualitative analysis of test performance. From a theoretical perspective, the findings provide evidence that semantic knowledge and the capacity to acquire rules are dissociable. Moreover, they exemplify the separable functional contributions to executive performance.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

*Correspondence should be addressed to Julie S Snowden, Cerebral Function Unit, Manchester Centre for Clinical Neurosciences, Salford Royal NHS Foundation Trust, Salford M6 8HD, UK (email: julie.snowden@manchester.ac.uk)

Executive impairments resulting from brain disease or injury have a major impact on functioning in daily life. Tests of executive function therefore play a crucial clinical role. They also have the potential to inform theoretical understanding of the functions of the frontal lobes (Robinson et al., 2015; Shallice & Burgess, 1996; Stuss, 2011; Stuss et al., 2000). The Brixton Spatial Anticipation Test (Burgess & Shallice, 1996, 1997) bears similarities to the traditional Wisconsin Card Sorting Test (Milner, 1963; Nelson, 1976; Stuss et al., 2000) in that it evaluates the ability to acquire and follow rules, to switch between rules, and to inhibit earlier response patterns. It has good reported validity for use in clinical populations (Van den Berg, Nys, Brands, & Ruis, 2009), and normative data are available for both younger and older adults (Bielak, Mansueti, Strauss, & Dixon, 2006; Burgess & Shallice, 1997; Spitoni et al., 2018; Van den Berg et al., 2009). It has the practical, clinical advantage of making minimal verbal demands, so is suitable for patients with communication disorders, poor literacy levels, or whose first language is not English.

Despite evidence for the test's sensitivity to focal frontal lesions (Burgess & Shallice, 1996, 1997; Reverberi, D'Agostini, Skrap, & Shallice, 2005), there has been surprisingly little systematic exploration of Brixton performance in the prototypical bilateral frontal lobe disease: behavioural variant frontotemporal dementia (bvFTD) (Neary et al., 1998; Rascovsky et al., 2011). bvFTD is a neurodegenerative disorder, with onset typically between 50 and 70 years, characterized by radical alterations in behaviour. It might be anticipated that the Brixton test would be highly sensitive to bvFTD. Yet, in a study of 11 bvFTD patients (Hornberger et al., 2010) overall Brixton performance differed only marginally from that of controls. The test also failed to distinguish bvFTD from a non-progressive disorder that mimics bvFTD (Hornberger, Piguet, Kipps, & Hodges, 2008), the so-called FTD phenocopy syndrome (Kipps, Hodges, Fryer, & Nestor, 2009). Given the known heterogeneity within bvFTD, evaluation of performance in a much larger cohort is required.

Aside from the paucity of data relating to test sensitivity, there is currently little information about the specificity of the Brixton test in identifying frontal lobe degenerative disease. Comparative studies of bvFTD and Alzheimer's disease (AD) reported no difference in Brixton performance in the two patient groups (Buhl, Stokholm, & Gade, 2013; Hornberger et al., 2010; Wong, Flanagan, Savage, Hodges, & Hornberger, 2014). AD is not, however, immune from frontal lobe involvement. Semantic dementia (SD), which is pathologically related to bvFTD, represents a purer comparison.

Semantic dementia is a disorder of conceptual knowledge, associated with prominent atrophy of the temporal lobes (Hodges & Patterson, 2007; Neary et al., 1998). A salient feature is difficulty in naming and understanding words, leading to the alternative designation 'semantic variant primary progressive aphasia' (Gorno-Tempini et al., 2011). The conceptual disorder is not, however, confined to language but progressively involves loss of understanding of information arising from all sensory modalities. Patients no longer recognize familiar faces (Josephs et al., 2008; Snowden, Thompson, & Neary, 2004) and voices (Luzzi et al., 2017) or understand the meaning of common objects (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Snowden, Griffiths, & Neary, 1994), non-verbal environmental sounds (Bozeat et al., 2000; Goll et al., 2010), smells (Luzzi et al., 2007), and tastes (Piwnica-Worms, Omar, Hailstone, & Warren, 2010). A pertinent question is whether the pervasive, multimodal semantic loss extends to the ability to identify, understand, and follow rules and to recognize rule shifts. This question has theoretical importance for understanding the semantic disorder in SD and the nature of conceptual knowledge. It also has clinical relevance with regard to the specificity of the Brixton test as an executive test of frontal lobe function.

The Brixton test is potentially informative because performance can be evaluated along multiple dimensions. The test comprises 10 numbered circles, one of which is coloured blue (Figure 1). The participant is required to anticipate the position of the blue circle on each successive page of a 55-page test booklet by identifying patterns that vary over the course of the task. The test incorporates nine patterns/rule shifts: three '+1' rules (e.g., the blue circle moves from position 1 to 2 to 3 to 4), two '-1' rules (e.g., it moves down from 4 to 3 to 2), three alternating rules (it switches respectively between 5 and 10, 4 and 10 and 8 and 9), and one constant rule in which the blue circle remains in the same position (9 to 9 to 9 to 9). In the standard procedure, performance is measured in terms of the total number of errors, the higher the score, the poorer the performance. The test also offers the opportunity to evaluate the effect on performance of the nature of the underlying rule and, as suggested by Crescentini et al. (2011), to distinguish between acquisition and maintenance of rules.

The test also readily lends itself to an analysis of the nature of errors. In their original studies, Burgess and Shallice (1996, 1997) distinguished between (1) perseverations, (2) plausible errors, based on an earlier or novel but viable rule, and (3) implausible responses, unrelated to a rule or previous response. The latter error type distinguished best patients with anterior from those with posterior hemisphere brain lesions with the anterior lesion group showing more implausible responses. Reverberi, D'Agostini, et al. (2005) and Reverberi, Lavaroni, Gigli, Skrap, and Shallice (2005) refined the error analysis further. A motivation was the recognition that responses that fall within the broad rubric of 'perseverations' are not all the same. They might constitute the repetition of the preceding response or the immediately preceding rule or a return to an earlier rule that is not the immediately preceding one.

Further distinctions between perseverations might plausibly be made. Consider the situation of the first rule shift in the Brixton test: the blue circle, having appeared in consecutive positions 2, 3, 4, 5, 6, now, instead of the anticipated position 7, moves back to position 5, and then consecutively to positions 4, 3, 2, and 1. That is, there is a shift from an '+1' to a '-1' rule. A hypothetical patient A, who fails to recognize the rule shift, might incorrectly select 6 when shown the blue circle in position 5, 5 when it is in position 4, 4 when it is in position 3, and so on. The patient is showing perseveration of the preceding +1 rule. Patient B, in contrast might, following the rule change, select position 8, then 9, then 10, despite the blue circle's respective positions of 5, 4, and 3. Patient B, like patient

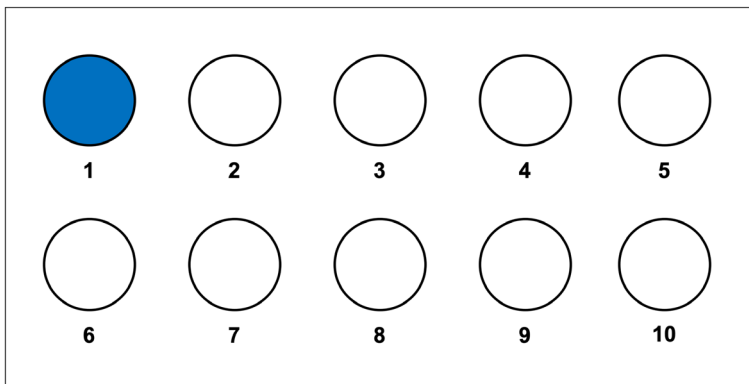


Figure 1. Brixton spatial anticipation Test. [Colour figure can be viewed at wileyonlinelibrary.com]

A, is showing perseveration of the preceding +1 rule. The difference is that whereas patient A's response for each trial is guided by the position of the blue circle, patient B's response is not. Hence, it would be important to distinguish between rule-based perseverative errors that are dictated by the position of the blue circle and rule-based perseverations that are dislocated from/independent of the blue circle's position but are influenced by the participant's own preceding response.

A further potentially important distinction, recognized by Burgess and Shallice (1996), is between logically plausible and implausible (random) incorrect responses. Consider the sequence 3, 2, 1 (−1 rule). Thereafter, the blue circle moves back to position 10. Yet, selection of position 6 is plausible, on the basis of a putative rule 'move one space in an anticlockwise direction' (see Figure 1). By contrast, selection of position 8 is implausible, since it bears no relationship to any viable rule and thus can be considered 'random'.

An additional type of response warrants consideration. Environmentally driven, stimulus-bound behaviour is a known feature of frontal lobe disease (Besnard et al., 2011; Lhermitte, 1983; Shallice, Burgess, Schon, & Baxter, 1989). It has been reported in bvFTD (Ghosh & Dutt, 2010; Ghosh, Dutt, Bhargava, & Snowden, 2013) and was included as a supportive feature in early diagnostic criteria for FTD (Neary et al., 1998). This raises the possibility of occurrence of 'stimulus-bound' error responses, in which the participant's response matches the position of the blue circle. The Brixton test includes one constant rule (9, 9, 9), so for those items, a response identical to the stimulus position is correct, but responses that match the blue circle's position might potentially arise erroneously elsewhere in the test.

The current study compared performance on the Brixton test in a consecutive series of patients diagnosed with bvFTD or SD. The study had the following principal aims: (1) to determine the value of the Brixton test in distinguishing between bvFTD and SD, disorders respectively of the frontal and temporal lobes, (2) to better understand the basis for impaired performance through systematic exploration of performance characteristics, and (3) to explore the value of error analysis in distinguishing frontal lobe degenerative disease. The findings ought to have clinical implications for understanding of bvFTD and SD and may inform theoretical understanding of executive functions and conceptual knowledge.

Methods

Participants

The study group comprised patients referred to a specialist diagnostic unit for early-onset dementias. The criteria for selection were that the patients (1) had a clinical diagnosis of bvFTD or SD and met established criteria for those conditions and (2) they had undergone neuropsychological evaluation that included the Brixton test.

Patients' overall functional capacity was evaluated using the revised Clinical Dementia Rating (CDR) scale, modified for use in FTD (Knopman, Weintraub, & Pankratz, 2011). The Mini-Mental State Examination provided a general measure of cognitive functioning. Background information for patients also included scores on a standard graded difficulty naming test (Graded Naming Test; McKenna & Warrington, 1980) and a less demanding locally developed naming test involving 40 pictures of highly familiar items (10 animals, 10 fruits/vegetables, 10 articles of clothing, 10 objects) drawn from the Snodgrass and Vanderwart (1980) corpus. This undemanding naming test yields ceiling-level

performance in healthy controls. A word–picture matching test using those same 40 items was also included.

One hundred and 10 patients fulfilled the criteria for the study: 76 with bvFTD and 34 with SD. Patients with bvFTD had presented to medical attention with a history of behavioural change, which included social disinhibition, reduced motivation and lack of self-care, loss of sympathy and empathy, increased preference for sweet foods, and repetitive behaviours and stereotypies. All patients fulfilled contemporary criteria for bvFTD (Rascovsky et al., 2011). Most bvFTD patients were physically well. However, 14 (18%) showed physical signs of amyotrophic lateral sclerosis (ALS), which is known to co-occur with bvFTD in a proportion of patients (Burrell et al., 2016; Neary et al., 1990; Saxon et al., 2017). Clinical brain imaging reports, available for 66 (87%) of cases, supported the clinical diagnosis. 55 bvFTD patients had been screened for genetic mutations linked FTD. Mutations were detected in 13 (24%): Nine had repeat expansions in the *C9orf72* gene, three in the *progranulin* gene, and one in the *MAPT* gene.

The SD patients presented to medical attention with symptoms indicative of semantic loss: problems in word comprehension and naming, difficulties in recognition of faces and objects. All showed temporal lobe atrophy; in 23 cases, this was most marked on the left side; in nine cases, it was more marked on the right side; and in two cases, there was no asymmetry of atrophy. All patients met criteria for SD (Neary et al., 1998). Most also fulfilled criteria for semantic variant of primary progressive aphasia (Gorno-Tempini et al., 2011), although the patients with right predominant atrophy had face recognition problems as an early symptom. All patients showed a multimodal disorder of semantic knowledge at the time of testing. In keeping with evidence that SD has a low familial incidence (Hodges et al., 2010), no patient had an identifiable genetic mutation.

Brixton tests had been administered by trained neuropsychologists with many years of clinical experience of assessing patients with forms of dementia. If patients had difficulty with verbal comprehension of test instructions, examiners adopted a visual gestural technique of pointing to the blue circle and then saying 'where next?' whilst pointing to the empty circles and indicating the next page. To circumvent potential problems in working memory, the examiner repeated the instruction 'where will be the blue one be next?' throughout the test.

Patients, or their consultees, provided written consent for clinical data to be used for research purposes. All patients are classed as 'vulnerable adults', demonstrating on cognitive test impairments in language and/or executive functions that potentially compromise decision-making capacity. Nevertheless, fundamental principles of the Mental Capacity Act (2005) are that people, including vulnerable adults, 'have the right to make their own decisions' and 'all practicable help must be given to enable them to do so'. Both the patient and their consultee (normally their spouse) were involved in the consent process, with information being presented in simplified form to the patient. Most patients signed their own consent form, whilst carers acted as a safeguard, confirming that the patient's agreement was consistent with their likely decision before they became ill.

Procedure

Brixton score sheets were extracted from the patients' clinic files by one author, and photocopied, ensuring that all personal and diagnostic identifying information was erased, and the copies were allocated a unique code. Coding of score sheets was then carried out, in a blinded fashion, by other assessors without knowledge of the patients or their diagnoses. A set of Brixton score sheets from patient groups not included in the study

were also examined in order to identify ambiguities in scoring of errors and to refine the system for classification, which occurred through an iterative process.

Quantitative scoring

1. An overall error score was calculated, in accordance with standard test instructions.
2. Performance on the first two trials following a rule shift (reasoning trials) and subsequent trials (post-reasoning trials) was examined separately. This follows the distinction made by Crescentini et al. (2011) between rule acquisition, when people are using reasoning/problem-solving skills to acquire the rule, and rule maintenance, when it would be expected that the rule would have been learned. It was anticipated that people with a degenerative dementia might be slower to acquire rules than the general population, so allowed for two trials, rather than one, for reasoning of the rule.
3. It was recorded whether or not the patient had successfully acquired each rule, operationally defined as two consecutive correct responses in post-reasoning trials. Set 5 (reverse sequence 10, 9, 8) is, however, short, with only 1 post-reasoning trial, (3 trials in total), so was omitted from this analysis.

Error analysis

The error classification followed the taxonomy described by Reverberi, D'Agostini, et al. (2005), Reverberi, Lavaroni, et al. (2005) with modifications and elaboration to allow for alternative error types that might potentially be applicable to bvFTD. In particular, a separation was made between rule-based responses that are influenced by the position of the blue circle (stimulus-related) and those that are not (stimulus-unrelated), as well as between responses that have no link to an underlying rule (non-rule-based). Error analysis was based on all trials, reasoning and post-reasoning.

1. Stimulus-related errors

- a. *Stimulus-bound*: The participant responds by pointing to the blue circle. This is not applicable for set 8, where the correct response is constant (i.e., 9, 9, 9) or trial 49 (see below).
- b. *Logical Inference*: A potentially plausible error response. Instances classifiable as a logical inference are as follows: (1) Trial 5: response 10 (consistent with putative 'move in a clockwise direction' rule); (2) Trial 11: response 6 (consistent with putative 'move in an anticlockwise direction' rule); (3) Trial 20: response 1 (consistent with alternation rule); (4) Trial 27: response 6 (consistent with 'anticlockwise direction' rule); (5) Trial 42: response 3 or 10 (consistent with alternation rule), response 8 (consistent with -1 rule); and (6) Trial 49: response 8 (consistent with constant rule).
- c. *Perseveration of rule type A*: Continued application of the rule from the immediately preceding set OR applied in the immediately preceding incorrect response. Note that for alternating sets, a repeated switch between adding and subtracting 1 (e.g., for set 7, responses 5,9,5,9 when shown the blue circle in positions 4,10,4,10) was classified as perseveration of rule type A after the first instance of the addition and subtraction.

- d. *Same rule*: Application of a rule that is not the immediately preceding rule or the rule applied in the immediately preceding response. Same rule errors apply regardless of whether or not they were successfully acquired when first exposed. Shifts back to an earlier rule mid-set, after the patient has achieved the current rule, constitute *same rule diversions*.

2. Rule-based, stimulus-unrelated errors

- e. *Perseveration of response*: Repetition of an incorrect response that is identical to the immediately preceding incorrect response. (e.g., continued production of the response '8'). Note that this error type does not apply to the first instance of the incorrect response.
- f. *Perseveration of rule type B*: Continued application of an incorrect rule, where responses relate to the position of the preceding response rather than the position of the blue circle. The 'rule' applied may relate to a previously acquired rule, a previously exposed rule not acquired, or an idiosyncratic rule.

3. Non-rule-based, non-stimulus-related errors

- g. *Random responses*: Response unrelated to a rule, preceding response, or blue circle position. Apparently random responses occurring mid-set after a rule has been achieved are classified as *Random diversions*.
- h. *Omissions*: No response is given.

Errors might feasibly be interpreted in more than one way. For example, perseveration of a preceding response (e.g., 8, 8, 8) in set 9 also constitutes perseveration of the preceding 'constant' rule. In order to deal with such ambiguities and ensure consistency of classification across the cohort, a flow chart for error classification was developed, shown in the Appendix.

Statistical analysis

The G* Power program (Faul, Erdfelder, Lang, & Buchner, 2007) was used to perform a sensitivity power analysis. All other analyses were carried out using IBM SPSS Statistics version 25.

t-Tests were applied for group comparisons where data were normally distributed. Effect sizes were calculated using Cohen's *d* formula $(M1 - M2)/S_p$, where *M* = group mean and *S_p* = pooled standard deviation. 95% confidence intervals were calculated using the procedure of Grissom and Kim (2005), described by Fritz, Morris, and Richler (2012).

Non-parametric analyses (Mann-Whitney *U*-tests and Spearman's correlations (*r_s*)) were used for Brixton and other cognitive data because of their skewed distribution. The Bonferroni-adjusted *p*-values were applied in analyses involving multiple comparisons. Effect sizes were calculated using the formula $r = z/\sqrt{n}$.

The chi-squared tests, or Fisher's exact test where appropriate, were used for categorical data. Effect sizes were calculated using Cramer's *V*: $\phi = \sqrt{(\chi^2/N)}$, with bootstrapping to determine 95% confidence intervals (CI).

Results

A sensitivity power analysis, based on $\alpha = 0.05$, $1 - \beta = 0.80$, and group sizes $N_1 = 76$ and $N_2 = 34$, yielded effect size $d = 0.53$, which corresponds to a medium effect

according to Cohen's classification for d (0.2, small; 0.5, medium; and 0.8, large). On this basis, a smallest effect size of interest for r and ϕ was identified as 0.3, as this corresponds to a Cohen standard medium effect (0.1, small; 0.3, medium; and 0.5, large).

All Brixton analyses, aside from those relating to genetics, involved the complete cohort $N_1 = 76, N_2 = 34$, whereas there were some missing values from background data.

Demographics and background cognitive data

The bvFTD and SD groups did not differ in gender distribution, age, duration of illness at test, CDR, or MMSE scores (Table 1). In most patients in both groups, CDR scores were consistent with a mild–moderate overall severity of dementia.

In keeping with expectation, there were highly significant differences in naming and word comprehension scores, superior performance being demonstrated by bvFTD patients (Table 2). The median performance in bvFTD fell within the normal range on the demanding Graded Naming Test (McKenna & Warrington, 1980) and approached ceiling level on the less demanding naming test and word–picture matching test. By contrast, SD patients were severely impaired, with scores on the Graded Naming Test at floor level.

Brixton performance: quantitative analysis

Number of errors

Comparisons between bvFTD and SD revealed the converse pattern to that shown for naming and comprehension tests. The performance of bvFTD patients, measured by the total number of errors, was significantly poorer than that of SD patients (Table 2).

Figure 2 shows the distribution of performance in the two patient groups, with classifications based on scaled scores and published test norms (Burgess & Shallice, 1997). More than half of bvFTD patients (53%) performed in the poor–impaired range, contrasting with only 6% of SD patients. SD patients showed the converse pattern of performance: 53% achieved scaled scores of 7 (high average) or above, compared with only 5% of bvFTD patients. In bvFTD, 34% of patients made more than 31 errors, in keeping with a scaled score of 1, compared with 3% of SD patients. Performance for individual patients, shown in Figure 3, illustrates the wide variation in test scores, which is most marked in bvFTD.

Reasoning versus post-reasoning trials

When responses were separated into reasoning trials, operationally defined as the initial two responses following a rule shift, and post-reasoning trials, the pattern was the same. bvFTD patients performed significantly more poorly than SD patients on both reasoning ($U = 535, \text{bonf } p < 0.001, r = .47$) and post-reasoning trials ($U = 388, \text{bonf } p < 0.001, r = .56$).

Rule attainment

Table 3 shows the percentage of the two patient groups who attained each rule. A lower percentage of bvFTD patients than SD patients acquired rules, irrespective of rule type. In bvFTD, only on +1 rules did the percentage reach above 75%. By contrast, in SD, the percentage reached 85% or above for all rules. The rule that bvFTD patients found hardest to acquire was Rule 7, which involves alternation between positions 10 and 4. Notably,

Table 1. Demographic characteristics and general functioning

Group	bvFTD	SD	χ^2 , t (df), Mann-Whitney U-test	p value	Effect size ϕ , d or r	95% CI Lower-upper
Males: females	47:29	20:14	$\chi^2 = 0.09$.76	$\phi = .03$	-0.22 to 0.17
Age at test mean (standard deviation)	62 (9)	64 (7)	t(85.5) = -1.15	.25	d = .21	-0.20 to 0.62
Years of illness at test mean (standard deviation)	3.3 (2.5)	3.3 (1.8)	t(103) = -0.02	.98	d = .005	-0.41 to 0.42
CDR revised/24 ^a median (range)	9 (3-16)	8.5 (4.5-12)	U = 1,090	.19	r = .12	
MMSE/30 mean (standard deviation)	24 (5)	22 (6)	t(97) = 1.7	.09	d = .38	-0.07 to 0.83

Note. Df = degrees of freedom; CI = confidence interval; CDR = modified Clinical Dementia Rating Scale Sum of Boxes (Knopman et al., 2011).

^aHigher score = greater severity

Table 2. Performance on Naming, Comprehension, and Brixton tests

Test	N ₁ /N ₂	bvFTD	SD	Mann–Whitney	p value	Effect size <i>r</i>
Graded Naming/30 median (range)	59/34	15 (3–28)	0 (0–5)	<i>U</i> = 10.0	<.0001 ^a	.83
Manchester naming/40 median (range)	63/31	38 (19–40)	22 (0–37)	<i>U</i> = 112	<.0001 ^a	.72
Word–picture match/40 median (range)	57/30	40 (37–40)	38 (18–40)	<i>U</i> = 320	<.0001 ^a	.62
Brixton total errors/54 Median (range)	76/34	25 (11–52)	13 (8–32)	<i>U</i> = 348	<.0001 ^b	.58
Brixton scaled score/10 median (range)	76/34	3 (1–7)	7 (1–9)	<i>U</i> = 373	<.0001 ^b	.58

Note. N₁ N₂ number of bvFTD (N₁) and SD (N₂) patients tested.

^aPoorer performance in SD.; ^bPoorer performance in bvFTD.

almost half of bvFTD patients had difficulty acquiring the constant rule (set 8), for which the correct response is 9 on successive trials.

Brixton test: Item 1 response

There is no correct answer for item 1 of the Brixton test. Nevertheless, the number 2 position is the most plausible and might be anticipated to be the most likely ‘default’ response. 42 of 76 (55%) bvFTD patients selected number 2 as their initial response compared with 25 of 34 (74%) SD patients. The findings point to a non-significant trend towards more non-conventional responding in bvFTD ($\chi^2(1) = 3.3, p = .07, \phi = .17, CI: -0.01$ to 0.34).

Correct responses occurring by chance

Sporadic correct responses, arising amidst a run of error responses, might be construed as arising by chance. Such ‘correct-by-chance’ responses were significantly more common in

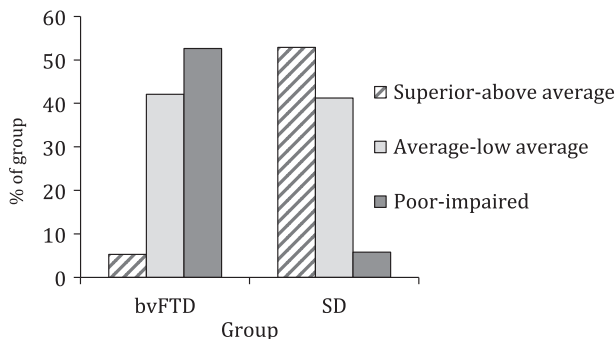


Figure 2. Percentage of the patient group performing at the superior–above-average level (scaled score 7–10 according to test norms), average–low average level (scaled score 6–4), and poor–impaired level (scaled score 3–1).

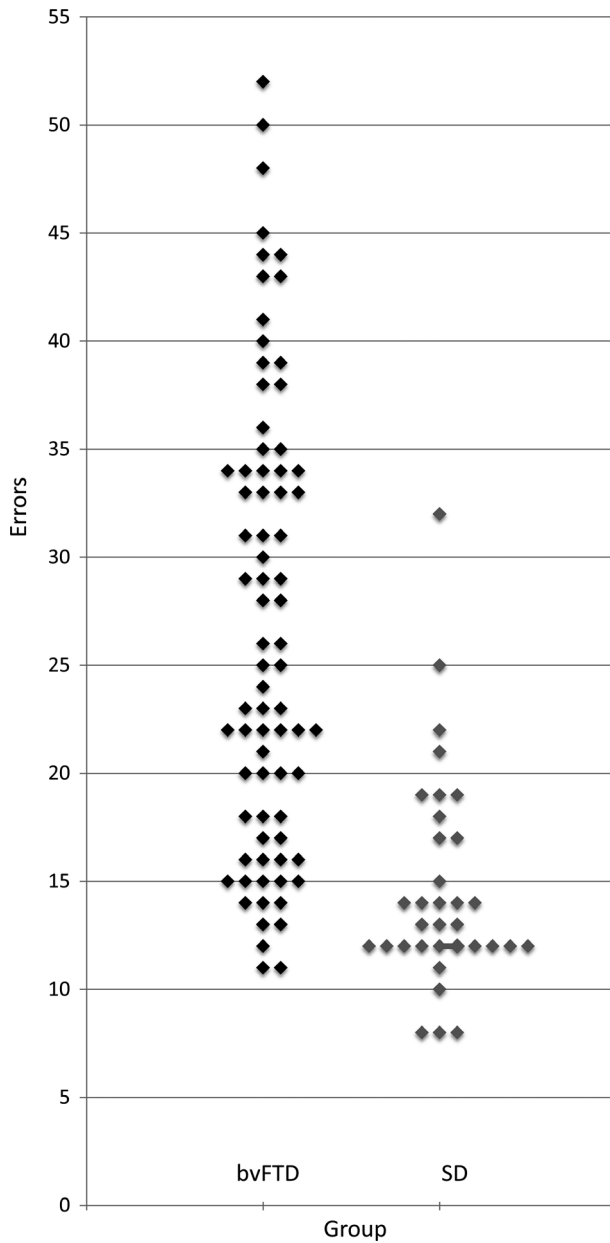


Figure 3. Scatterplots showing Brixton error scores for individual bvFTD and SD patients (maximum possible error score = 54).

the bvFTD group ($U = 784.5, p < .0001, r = .39$). They occurred in 32 of 76 (43%) bvFTD patients but only in 1 of 34 (3%) SD patients.

Brixton performance: analysis of errors

Table 4 shows the breakdown of errors in the two groups.

Table 3. Percentage of patients who acquired rule

Brixton test condition			% of group acquiring rule		Group comparison			
Set	Rule	Example	bvFTD (n = 76)	SD (n = 34)	χ^2 (df = 1)	p_{bonf} value	Effect size ϕ	95% CI Lower–upper
1	Add I	2,3,4	84	91	FExact	1.0	.09	–0.09 to 0.24
2	Subtract I	3,2,1	66	94	9.9	.02	.30	0.16–0.42
3	Alternate	10,5,10	58	97	17.2	<.001	.40	0.28–0.51
4	Add I	8,9,10	87	100	FExact	.24	.21	0.14–0.28
6	Add I	2,3,4	78	100	9.0	.02	.29	0.20–0.37
7	Alternate	4,10,4	37	88	24.9	<.001	.48	0.32–0.62
8	Constant	9,9,9	58	94	14.4	.001	.36	0.22–0.48
9	Alternate	8,9,8	49	97	23.8	<.001	.47	0.34–0.57

Note. *df* = degrees of freedom; FExact = Fisher's exact test; p_{bonf} value = Bonferroni corrected for multiple comparisons.

Stimulus-related errors

Stimulus-bound errors were rare in both groups. Perseverations of rule type A were numerically more common in bvFTD, although the group difference did not reach corrected levels of significance. Same rule errors occurred with equal frequency in both groups. That included diversions to an alternative rule after a rule had been acquired.

Rule-based errors unconnected to the stimulus

Perseverations of a preceding response and perseverations of a rule unconnected to the stimulus (blue circle) were significant indicators of bvFTD (Table 4). Perseveration of rule type B errors was exceedingly rare in SD and never constituted the application of an idiosyncratic rule, not previously exposed.

Table 4. Error breakdown in bvFTD and SD

Error type	bvFTD		SD		Mann–Whitney		Effect size <i>r</i>
	Median	Range	Median	Range	<i>U</i>	p_{bonf} value	
Stimulus-related errors (response related to position of blue circle)							
Stimulus-bound	0	0–13	0	0–2	1,108	.74	.16
Logical inference	0	0–4	1	0–3	998	.28	.20
Perseveration of rule A	3	0–32	2	0–24	929	.14	.23
Same rule	9	0–16	8	5–15	1,122	1.0	.11
Rule-based stimulus-unrelated errors (response related to previous response not blue circle)							
Perseveration of response	0	0–41	0	0–3	936	.03	.28
Perseveration of rule B	0	0–37	0	0–1	931	.02	.29
Non-rule-based, non-stimulus-related errors							
Random response	3	0–48	1.5	0–10	818	.02	.30
Omissions	0	0–19	0	0–1	1,278	1.0	.03

Note. Group comparisons that remain significant when corrected for multiple comparisons are shown in bold.

Non-rule-based responses

bvFTD patients produced more random errors than SD patients. Failures to respond were rare and did not distinguish the groups.

Interrelationship between error types

Notwithstanding the significant group differences between bvFTD and SD, there was substantial variation within the bvFTD cohort. For example, a majority of bvFTD patients made no rule-based, stimulus-unrelated errors at all (hence the median error score of 0), yet at the other extreme, such errors dominated patient performance. Table 5 shows the inter-correlation between error types in bvFTD. There were strong associations between stimulus-unrelated responses (perseveration of response and perseveration of rule type B) and random errors. Stimulus-unrelated and random errors showed inverse correlations with stimulus-related (perseveration of rule type A and same rule) errors. In the SD group, there were no significant correlations between error types.

Brixton performance in bvFTD: relationship to clinical factors

Brixton performance was impaired in a high proportion of bvFTD patients but not all. Correlational analyses were carried out with a view to identifying factors influencing performance.

Total error score

There was no significant correlation between the Brixton total error score in bvFTD and patients' age ($r_s = .09, p = .47$) or duration of illness ($r_s = .01, p = .93$). The presence of ALS did not affect performance ($U = 425, p = .90, r = .01$), nor did the presence of a

Table 5. Correlation between error types produced in bvFTD

Error type	Stimulus (blue circle)-related errors				Rule-based stimulus-unrelated errors		Non-rule-based errors	
	SB	LI	P-rul A	SR	P-resp	P-rul B	Ran	Om
SB	–	–0.06	–0.06	0.04	0.24	0.25	0.27	0.02
LI	–	–	–0.07	0.16	–0.19	–0.26	0.004	–0.04
P-rul A	–	–	–	0.14	–0.25	–0.14	–0.48***	–0.13
SR	–	–	–	–	–0.31	–0.32*	–0.20	0.08
P-resp	–	–	–	–	–	0.36**	0.52***	–0.15
P-rul B	–	–	–	–	–	–	0.46***	–0.04
Ran	–	–	–	–	–	–	–	–0.07
Om	–	–	–	–	–	–	–	–

Note. Correlations of interest, which are >0.3 and which remain significant when corrected for multiple comparisons, are shown in bold.

SB = stimulus bound; LI = logical inference; P-rule A = perseveration of rule type A; SR = same rule; P-resp = perseveration of response; P-RuleB = perseveration of rule type B; Ran = random; Om = omission.

*_{bonf} $p < 0.05$; **_{bonf} $p < 0.01$; ***_{bonf} $p < 0.001$.

genetic mutation (present in 13 of 55 screened patients; $U = 260, p = .79, r = .04$). There was, however, a significant association between Brixton scores and CDR and MMSE scores. Poorer Brixton performance was associated with more severe CDR ($r_s = .59, p < .001$) and lower MMSE scores ($r_s = -.61, p < .001$). This contrasts with the finding in SD of no significant relationship between Brixton errors and either CDR ($r_s = .27, p = .13$) or MMSE ($r_s = -.29, p = .15$) scores.

Error types

In keeping with findings for the total error score, there was no significant relationship between error subtypes and demographic factors, including illness duration. There were, however, significant inverse correlations, in bvFTD only, between both CDR and MMSE scores and the frequency of perseveration of rule type B errors ($r_s = .49, \text{bonf}p < .0001$ and $r_s = -.58, \text{bonf}p < .001$, respectively). There was a positive correlation between MMSE scores and the number of logical inference errors ($r_s = .40, \text{bonf}p = .004$).

Discussion

The study examined the potential of the Brixton test to distinguish between bvFTD and SD, neurodegenerative disorders with predominant involvement respectively of the frontal and the temporal lobes. The study sought also to improve understanding of the basis for performance failure through systematic analysis of performance characteristics and nature of errors.

Sensitivity and specificity

The Brixton test elicited striking differences in performance in patients with bvFTD and SD. More than half of the large bvFTD cohort performed within the poor–impaired range, according to standard published classification (Burgess & Shallice, 1997), contrasting with 6% of SD patients. Group differences were comparable for both reasoning and post-reasoning trials, excluding the possibility that bvFTD patients were simply slower to acquire rules. When bvFTD patients had difficulty shifting to a new rule, then that difficulty was typically maintained across all trials within a set. Difficulty was, moreover, apparent across all rule types, albeit with varying magnitude. Performance was particularly poor for alternating rules, paralleling previous findings in FTD of impairment in object alternation tasks (Freedman et al., 2013). It was relatively better for the +1 rule, which is the first rule applied in the test.

The SD patients were matched to the bvFTD patients for gender, age, duration of symptoms, CDR, and MMSE scores, so that group differences could not be attributed to demographic factors, stage of illness, or severity of disease. Indeed, the SD patients, predictably, performed significantly more poorly than the bvFTD patients on measures of naming and word comprehension. The findings suggest that the Brixton test is both sensitive to bvFTD and has specificity in distinguishing a predominantly ‘frontal’ from a ‘temporal’ lobe neurodegenerative disease. Such a conclusion aligns with findings from an eye-tracking task in bvFTD and SD using an analogue of the Brixton procedure (Primativo et al., 2017).

The demonstration of preserved Brixton performance in SD patients is instructive. Patients with SD exhibit widespread loss of conceptual understanding about the world

that includes difficulty understanding the meaning of words, objects, and other sensory stimuli. It is a pertinent question whether the conceptual loss extends to the understanding of rules governing the movement of a spatial stimulus. On the contrary, the data point to a striking area of preservation. Overall, SD patients were able to grasp the nature of the task, despite problems in language comprehension, and they were quick to identify rules and detect rule shifts. The findings complement those showing other domains of relative preservation in SD. Patients may display good numerical skills (Crutch & Warrington, 2002; Papagno, Semenza, & Girelli, 2013), although preservation of number knowledge is not absolute (Julien, Thompson, Neary, & Snowden, 2008; Luzzi, Cafazzo, Silvestrini, & Provinciali, 2013). Perhaps of particular relevance, they commonly enjoy puzzles such as jigsaws (Green & Patterson, 2009) and number games such as sudoku (Papagno et al., 2013), which involve systematization and logical thought. SD patients are known to pursue such activities until relatively late in the course of disease. On a practical level, the findings in SD highlight the clinical value of an executive test such as the Brixton that minimizes language demands and hence can be disentangled from patients' pervasive difficulty in word semantics.

The striking group differences demonstrated by the current study contrast with earlier findings that the Brixton test did not distinguish bvFTD from AD (Buhl et al., 2013; Hornberger et al., 2010; Wong et al., 2014). There are plausible reasons why that might be so. Group sizes were small in those earlier studies so might not be representative of the bvFTD population. AD patients, particularly the elderly, are not immune from executive impairments (Swanberg, Tractenberg, Mohs, Thal, & Cummings, 2004). Moreover, the Brixton test makes additional demands on spatial functioning and on working memory, both of which are commonly affected in AD (Snowden et al., 2007; Stopford, Snowden, Thompson, & Neary, 2007) but are spared in SD. In the comparative studies of bvFTD and AD, group comparisons were based on test scores alone. It is reasonable to speculate that qualitative analysis of errors might have revealed a distinct performance profile in the two groups. In a study in which performance in bvFTD and AD was compared on the Tower subtest of the Delis–Kaplan Executive Function System (D-KEFS) battery (Delis, Kaplan, & Kramer, 2001), overall scores did not distinguish bvFTD from AD (Carey et al., 2008). By contrast, the pattern of errors discriminated the groups: Rule violation errors occurred significantly more often in bvFTD. Those authors' findings highlighted the importance of characterizing component processes of performance failure in the cognitive assessment of FTD and AD.

Error analysis

The error analysis in the current study elicited differences between bvFTD and SD. In particular, bvFTD patients were more likely to make perseverative and random errors. The group difference in perseverative errors was particularly marked for perseverations of the preceding response or perseveration of the preceding rule type B, indicating specificity of these types of perseverative errors for frontal lobe disease. That is, bvFTD, but not SD, patients had a tendency to become disengaged from the blue circle stimulus. Notably, same rule errors, which might also be construed as a form of perseveration since they constitute a shift back to an earlier rule, did not differentiate the groups. The findings reinforce the view (Reverberi, D'Agostini, et al. (2005), Reverberi, Lavaroni, et al. (2005) that errors that can broadly be classed as perseverations are not all the same. Indeed, whilst there was a strong inter-correlation in the bvFTD group between perseverations of

a response and perseverations of a rule type B, there were no, or inverse, correlations with perseverations of rule type A and same rule errors.

Perseverations of a response and perseverations of a rule type B were also strongly associated with random errors. What those three error types have in common is that they reflect a failure of adherence to task goal: The participant, although required to predict the position of the blue circle, responds in a way that has no bearing on the blue circle at all. By contrast, same rule and perseveration of rule type A errors, although pointing to failure of set shifting and rule abstraction, nevertheless provide evidence of attention to the blue circle and hence regard to task goal. The findings point to a dissociable aspects of executive performance breakdown in bvFTD.

The finding of a correlation between stimulus-unrelated responses in bvFTD and both CDR and MMSE scores might reasonably suggest that these errors are a marker of disease severity. This is a plausible explanation although there is a need for caution in automatically interpreting them as such. A patient who disregards the constraints imposed by the Brixton task (i.e., the position of the blue circle) might similarly disregard task goals in the MMSE. For example, a bvFTD patient when asked for today's date might instead give the date of their birthday. Clinically, such errors do indeed occur. They show that MMSE performance is coloured by patients' executive impairments and so should not be considered an independent marker of disease severity. Similarly, the CDR is not wholly independent of executive skills. Notably, there was no significant correlation between stimulus-unrelated errors and disease duration, which is one marker of stage of illness. The possibility exists that distinct error profiles reflect not merely severity of disease but also phenotypic variation within bvFTD.

Responses that are identical to the position to the blue circle were included as a potential error type on the grounds that stimulus-bound behaviours are an established feature of frontal lobe disease (Ghosh et al., 2013; Lhermitte, 1983). Interestingly, these errors were rare and – as noted above – some responses bore no relation to the blue circle at all. Reinforcing the point, bvFTD showed difficulty acquiring the 'constant' rule whereby the correct response location matched that of the blue circle. Thus, in this relatively large bvFTD cohort, stimulus-boundedness was not a notable performance characteristic.

The current study measured separately performance for the first trials following a rule shift (reasoning trials) and subsequent trials (post-reasoning trials). This differentiation was drawn from the distinction made by Crescentini et al. (2011) between rule acquisition and rule following. Those authors showed, in an fMRI study of healthy adults who carried out a computerized adaptation of the Brixton test, differential regions of activation during rule acquisition and rule following. In the current study, the distinction between initial and later trials following a rule shift proved relatively uninformative, presumably by virtue of the fact that rule acquisition in bvFTD was so poor. For many bvFTD patients, all trials constituted reasoning or rule acquisition trials.

Clinical significance and ecological validity

The study findings highlight the value of the Brixton test in the assessment of neurodegenerative disease of the frontal lobes. This conclusion is clinically important. In recent years, there has been increasing concern regarding the ecological validity of traditional, clinic-based tests of executive function like the Brixton (Chan, Shum, Touloupoulou, & Chen, 2008; Manchester, Priestly, & Jackson, 2004; Poletti, Cavallo, & Adenzato, 2017). A central problem is that traditional tests, by definition, provide

structure. Precise test instructions and the nature of the test materials constrain participants' responses. Moreover, tests are administered in a quiet environment free from distraction. In consequence, people may perform well, whilst demonstrating considerable difficulty dealing with the open-ended and conflicting demands of everyday life. Recognition of these limitations has led to attempts to develop tasks that better mirror everyday function (Alderman, Burgess, Knight, & Henman, 2003; Burgess, Alderman, Forbes, & Costello, 2006; Lamberts, Evans, & Spikman, 2010).

Reports of ecological validity of the Brixton test have been variable. One study of stroke patients found the test not only to be a sensitive measure but also to have ecological validity for identifying cognitive functional outcomes (Vordenberg, Barrett, Doninger, Contardo, & Ozoude, 2014). However, other studies of brain-injured people have yielded only modest ecological validity, when measured against behavioural reports from informants (Odhuba, van den Broek, & Johns, 2005; Wood & Liossi, 2006). The latter findings are noteworthy. Nevertheless, they need to be interpreted in context. Executive tasks do not measure a unitary construct (Alexander, Stuss, Picton, Shallice, & Gillingham, 2007; Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Shallice & Burgess, 1996; Stuss, 2011; Testa, Bennett, & Ponsford, 2012), and performance on different tasks may dissociate. A factor analytic study of 19 executive tests in a healthy adult population (Testa et al., 2012) showed that the Brixton test loaded on to the same factor as the Wisconsin Card Sorting Test (Milner, 1963; Nelson, 1976) and the Zoo map subtest of the Behavioural Assessment of the Dysexecutive Syndrome (BADS) (Wilson, Alderman, Burgess, Emslie, & Evans, 1996), which the authors defined as a task analysis factor, involved in problem-solving and reasoning. Other executive tasks loaded on to separate factors. In keeping with such fractionation, whereas some patients with frontal lobe lesions perform poorly on the Brixton test (Telling, Meyer, & Humphreys, 2010), others do not (Kapur et al., 2009).

Just as components of executive function are dissociable in frontal lobe disease so too are behavioural changes and executive impairments (Stuss, 2011). In studies of bvFTD, executive and behavioural measures have shown only modest association and distinct, and overlapping neural correlates (Gansler, Huey, Pan, Wasserman, & Graffman, 2017). Evidence that performance on the Brixton test is a poor predictor of behavioural change, elicited by questionnaire or structured interview, might simply reflect the fact that the two are measuring distinct components of frontal lobe function. Perhaps a more apt marker of everyday functioning, against which the Brixton test could be evaluated, would be practical indices of patients' reasoning, decision-making, and judgement in daily life. In any event, it is of clinical relevance that despite inherent limitations of clinic-based executive tests, the Brixton test proved sensitive and informative in the investigation of bvFTD.

Conclusion

The Brixton test is a valuable clinical tool that effectively distinguishes between bvFTD and SD, associated respectively with prominent frontal and temporal lobe degeneration. The study highlights the clinical diagnostic value of qualitative analysis of test performance. The different error patterns within bvFTD demonstrate the dissociative nature of different forms of perseveration and reinforce the notion of separable contributions to executive performance.

Acknowledgements

We thank the British Psychological Society Division of Neuropsychology and the British Neuropsychological Society for providing support to Sulakshana P. Rao under the International Fellowship Scheme organized by Prof Narinder Kapur. The study was undertaken during SPR's visit to the Manchester Centre for Clinical Neurosciences as part of her UK travel fellowship.

Conflicts of interest

All authors declare no conflict of interest.

Author contributions

Sulakshana P. Rao (Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Writing – review & editing) Ranita Nandi (Validation) Aparna Dutt (Methodology; Supervision; Validation; Writing – review & editing) Narinder Kapur (Conceptualization; Funding acquisition; Methodology; Writing – review & editing) Jennifer M. Harris (Data curation; Project administration; Writing – review & editing) Jennifer C. Thompson (Validation; Writing – review & editing) Julie Snowden, PhD (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Supervision; Validation; Writing – original draft).

Ethical approval

Ethical approval was obtained for clinical data, acquired in the Cerebral Function Unit, Manchester Centre for Clinical Neurosciences, to be used for research purposes (UK National Research Ethics Committee reference: 09/h0906/53+5).

Data availability statement

The data set that underpins the study is available in a Mendeley data repository at <http://dx.doi.org/10.17632/pjpxdmw84h.1>

References

- Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of International Neuropsychological Society*, *9*, 31–44. <https://doi.org/10.1017/S1355617703910046>
- Alexander, M. P., Stuss, D. T., Picton, T., Shallice, T., & Gillingham, S. (2007). Regional frontal injuries cause distinct impairments in cognitive control. *Neurology*, *68*, 1515–1523. <https://doi.org/10.1212/01.wnl.0000261482.99569.fb>
- Besnard, J., Allain, P., Aubin, G., Chauviré, V., Etcharry-Bouyx, F., & Le Gall, D. (2011). A contribution to the study of environmental dependency phenomena: The social hypothesis. *Neuropsychologia*, *49*, 3279–3294. <https://doi.org/10.1016/j.neuropsychologia.2011.08.001>
- Bielak, A. A. M., Mansueti, L., Strauss, E., & Dixon, R. A. (2006). Performance on the Hayling and Brixton tests in older adults: Norms and correlates. *Archives of Clinical Neuropsychology*, *21*, 141–149. <https://doi.org/10.1016/j.acn.2005.08.006>

- Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*, 1207–1215. [https://doi.org/10.1016/S0028-3932\(00\)00034-8](https://doi.org/10.1016/S0028-3932(00)00034-8)
- Buhl, C., Stokholm, J., & Gade, A. (2013). Clinical utility of short social cognition tests in early differentiation of behavioral variant frontotemporal dementia from Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders Extra*, *3*, 376–385. <https://doi.org/10.1159/000355123>
- 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. <https://doi.org/10.1017/S1355617798466037>
- Burgess, P. W., Alderman, N., Forbes, C., & Costello, C. (2006). The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, *12*, 194–209. <https://doi.org/10.1017/S1355617706060310>
- Burgess, P. W., & Shallice, T. (1996). Bizarre responses, rule detection and frontal lobe lesions. *Cortex*, *32*, 241–259. [https://doi.org/10.1016/s0010-9452\(96\)80049-9](https://doi.org/10.1016/s0010-9452(96)80049-9)
- Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Pearson Clinical UK, Retrieved from www.pearsonclinical.co.uk.
- Burrell, J. R., Halliday, G. M., Kril, J. J., Ittner, L. M., Götz, J., Kiernan, M. C., & Hodges, J. R. (2016). The frontotemporal dementia-motor neuron disease continuum. *Lancet*, *388*, 919–931. [https://doi.org/10.1016/S0140-6736\(16\)00737-6](https://doi.org/10.1016/S0140-6736(16)00737-6)
- Carey, C. L., Woods, S. P., Damon, J., Halabi, C., Dean, D., Delis, D. C., ... Kramer, J. H. (2008). Discriminant validity and neuroanatomical correlates of rule monitoring in frontotemporal dementia and Alzheimer's disease. *Neuropsychologia*, *46*, 1081–1087. <https://doi.org/10.1016/j.neuropsychologia.2007.11.001>
- Chan, R. C. K., Shum, D., Touloupoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*, 201–216. <https://doi.org/10.1016/j.acn.2007.08.010>
- Crescentini, C., Seyed-Allaei, S., De Pisapia, N., Jovicich, J., Amati, D., & Shallice, T. (2011). Mechanisms of rule acquisition and rule following in inductive reasoning. *The Journal of Neuroscience*, *31*, 7763–7774. <https://doi.org/10.1523/JNEUROSCI.4579-10.2011>
- Crutch, S. J., & Warrington, E. K. (2002). Preserved calculation skills in a case of semantic dementia. *Cortex*, *38*, 389–399. [https://doi.org/10.1016/s0010-9452\(08\)70667-1](https://doi.org/10.1016/s0010-9452(08)70667-1)
- Delis, D. C., Kaplan, E. B., & Kramer, J. H. (2001). *The Delis-Kaplan executive function system*. San Antoni, TX: The Psychological Corporation.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. <https://doi.org/10.3758/BF03193146>
- Freedman, M., Binns, M., Black, S., Levine, B., Miller, B., Ramirez, J., ... Stuss, D. T. (2013). Object alternation. A novel probe of medial frontal function in frontotemporal dementia. *Alzheimer's Disease and Associated Disorders*, *27*, 316–323. <https://doi.org/10.1097/WAD.0b013e318293b546>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations and interpretation. *Journal of Experimental Psychology: General*, *141*, 2–18. <https://doi.org/10.1037/a0024338>
- Gansler, D. A., Huey, E. D., Pan, J. J., Wasserman, E., & Graffman, J. H. (2017). Assessing the dysexecutive syndrome in dementia. *Journal of Neurology, Neurosurgery and Psychiatry*, *88*, 254–261. <https://doi.org/10.1136/jnnp-2016-313576>
- Ghosh, A., & Dutt, A. (2010). Utilisation behaviour in frontotemporal dementia. *Journal of Neurology, Neurosurgery and Psychiatry*, *81*, 154–156. <https://doi.org/10.1136/jnnp.2008.160416>
- Ghosh, A., Dutt, A., Bhargava, P., & Snowden, J. (2013). Environmental dependency behaviours in frontotemporal dementia: Have we been underrating them? *Journal of Neurology*, *260*, 861–868. <https://doi.org/10.1007/s00415-012-6722-0>

- Goll, J. C., Crutch, S. J., Loo, J. H. Y., Rohrer, J. D., Frost, C., Bamiou, D. E., & Warren, J. D. (2010). Non-verbal sound processing in the primary progressive aphasias. *Brain*, *33*, 272–285. <https://doi.org/10.1093/brain/awp235>
- Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., . . . Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, *76*, 1006–1014. <https://doi.org/10.1212/WNL.0b013e31821103e6>
- Green, H. A., & Patterson, K. (2009). Jigsaws—a preserved ability in semantic dementia. *Neuropsychologia*, *47*, 569–576. <https://doi.org/10.1016/j.neuropsychologia.2008.10.015>
- Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach* (pp. 59–60). New York, NY: Psychology Press.
- Hodges, J. R., Mitchell, J., Dawson, K., Spillantini, M. G., Xuereb, J. H., McMonagle, P., . . . Patterson, K. (2010). Semantic dementia: Demography, familial factors and survival in a consecutive series of 100 cases. *Brain*, *133*, 300–306. <https://doi.org/10.1093/brain/awp248>
- Hodges, J. R., & Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *Lancet Neurology*, *6*, 1004–1014. [https://doi.org/10.1016/S1474-4422\(07\)70266-1](https://doi.org/10.1016/S1474-4422(07)70266-1)
- Hornberger, M., Piguet, O., Kipps, C., & Hodges, J. R. (2008). Executive function in progressive and nonprogressive behavioral variant frontotemporal dementia. *Neurology*, *71*, 1481–1488. <https://doi.org/10.1212/01.wnl.0000334299.72023.c8>
- Hornberger, M., Savage, S., Hsieh, S., Mioshi, E., Piguet, O., & Hodges, J. R. (2010). Orbitofrontal dysfunction discriminates behavioral variant frontotemporal dementia from Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *30*, 547–552. <https://doi.org/10.1159/000321670>
- Josephs, K. A., Whitwell, J. L., Vemuri, P., Senjem, M. L., Boeve, B. F., Knopman, D. S., . . . Jack, C. R. (2008). The anatomic correlate of prosopagnosia in semantic dementia. *Neurology*, *71*, 1628–1633. <https://doi.org/10.1212/01.wnl.0000334756.18558.92>
- Julien, C. L., Thompson, J. C., Neary, D., & Snowden, J. S. (2008). Arithmetic knowledge in semantic dementia: Is it invariably preserved? *Neuropsychologia*, *46*, 2732–2744. <https://doi.org/10.1016/j.neuropsychologia.2008.05.010>
- Kapur, N., Hutchinson, P., Berry, E., Hawkins, K., Llewellyn, D., & Wilson, B. (2009). Executive dysfunction in a case of transoral-frontal self-inflicted gunshot injury. *Brain Injury*, *23*, 985–989. <https://doi.org/10.3109/02699050903373469>
- Kipps, C., Hodges, J. R., Fryer, T. D., & Nestor, P. J. (2009). Combined magnetic resonance imaging and positron emission tomography brain imaging in behavioural variant frontotemporal degeneration: Refining the clinical phenotype. *Brain*, *132*, 2566–2578. <https://doi.org/10.1093/brain/awp077>
- Knopman, D. S., Weintraub, S., & Pankratz, V. S. (2011). Language and behavior domains enhance the value of the clinical dementia rating scale. *Alzheimer's & Dementia*, *7*, 293–299. <https://doi.org/10.1016/j.alz.2010.12.006>
- Lamberts, K. F., Evans, J. J., & Spikman, J. M. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology*, *32*, 56–65. <https://doi.org/10.1080/13803390902806550>
- Lhermitte, F. (1983). "Utilization behaviour" and its relation to lesions of the frontal lobes. *Brain*, *106*, 237–255. <https://doi.org/10.1093/brain/106.2.237>
- Luzzi, S., Baldinelli, S., Ranaldi, V., Fabi, K., Cafazzo, V., Fringuelli, F., . . . Gainotti, G. (2017). Famous faces and voices: Differential profiles in early right and left semantic dementia and in Alzheimer's disease. *Neuropsychologia*, *94*, 118–128. <https://doi.org/10.1016/j.neuropsychologia.2016.11.020>
- Luzzi, S., Cafazzo, V., Silvestrini, M., & Provinciali, L. (2013). Arithmetic knowledge in early semantic dementia. *Neurological Sciences*, *34*, 1613–1619. <https://doi.org/10.1007/s10072-013-1294-z>
- Luzzi, S., Snowden, J. S., Neary, D., Coccia, M., Provinciali, L., & Lambon Ralph, M. A. (2007). Distinct patterns of olfactory impairment in Alzheimer's disease, semantic dementia, frontotemporal dementia and corticobasal degeneration. *Neuropsychologia*, *45*, 1823–1831. <https://doi.org/10.1016/j.neuropsychologia.2006.12.008>

- Manchester, D., Priestly, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, *18*, 1067–1081. <https://doi.org/10.1080/02699050410001672387>
- McKenna, P., & Warrington, E. K. (1980). *Graded naming test*. Cambridge, UK: Cambridge Cognition.
- Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*, 90–100. <https://doi.org/10.1001/archneur.1963.00460070100010>
- Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., . . . Benson, D. F. (1998). Frontotemporal lobar degeneration. A consensus on clinical diagnostic criteria. *Neurology*, *51*, 1546–1554. <https://doi.org/10.1212/wnl.51.6.1546>
- Neary, D., Snowden, J. S., Mann, D. M. A., Northen, B., Goulding, P. J., & MacDermott, N. (1990). Frontal lobe dementia and motor neuron disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *53*, 23–32. <https://doi.org/10.1136/jnnp.53.1.23>
- Nelson, H. E. (1976). A modified card sorting task sensitive to frontal lobe defects. *Cortex*, *12*, 313–324. [https://doi.org/10.1016/S0010-9452\(76\)80035-4](https://doi.org/10.1016/S0010-9452(76)80035-4)
- 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. <https://doi.org/10.1348/014466505X29431>
- Papagno, C., Semenza, C., & Girelli, L. (2013). Meeting an “impossible challenge” in semantic dementia: Outstanding performance in numerical Sudoku and quantitative number knowledge. *Neuropsychology*, *27*, 680–690. <https://doi.org/10.1037/a0034457>
- Piwica-Worms, K. E., Omar, R., Hailstone, J. C., & Warren, J. D. (2010). Flavour processing in semantic dementia. *Cortex*, *46*, 761–768. <https://doi.org/10.1016/j.cortex.2009.07.002>
- Poletti, M., Cavallo, M., & Adenzato, M. (2017). Detecting dysexecutive syndrome in neurodegenerative diseases: Are we using an appropriate approach and effective tools? *Journal of Neurology, Neurosurgery and Psychiatry*, *88*, 195. <https://doi.org/10.1136/jnnp-2016-314744>
- Primativo, S., Clark, C., Yong, X. X., Firth, N. C., Nicholas, J., Alexander, D., . . . Crutch, S. J. (2017). Eyetracking metrics reveal impaired spatial anticipation in behavioural variant frontotemporal dementia. *Neuropsychologia*, *106*, 328–340. <https://doi.org/10.1016/j.neuropsychologia.2017.10.014>
- Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., . . . Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, *134*, 2456–2477. <https://doi.org/10.1093/brain/awr179>
- Reverberi, C., D’Agostini, S., Skrap, M., & Shallice, T. (2005). Generation and recognition of abstract rules in different frontal lobe subgroups. *Neuropsychologia*, *43*, 1924–1937. <https://doi.org/10.1016/j.neuropsychologia.2005.03.004>
- Reverberi, C., Lavaroni, A., Gigli, G. L., Skrap, M., & Shallice, T. (2005). Specific impairments of rule induction in different frontal lobe subgroups. *Neuropsychologia*, *43*, 460–472. <https://doi.org/10.1016/j.neuropsychologia.2004.06.008>
- Robinson, G. A., Cipolotti, L., Walker, D. G., Biggs, V., Bozzali, M., & Shallice, T. (2015). Verbal suppression and strategy use: A role for right lateral prefrontal cortex? *Brain*, *2015*(138), 1084–1096. <https://doi.org/10.1093/brain/awv003>
- Saxon, J. A., Harris, J. M., Thompson, J. C., Jones, M., Richardson, A. M. T., Langheinrich, T., . . . Snowden, J. S. (2017). Semantic dementia, progressive non-fluent aphasia and their association with amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, *88*, 711–712. <https://doi.org/10.1136/jnnp-2016-314912>
- Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and the temporal organization of behavior. *Philosophical Transactions of the Royal Society B*, *351*, 1405–1412. <https://doi.org/10.1098/rstb.1996.0124>
- Shallice, T., Burgess, P., Schon, P., & Baxter, D. (1989). The origins of utilization behaviour. *Brain*, *112*, 1587–1598. <https://doi.org/10.1093/brain/112.6.1587>

- Snodgrass, J. G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215. <https://doi.org/10.1037/0278-7393.6.2.174>
- Snowden, J. S., Griffiths, H., & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology*, *11*, 265–288. <https://doi.org/10.1080/02643299408251976>
- Snowden, J. S., Stopford, C., Julien, C., Thompson, J. C., Davidson, Y., Gibbons, L., . . . Mann, D. M. A. (2007). Cognitive phenotypes in Alzheimer's disease and genetic risk. *Cortex*, *43*, 835–885. [https://doi.org/10.1016/s0010-9452\(08\)70683-x](https://doi.org/10.1016/s0010-9452(08)70683-x)
- Snowden, J. S., Thompson, J. C., & Neary, D. (2004). Knowledge of famous faces and names in semantic dementia. *Brain*, *127*, 1–13. <https://doi.org/10.1093/brain/awh099>
- Spitoni, G. F., Bevacqua, S., Cerini, C., Ciurli, P., Piccardi, L., Guariglia, P., . . . Antonucci, G. (2018). Normative data for the Hayling and Brixton tests in an Italian population. *Archives of Clinical Neuropsychology*, *33*, 466–476. <https://doi.org/10.1093/arclin/acx072>
- Stopford, C. L., Snowden, J. S., Thompson, J. C., & Neary, D. (2007). Distinct memory profiles in Alzheimer's disease. *Cortex*, *43*, 846–857. [https://doi.org/10.1016/s0010-9452\(08\)70684-1](https://doi.org/10.1016/s0010-9452(08)70684-1)
- Stuss, D. T. (2011). Functions of the frontal lobes: Relation to executive functions. *Journal of the International Neuropsychological Society*, *17*, 759–765. <https://doi.org/10.1017/S1355617711000695>
- Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., . . . 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. [https://doi.org/10.1016/s0028-3932\(99\)00093-7](https://doi.org/10.1016/s0028-3932(99)00093-7)
- Swanberg, M. M., Tractenberg, R. E., Mohs, R., Thal, L. J., & Cummings, J. L. (2004). Executive dysfunction in Alzheimer's disease. *Archives of Neurology*, *61*, 556–560. <https://doi.org/10.1001/archneur.61.4.556>
- Telling, A. L., Meyer, A. S., & Humphreys, G. W. (2010). Distracted by relatives: Effects of frontal lobe damage on semantic distraction. *Brain and Cognition*, *73*, 208–214. <https://doi.org/10.1016/j.bandc.2010.05.004>
- Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Archives of Clinical Neuropsychology*, *27*, 213–224. <https://doi.org/10.1093/arclin/acr112>
- Van den Berg, E., Nys, G. M. S., Brands, A. M. A., & Ruis, C. (2009). The Brixton Spatial Anticipation Test as a test for executive function: Validity in patient groups and norms for older adults. *Journal of the International Neuropsychological Society*, *15*, 696–703. <https://doi.org/10.1017/S1355617709990269>
- Vordenberg, J. A., Barrett, J. J., Doninger, N. A., Contardo, C. P., & Ozoude, K. A. (2014). Application of the Brixton spatial anticipation test in stroke: Ecological validity and performance characteristics. *Clinical Neuropsychology*, *28*, 300–316. <https://doi.org/10.1080/13854046.2014.881555>
- Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. London, UK: Pearson Clinical UK.
- Wong, S., Flanagan, E., Savage, G., Hodges, J., & Hornberger, M. (2014). Contrasting prefrontal cortex contributions to episodic memory function in behavioural variant Frontotemporal Dementia and Alzheimer's Disease. *PLoS One*, *9*, 1–13. <https://doi.org/10.1371/journal.pone.0087778>
- Wood, R. L., & Liossi, C. (2006). The ecological validity of executive tests in a severely brain injured sample. *Archives of Clinical Neuropsychology*, *21*, 429–437. <https://doi.org/10.1016/j.acn.2005.06.014>

Appendix

Flow chart for error classification

Qu. 1. Is the response identical to the correct answer?

YES → Score as **CORRECT** (even if response appears correct by chance)

NO → Score as **ERROR**. Go to Qu. 2

Qu. 2. Does the error response bear an identifiable relationship to the position of the blue circle/stimulus?

YES → It is a **STIMULUS-RELATED** error. Go to Qu. 3

NO → Go to Qu. 7

Qu. 3. Is the error response identical to the stimulus?

YES → It is a **Stimulus Bound** error

NO → Go to Qu. 4

Qu. 4. Is the error response a plausible interpretation of current rule? for example, responds 6 after 3,2,1 on basis of putative 'move anticlockwise' rule

YES → It is a **Logical Inference** error

NO → Go to Qu. 5

Qu. 5. Is the relationship of the response to the stimulus the same as for the preceding set or the immediately preceding response (or in the case of alternating rules the immediately preceding pair of responses)?

YES → It is a **Perseveration of Rule – type A** error

NO → Go to Qu. 6

Qu. 6. Does the error constitute a shift to an alternative rule, which is not the immediately preceding one?

YES → It is a **Same Rule** error. Go to Qu. 6a

NO → Go to Qu.7

Qu. 6a. Does the shift to an earlier rule occur within a set for which the correct rule has been acquired?

YES → **It is a Same Rule Diversion**

NO → It is a Same Rule error

Qu. 7. Does the response bear an identifiable relationship to the preceding response?

YES → It is a **RULE-BASED, STIMULUS-UNRELATED** error. Go to Qu. 8

NO → Go to Qu. 10

Qu. 8. Is the response identical to the immediately preceding response?

YES → It is a **Perseveration of Response** error

NO → Go to Qu. 9

Qu. 9. Is the relationship of the response to the preceding response consistent with an identifiable rule? (This may an earlier rule, acquired or not acquired, or an idiosyncratic rule)

YES → It is a **Perseveration of Rule –type B** error

NO → Go to Qu. 10

Qu. 10. Is the response unrelated to an identifiable rule?

YES → It is a **NON-RULE BASED** error. Go to Qu. 11

Continued

NO → Re-check earlier questions

Qu. 11. Does the response appear arbitrary?

YES → It is a **Random** error. Go to Qu. 11a

NO → Go to Qu. 12

Qu. 11a → Does the random response occur within a set for which the correct rule has been acquired?

YES → It is a **Random Diversion** error

NO → It is a Random error

Qu. 12. Is the response for the trial missing?

YES → It is an **Omission** error

NO → Re-check earlier questions
