ELSEVIER

Neuropsychologia 43 (2005) 1924–1937

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Generation and recognition of abstract rules in different frontal lobe subgroups☆

Carlo Reverberi [a,b,*], Serena D'Agostini [c], Miran Skrap [c], Tim Shallice [a,d]

[a] *International School for Advanced Studies (SISSA–ISAS), Trieste, Italy*
[b] *Department of Psychology, Università Milano–Bicocca, Piazza Ateneo Nuovo 1, 20126 Milano, Italy*
[c] *Ospedale Santa Maria della Misericordia, Udine, Italy*
[d] *Institute of Cognitive Neuroscience, London, UK*

Received 9 October 2004; received in revised form 15 February 2005; accepted 2 March 2005
Available online 12 April 2005

Abstract

The Left Lateral cortex is known to have a role in inductive reasoning tasks. A more specific hypothesis on its role is that it is crucial in the generation of new abstract rules, rather than in the selection and implementation of a specific rule among a set of previously learned ones. Two new tests – the Generation of Hypotheses test and the Recognition of the Rule test – were administered to 46 patients with focal damage to the frontal cortex. Patients were divided in three frontal subgroups: Left Lateral, Right Lateral and Medial. On the basis of the new hypothesis, it was predicted that (i) the Left Lateral subgroup would fail in the Generation of Hypotheses test but would show spared performance on the Recognition of the Rule test and that (ii) the other frontal subgroups would perform normally on both tests. The findings on the Left Lateral and Right Lateral frontal subgroup were consistent with the predictions. This suggests that the Left Lateral frontal cortex is critical specifically for the generation of hypotheses in inductive reasoning. The Medial frontal subgroup, in contrast with our expectations, was impaired on Generation test; two hypotheses have been raised to explain this finding.

Keywords: Inductive reasoning; Frontal cortex; Working memory; Hypothesis generation; Executive functions; Thinking

1. Introduction

The Wisconsin Card Sorting test (WCST, Milner, 1964; Stuss et al., 2000) is a well-known task in which participants are required to sort cards according to a criterion (colour, form or number), which they learn through a process of trial and error, and then shift to a new criterion following a schedule determined by the examiner. In order to perform within the normal range in this task, the available neuropsychological evidence places a crucial role on frontal lobe structures (Drewe, 1974; Milner, 1964; Stuss et al., 2000, see also Demakis, 2003 for a recent systematic review), although other more posterior structures may well be involved too (Stuss et al., 2000).

Once it has been established that a test has a fairly good anatomical specificity, the next step is to understand why a particular class of patients is impaired (i.e. to determine which are the cognitive functions that cause the failure) and, more crucially, if these cognitive functions putatively involved can dissociate from each other and localise to different structures inside the frontal lobes. Traditionally rule abstraction (inductive reasoning) was viewed as a key component of performance on sorting tasks using similar attentional dimensions to the WCST (e.g. Cicerone, Lazar, & Shapiro, 1983); more recently, a deficit in switching/inhibition of central sets is the hypothesis most often considered for the frontal impairment in WCST (Dias, Robbins, & Roberts, 1996; Owen, Roberts, Polkey, Sahakian, & Robbins, 1991; Stuss & Levine, 2002; Warrington, 2000). It is, nevertheless, widely acknowledged that the WCST is a multi-component test: other frontal func-

☆ The study was carried out in the Santa Maria della Misericordia Hospital (Udine, Italy), in SISSA (Trieste, Italy) and in Università Milano–Bicocca (Milano, Italy).

* Corresponding author. Tel.: +39 02 64486776; fax: +39 02 64486706.
E-mail address: carlo.reverberi@unimib.it (C. Reverberi).

0028-3932/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2005.03.004

tions that one can hypothesize to affect performance on the test are (Burgess & Shallice, 1996; Dehaene & Changeux, 1991; Kimberg & Farah, 1993; Rolls, 2000; Stuss et al., 2000):

(i) application and following of rules;
(ii) utilisation of feedback to guide behaviour;
(iii) working memory;
(iv) monitoring and checking;
(v) impulsivity.

The WCST is not a task which is well suited to disentangling the effects of the different factors that may affect the performance of different types of patient. Given its structure it would be very difficult to specify alternative patterns of failure that could unambiguously be attributed to one of the above factors or a combination of them.

Abstracting a rule – induction – is also a major component of another concept attainment task, the Brixton Spatial Rule Attainment task (Burgess & Shallice, 1996). Recently our group (Reverberi, Lavaroni, Gigli, Skrap, & Shallice, 2005) devised a new version of the Brixton task, with three aims: to check if any of the above mentioned functions is compromised in a frontal patient sample, to examine if the same functions dissociate from each other and, finally, to investigate if they localize to different sub-regions of the frontal lobes. The new version of the Brixton test is split into two halves: in the first half, which is similar to the standard version (Burgess & Shallice, 1996), participants are presented with a card containing a 2×5 display of circles only one of which is coloured blue. The participants must predict which circle will be blue on the next card. The rules, which have to be attained, pertain to the relation among succeeding stimuli. In the second half an interference procedure is introduced for each of the rules presented, to probe the appearance of error kinds theoretically related to a monitoring deficit.

Using a lesion classification procedure introduced by Stuss et al. (1998), Reverberi and collaborators were able to show that only two frontal subgroups were significantly impaired on the Brixton task: the Left Lateral and the Inferior Medial. Moreover, they showed that only an induction deficit, and not an impairment of the other functions mentioned above in the discussion about WCST, was able to account for the difficulties that the Left Lateral subgroup had on the task. In the remaining frontal subgroups (Inferior and Superior Medial, Right Lateral) the Brixton test score either showed no deficit or could be explained by one of the alternative hypotheses. Thus, if the interpretation of the Left Lateral group impairment in terms of a difficulty in induction is correct, the Left Lateral frontal cortex should be viewed as a crucial neural substrate of inductive reasoning, in contrast to the marginal role of the other frontal regions.

Another aspect of the behaviour of frontal patients noted qualitatively by Reverberi, Lavaroni, et al. (2005) suggested an even more specific cause for the impairment of the Left Lateral group. They observed that these patients often complained that they had difficulty in finding a possible rule, and indeed frequently doubted that one was actually present. This comment suggested that the impairment of these patients arose in the generation of hypotheses rather than in a subsequent stage of the inductive reasoning cascade, i.e. the selection between alternative hypotheses. Thus, the Left Lateral patients did not seem to fail because they were unable to choose the right rule efficiently among a set of alternatives, but because they had a more basic problem in generating rules.

The inferences from these clinical observations can be tested in a straightforward way. If the Left Lateral frontal cortex is crucial for generating alternative hypotheses, it would be expected that the patients with a lesion to this region:

(i) would fail in a test of Generation of Hypotheses;
(ii) would perform normally on the Brixton task if all the rules that might be used were shown to the patients before the administration of the test. Of course, a deficit on functions other than the abstraction one necessary to carry out the task (e.g. a deficit to working memory) needs also to be excluded.

Finally, it would be also predicted that the performance of patients with lesions in other regions of the frontal lobes (i.e. Right Lateral, Superior and Inferior Medial aspects) without deficits of working memory would be spared in both tasks.

In the present study, we tested the first prediction by asking participants to produce, in a Brixton-like setting, as many different rules as they could for the movement of a blue circle in a 2×6 array. To test the second prediction, we devised a "Recognition of the Rule" task also based on the Brixton test paradigm. Given the structure of the Recognition test, it was also possible to check on a second patient sample the hypotheses other than an induction deficit that had been considered and rejected in our earlier study for the Left Lateral impairment (Reverberi, Lavaroni, et al., 2005).

The same Working Memory task as in the Reverberi, Lavaroni, et al. (2005) study was also administered in order to assess the ability to store relevant information temporarily: this is also a necessary ability for the Recognition of the Rule task. Finally, we explored the relationship of the WM test with three standard clinical short-term memory tests: Digit Span Backward, Digit Span Forward and the Corsi test.

2. Material and methods

2.1. Participants

Forty-six patients with a single focal brain lesion as determined by a CT or an MRI scan were recruited from the Neurological and Neurosurgical wards of Ospedale Civile in Udine (Italy); all patients gave their consent to participate to the study. The study was approved by the ethical committee of Scuola Internazionale Superiore di Studi Avanzati–International School for Advanced Studies (SISSA–ISAS). The aetiology of the patient sample was mixed: stroke, neoplasm and arachnoid cyst (Table 1). Exclu-

Table 1
Aetiology for each lesion group

	MED	LL	RL	Patients overall
Arachnoid cyst	1			1
Glioma high grade	1		4	5
Glioma low grade	8	1	1	10
Meningioma	13	8	6	27
Stroke	1	2		3

The absolute frequencies of patients included in the study. MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal.

sion criteria were the presence in the clinical history of psychiatric disorders, substance abuse or previous neurological disease, neuroradiological evidence of diffuse brain damage, and age less than 18 or more than 70. The time since the lesion ranged between 7 and 1579 days (Table 2); this did not significantly differ between the lesion subgroups [Kruskal–Wallis test, $\chi^2(2) = 0.352$, $P > 0.1$] (the starting point considered in the case of neoplasm is the day of surgery). Among patients, only two had been diagnosed as mild Broca aphasics. Twenty-seven normal control volunteers also participated in the study. The controls were matched with the patients for age and educational level. There were no significant differences between the frontal patients overall and the controls for age [$F(1,71) = 0.196$, $P > 0.1$] or education [$F(1,71) = 0.080$, $P > 0.1$].

2.2. Neuroradiological assessment

For all patients, a CT or an MRI scan was available (although in one case it was later lost). Following the general procedure of Stuss et al. (1998), the patients were assigned to three anatomically defined subgroups depending on their lesion site (Fig. 1): Medial region (MED), in which the lesion involves the orbital surface and/or the medial surface of one or both frontal lobes, Left Lateral (LL) and Right Lat-

Table 2
Demographic variables for each lesion group and for control subjects

	MED	LL	RL	Patients overall	CTL
N	24	11	11	46	27
Age [mean (S.D.)]	51 (10)	52 (13)	48 (12)	50 (11)	49 (10)
Education [mean (S.D.)]	9.17 (3.07)	9.00 (2.97)	9.27 (3.72)	9.15 (3.14)	9.37 (3.27)
Days from onset [median (range)]	228 (7–1507)	626 (7–1579)	231 (7–1314)	252 (7–1579)	

MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal; CTL, Controls.

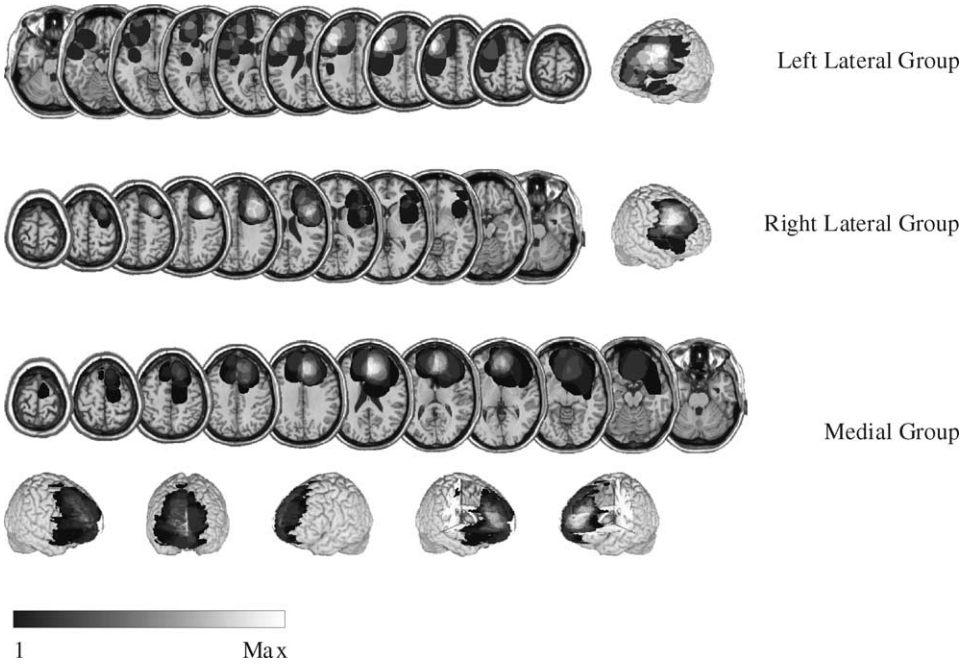

Fig. 1. Overlay lesion plots for the three lesion subgroups. The number of overlapping lesions in each voxel is illustrated on a grey scale: the lighter is a point on the plot, the higher the number of patients with that voxel damaged. The grey scale is devised so that the white colour codes for the maximal overlap in each lesion subgroup; e.g. in the Left Lateral group the maximal number of patients with a lesion to the same voxel across the whole brain is six; thus, the white colour in the grey scale for LL group will code for six patients having the lesion (maximal overlap for RL is 8, for MED is 14). Talairach *z*-coordinates (Talairach & Tournoux, 1988) of each transverse in all plots section are 45, 55, 65, 75, 85, 95, 105, 115, 125, 135, 145 (see Fig. 4 supplementary material). The whole collection of templates for each patient and a coloured version of the overlay lesion plots analogous to the present ones are available in the online supplementary material.

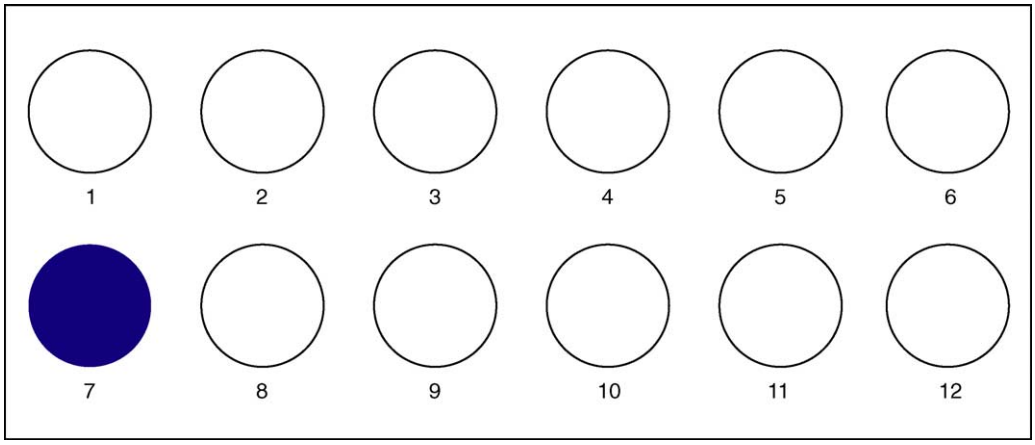

Fig. 2. Generation of Hypotheses test: example of a card.

eral (RL), which have unilateral damage to the frontal lobe convexity (in the present study, we did not split the Medial group into Superior and Inferior because there were too few patients with a lesion to the inferior frontal lobe in the series). In order to classify lesions, the scans were evaluated by a senior neuroradiologist blind to the experimental results.

All patient lesions were mapped using the free MRIcro (www.mricro.com) software distribution (Rorden & Brett, 2000) and were drawn manually – by a senior neuroradiologist blind to experimental results – on slices of a T1-weighted template MRI scan from the Montreal Neurological Institute (www.bic.mni.mcgill.ca/cgi/icbm_view). This template is approximately oriented to match Talairach space (Talairach & Tournoux, 1988) and is distributed with MRIcro. The lesion maps for each individual patient are available on the online supplementary material (see Fig. 6 and Table 1, supplementary material).

2.3. Materials and procedure

2.3.1. Generation of Hypotheses

The participants were required to generate, in 10 min, as many different "rules" as they could for the movement of a blue circle in an array similar to the one used in the Brixton Spatial Rule Attainment Task (Burgess & Shallice, 1996). Each display contained 2 × 6 numbered circles (1–6 first row, left to right; 7–12 second row, left to right); only one being blue, the rest being white (Fig. 2).

On each cycle, the task required the participants to (i) generate a rule, (ii) show it to the experimenter and (iii) describe it verbally, until the available time of 10 min is finished (Fig. 3). Participants were informed that the rules need not be particularly "strange or original", but on the contrary it is useful to show any rule that comes to mind, even the more trivial. It was stressed that the rules need to differ from each other and that as many as possible are required. Verbal responses were *only* used to build the classification system (see below).

To allow participants to demonstrate the generated rules, they were presented on a LCD touch screen monitor with a series of the above-described cards. Participants were informed that they could control the movement of the blue circle by touching on the screen the position toward which they wanted to move it on the next card. Thus, for example, if s/he touched the circle in position 5, the present card turned and a new one was showed with the blue circle in position 5. After six touches, the participant was invited to verbally describe the rule s/he had just demonstrated. Once the participant finished describing the rule, the experimenter pressed a key to begin the next trial, which was signalled by a message on the screen ("invent a new rule"). The position of the blue circle on the first card of each trial was determined by the experimenter using a random ordering which was fixed across participants. Thus, for example, if a participant wanted to demonstrate the "−1" rule, s/he had to apply it to the first card (say 8) and then to the succeeding ones, producing the following sequence: (8)-7-6-5-4-3-2. Reaction times for each

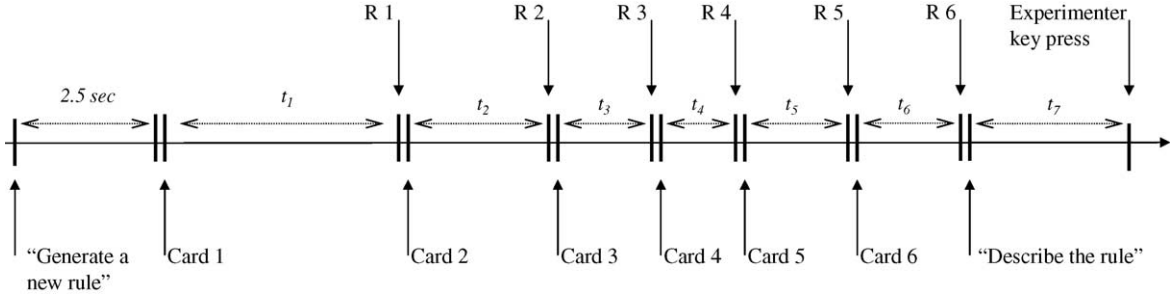

Fig. 3. Generation of Hypotheses test: schema of a single trial of the task. R, response; *t*, time.

response were collected for the time between the beginning of a new trial (the experimenter's key press) and the first choice of position by the participant.

In order to score a rule, *first*, it was necessary to determine if the sequence of blue circle positions produced by the participant was an instance of a recognisable and admissible rule. *Second*, the rule had to be classified in order to detect possible multiple instances of the same rule (i.e. repetitions) generated by the same participant.

The criteria for considering a sequence as an *admissible* rule are:

(i) Each cycle of the rule had to be accomplished within three cards. In other words, if you consider a rule as a succession of elementary spatial transitions (e.g. "diagonal shift", "contiguous horizontal shift") the same transition as the first had to occur in no more than three moves and be followed in order by the same set of transitions (e.g. 6-12-11-4-10-9-2 is an example of the three move rule: "first vertical then horizontal then diagonal");
(ii) There had to be less than three errors in the application of the rule.

These two restrictions are important in order to be sure that the participant is actually applying a rule that s/he has in mind, as the three card maximum for a complete cycle allows at least two full instantiations of the rule in the seven card-six move window considered in the experiment. Thus, by applying these criteria we have excluded, for example:

- Rules composed of more than three elementary transitions, e.g. "horizontal left then vertical then horizontal left omitting one then vertical" or "horizontal right omitting two then three times horizontal left";
- Series of geometrical figures: usually there is not a repetitive pattern in the window considered. For example, a "series of rectangles": 4-5-6-12-11-10-1;
- Sequences without any repetitive recognisable pattern ("incomprehensible sequences" in Table 3). For example, the sequence 1-3-6-12-6-3-1 described by the participant as a series of arithmetic operations: "1 + 2 = 3 then 3 × 2 = 6 then 6 × 2 = 12 then 12/2 = 6", i.e. 1-3-6-12-6-3-1.

It is important to note that verbal utterances of either control participants or patients were not used in order to score a rule as valid or invalid.

Finally, in order to *classify* the sequences, we tried to adhere as far as possible to the implicit rule taxonomy produced by the Control Group (Table 3): the collected verbal descriptions of control participants were used for this purpose. A

Table 3
Generation of Hypotheses task: proportions of rules generated for each type

Rule	Example	CTL (%)	FP (%)
Add or Subtract more than 2 (">2")	4-8-12-4-8-12-4	12.8	14.7
Add or Subtract 2 ("2")	2-4-6-8-10-12-2	11.7	12.7
Add or Subtract 1 ("1")	8-7-6-5-4-3-2	10.7	15.4
Diagonal	2-9-4-11-6-7	6.0	9.9
Top down	1-8-2-9-3-10	5.5	5.9
Greek fret	11-5-4-10-9-3	3.8	5.0
Alternation	10-5-10-5-10-5	3.6	2.4
1; diagonal	8-9-4-5-12-7	3.0	1.3
1; 2	6-4-5-3-4-2-3	2.7	0.4
1; >2	12-11-8-7-4-3-12	2.5	0.7
2; >2	1-3-12-2-11-1	2.5	
Series on two rows	1-6-2-7-3-8-4-9	1.9	1.1
Vertical	12-6-10-4-8-2-12	1.4	1.3
1; diagonal; vertical	6-12-11-4-10-9-2	1.4	0.2
Extremes	12-6-1-7-12-6-1-7	1.1	2.0
1; 2 (*x*)	1-2-3-5-6-7-9-10	1.1	0.4
1; vertical (*x*)	12-11-10-4-3-2-8	1.1	
Square	9-3-4-10-9-3-4-10	0.8	1.7
>2; 1 (*x*)	7-10-9-8-11-10-9	0.8	
1; same	1-6-2-6-3-6-4-6	0.8	
2; vertical	6-12-10-4-2-8-12	0.5	0.6
Same	4-4-4-4-4-4-4-4	0.5	
2; same	7-1-5-1-3-1-1	0.3	
Alternation on three	12-7-4-12-7-4-12	0.3	
Parallel diagonal	6-5-12-4-11-3-10	0.3	
>2; >2	6-10-7-11-8-12-9		0.7
Progression	7-8-10-1-5-10-4		0.7
>2; vertical	12-3-9-12-6-3-9		0.4
Triangle	12-11-5-12-11-5-12		0.2
Inadmissible and incomprehensible	8-11-3-1-7-9-12	6.8	7.9
Inadmissible but comprehensible	6-8-10-9-8-10-12	16.1	14.2

CTL, Control Group; FP, frontal patients overall.

few notable exceptions need to be pointed out; specifically, in contrast to the opinion of control participants, we scored as identical the following rules (Table 3): (i) ±3 (e.g. horizontal shift from position 1 to 4 and vice versa); ±4; ±5 an so on; (ii) alternative orders of the same basic transitions, for example vertical then +1 then −3 is the same as −3 then vertical than +1; (iii) alternative directions of the same basic transition, e.g. a horizontal shift jumping one circle to the right (+2) or to the left (−2). In these three cases, the putative new rule only involved a recurrent application of the same basic ideas ("jump some circles", "combine these three rules") with only a marginally different additional manipulation.

2.3.2. Recognition of the Rule

The test is composed of two parts: a training phase and testing phase. The testing phase is similar to the Brixton Spatial Rule Attainment Task (Burgess & Shallice, 1996).

2.3.2.1. Training phase. An example of each of the seven rule kinds to be used was shown on a computer screen. After the presentation of each example the participants are required to reproduce it on a single Brixton card and describe it verbally. If a participant cannot do this, the procedure was repeated until s/he demonstrated that s/he had understood the rationale of the rule. During the following test phase a synopsis of the relevant rule kinds with the same examples presented during the training were visible to the participant on a on a paper sheet.

2.3.2.2. Test phase. Seventy cards were presented, one at time, on a touch screen monitor. Each card contained a 2 × 5 array of numbered circles (1–5 first row, left to right; 6–10 second row, left to right); only one was blue, the rest being white. The blue circle moved from one card to the next following 11 rules of seven different kinds. On average, a rule changed after six cards (range 5–8), without any explicit warning (Table 4). The participant's task was to touch the circle where s/he thought the blue circle would be on the card following the one currently presented. Participants were told that the coloured circle never moves randomly and that rules change without warning. An example of a series of answers scored as correct is, for the first 10 cards (Table 4), the following: ignored-4-5-6-7-6-2-6-2-6. Note that we counted the first card, which obeyed a new rule, as correct if it followed the preceding rule in force. For example, the response considered correct for the last card of the first rule (Table 4) is "7" even though the card that actually occurs next has the blue circle in position 2.

Table 4
Recognition of the Rule task: cards and rules used

Cards	Rule	*N* cards in the rule
2-3-4-5-6	+1	5
2-6-2-6-2-6-2	Alt (2–6)	7
4-6-8-10-2-4-6	+2	7
5-4-3-2-1-10	−1	6
5-9-4-8-3-7-2-6	Top down	8
6-6-6-6-6	Stay	5
7-1-6-7-1-6-7-1	Triangle	8
5-1-5-1-5-1	Alt (6–10)	6
10-9-8-7-6	−1	5
1-7-2-8-3-9-4	Top down	7
4-4-4-4-4-4	Stay	6

The numbers in the first column refers to the position of the circle in the 2 row × 5 column array.

2.3.3. Working Memory

A test (Reverberi, Lavaroni, et al., 2005) to assess the ability of participants to process the Working Memory requirements of the Recognition of the Rule test was also given. Two card types were used: one being the same as in the Recognition of the Rule task and the other having a red-filled circle instead of a blue-filled one. Three cards with a randomly positioned blue circle were shown to participants one at time. Four cards with a red coloured circle, which they had to touch, followed. Finally they had to state the positions of the three blue-filled circles. Ten trials were administered to each participant. We also administered the Digit Span Forward and Backward (Wechsler, 1997) to both the control participants and the patients and the Corsi test (Spinnler & Tognoni, 1987) to the patients only.

2.4. Variables

In the *Generation of Hypotheses* test we analysed the following variables:

(i) the number of new rules generated in the given time;
(ii) the number of repetitions of rules of the same taxonomical class (Table 3);
(iii) the number of "inadmissible" or "incomprehensible" sequences;
(iv) the generation time for the new rules: the time between the presentation of the "generate a new rule" instruction and the beginning of the demonstration of the rule (i.e. 2.5 s + t_1; see Fig. 3).
(v) The demonstration time for the new rules: the time between the response to the first card and the end of the demonstration of the new rule (i.e. the sum of 2.5 s and RTs from t_1 to t_6; Fig. 3).

In the *Recognition of the Rule* test we analysed basically the same variables we explored in our preceding study on the Brixton test (Reverberi, Lavaroni, et al., 2005). We omitted only so-called "Bizarre Errors" because the definition of this kind of incorrect response could not be applied in the present setting. The variables are:

(i) *Correct responses (RecRule)*.
(ii) *Perseveration of the response (PRe)*: an incorrect response, which is the same as the immediately preceding one (e.g. incorrect response *n*: 5, incorrect response *n* + 1: again 5).
(iii) *Perseveration of the preceding rule (PPRu)*: an incorrect response in which the rule that preceded the currently

active one is applied. When the correct rule has not been attained, each Brixton rule has its own rate of utilisation as an "attempt" for each participant (e.g. some could tend to use more often "+1", others "−1" as the first try); to measure this kind of error appropriately it is necessary to estimate how the baseline rate of production of a particular rule A is modified by the fact that the rule *n* was the last active one. The index used is an odds ratio: the conditional probability that trial *n* obeys rule A given that rule A was the preceding active rule needs to be divided by the conditional probability that trial *n* obeys rule A given that the preceding active rule was not rule A.

(iv) *Same rule (SR)*: incorrect responses on which the participant continues to apply the same incorrect rule, even when s/he has been negatively reinforced (e.g. the participant continues to use a +1 rule even after the first unsuccessful attempt; thus, when the alternation rule is active, a pattern such as this could be obtained: S: 6 R: 7; S: 2 R: 3; S: 6 R: 7 and so on);

(v) *Move errors*: where a subject has correctly attained a rule, but then goes on to make an error. Treating at least two successive correct responses as evidence that the participant has attained a rule, we calculated the number of times each subject subsequently made an error before the rule changed. We considered the ratio of the number of move errors to the number of attained rules.

For each error type apart from PPRu and move, the rate at which each error type occurs was evaluated. This taxonomy is neither mutually exclusive nor exhaustive.

(vi) Generate and recognise (GenRec): we selected on the Recognition test the subset of rules that were formally identical to the ones produced by a particular participant in her/his preceding Generation of Hypotheses test. The proportion of attained rules in this subset is then computed.

2.5. Statistical analysis

2.5.1. Group analysis

The raw data were first checked for conformity to the normal distribution using the Kolmogorov–Smirnov test and for homogeneity of variance by the Levene test. Variables differing significantly from the normal distribution or having inhomogeneous variances between groups underwent logarithmic transformation. If one of the assumptions necessary to apply the analysis of the covariance (ANCOVA) was still not valid after transformation, a non-parametric test, the Mann–Whitney was used. In this latter case *P*-values were estimated using the exact method. Where an ANCOVA was carried out, the effects on the dependent variables were evaluated by covarying for age and education. Given our expectation on the direction of the effects for most of variables considered, we generally used one-tailed tests if not otherwise specified. Effects were considered significant at the $P < 0.05$ level.

2.5.2. Correlation analysis

We evaluated the correlations between a set of potentially relevant variables. A multiple regression analysis was performed with the variables of interest. The percentage of the variation (R^2) explained by demographic factors was removed and *F* statistics were calculated for each relevant variable. Effects were considered significant at the $P < 0.05$ level, two-tailed.

3. Results

3.1. Generation of Hypotheses test

Education is the only factor significantly affecting the number of rules score in controls [$R^2 = 0.16$, $F(1,25) = 4.68$, $P < 0.05$ two-tailed] while age is the only one in patients [$R^2 = 0.10$, $F(1,44) = 4.91$, $P < 0.05$ two-tailed]. A regression analysis has been performed with the logarithm of the days from onset (Table 2) as predictor, age and years of education as covariates and number of rules score as a dependent variable. The proportion of the variance explained by days from onset was negligible [$R^2 = 0.003$, $F(1,42) = 0.163$, $P > 0.1$]. Moreover, we tested the effect of the days from onset variable on the number of generated rules by splitting each lesion subgroup into acute cases (<2 months from onset) and chronic cases (>2 months from onset); the effect was not significant either in the Frontal Group overall or in each of the subgroups considered individually (Fig. 8, supplementary material). The possible effects of differences in aetiology were evaluated by means of an ANCOVA, with demographic factors as covariates. Apart from arachnoid cyst ($n = 1$, not included) four groups were identified: meningioma ($n = 27$), high grade glioma ($n = 5$), low grade glioma ($n = 10$) and stroke ($n = 3$). A difference in aetiology did not affect the number of rules score [$F(3,39) = 0.870$, $P > 0.1$].

The Frontal Group overall (Fig. 4) generated significantly less new rules than the Control Group [$F(1,69) = 10.023$,

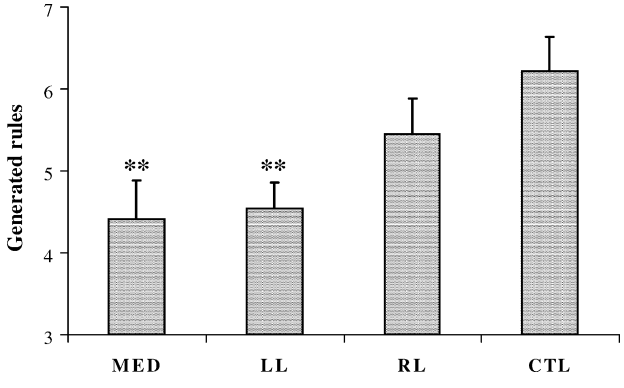

Fig. 4. Generation of Hypotheses test: performance of patient subgroups. MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal; CTL, Control Group.

Table 5
Generation of Hypotheses task: performance of each lesion subgroup and the Control Group

	MED	LL	RL	Patients Overall	CTL
Number of rules	4.42** (2.28)	4.55** (1.04)	5.45 (1.44)	4.70** (1.88)	6.22 (2.15)
Repetitions	3.63 (2.24)	4.91 (2.81)	4.36 (2.54)	4.11 (2.46)	4.15 (3.78)
Inadmissible or incomprehensible rules	2.92 (2.45)	2.27 (2.83)	2.45 (2.30)	2.65 (2.47)	3.11 (2.61)
Generation time	3.34 (1.15)	3.20 (1.02)	3.07 (1.08)	3.24 (1.09)	3.20 (1.14)
Show time	2.47 (0.84)	2.83 (1.38)	2.08 (0.78)	2.47 (1.00)	2.21 (0.90)

Values significantly different from Control Group are underlined; *$P < 0.05$; **$P < 0.01$. Averages with S.D. in parentheses are reported. MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal; CTL, Controls.

$P < 0.01$]. At the frontal subgroup level both the Left Lateral [Mann–Whitney, $z = 2.4$, $P < 0.01$] and the Medial [$F(1,47) = 8.241$, $P < 0.01$] subgroups were significantly impaired; by contrast the Right Lateral group did not differ significantly from the Control Group [$F(1,34) = 1.408$, $P > 0.1$].

The number of repetitions score (Table 5) was not higher than the Control Group either for the Frontal Group overall [$F(1,69) = 0.001$, $P > 0.1$] or for any of the frontal subgroups [LL: $F(1,34) = 0.605$, $P > 0.1$; RL: $F(1,34) = 0.044$, $P > 0.1$; MED: $F(1,47) = 0.231$, $P > 0.1$].

In order to check if the smaller number of rules produced could be due to a tendency in frontal groups to generate more sequences that would, on later analysis, be rejected as inadmissible, we estimated the number of such proposed rules in each of the frontal subgroups and the Control Group: none of the subgroups showed an increased number of inadmissible or incomprehensible rules.

Finally, we evaluated if there were differences between groups over the two major sections in which the time given to each participant was divided, i.e. the time for generating a new rule and the time to show it. This is a necessary step, since a reduction in the time available to think of new rules may be present despite the fixed overall granted time, for example because patients might need a longer time to verbally describe the invented rules or to demonstrate/apply them. The frontal patients overall (Table 5) did not differ significantly from controls either in the generation [$F(1,69) = 0.26$, $P > 0.1$] or in the demonstration time [$F(1,69) = 1.125$, $P > 0.1$]; the same pattern was found at the subgroup level, in particular for the two subgroups with a poor performance on the number of rules measure [generation: LL: $F(1,34) = 0.002$, $P > 0.1$; MED: $F(1,47) = 0.241$, $P > 0.1$; show time: LL: $F(1,34) = 2.281$, $P > 0.1$; MED: $F(1,47) = 0.989$, $P > 0.1$]

3.2. Recognition of the Rule test

Age was the only factor affecting the proportion of correct responses (RecRule) significantly in controls [$R^2 = 0.19$, $F(1,25) = 5.79$, $P < 0.05$ two-tailed], while in patients both age [$R^2 = 0.44$, $F(1,44) = 34.09$, $P < 0.001$ two-tailed] and education [$R^2 = 0.36$ $F(1,44) = 24.51$, $P < 0.001$ two-tailed] were significant. A regression analysis was performed with the logarithm of the days from onset (Table 2) as predictor, age and years of education as covariates and RecRule score as a dependent variable. The proportion of variance explained by days from onset was small but significant [$R^2 = 0.047$, $F(1,42) = 4.565$, $P < 0.05$]. However, by testing the effect of the days from onset variable by splitting each lesion subgroup into acute cases (<2 months from onset) and chronic cases (>2 months from onset) we did not obtain a significant effect either in the Frontal Group overall or in any of the lesion subgroups (Fig. 9, supplementary material). The possible effects of differences in aetiology were evaluated by means of an ANCOVA, with demographic factors as covariates. Apart from arachnoid cyst ($n = 1$, not included) four groups were identified: meningioma ($n = 27$), high grade glioma ($n = 5$), low grade glioma ($n = 10$) and stroke ($n = 3$). A difference in aetiology significantly affected RecRule score [$F(3,43) = 3.433$, $P < 0.05$]. Patients with stroke and high grade glioma tended to have lower scores; in a post hoc analysis none of the pairwise comparisons were significant (LSD method).

The Frontal Group overall (Table 6) gave significantly fewer correct responses on Recognition of the Rule score (RecRule) than did the Control Group [$F(1,69) = 3.393$, $P < 0.05$]. However, at the subgroup level the Right Lateral and the Medial groups showed only a trend toward a deficit [RL: $F(1,34) = 2.359$, $P < 0.1$; MED: $F(1,47) = 1.880$,

Table 6
Recognition of the Rule task: error types for each lesion subgroup and the Control Group

	MED	LL	RL	Patients overall	CTL
Recognition of the Rule, (prop. of errors)	0.49 (0.24)	0.48 (0.23)	0.46 (0.21)	0.47* (0.22)	0.39 (0.18)
PRe errors	0.07 (0.07)	0.11 (0.05)	0.09 (0.05)	0.08 (0.07)	0.09 (0.07)
PPRu (odds ratio)	1.37 (1.02)	1.99 (1.95)	1.26 (0.88)	1.49 (1.28)	1.08 (0.71)
SR	0.16* (0.07)	0.12 (0.06)	0.13 (0.07)	0.14 (0.07)	0.12 (0.06)
Move errors	0.16 (0.18)	0.14 (0.10)	0.21 (0.31)	0.17 (0.20)	0.11 (0.18)
Generated and recognised	0.61 (0.43)	0.64 (0.38)	0.68 (0.30)	0.63 (0.39)	0.79 (0.32)

PRe, perseveration of the response; PPRu, perseveration of the preceding rule; SR, same rule error. Values significantly different from Control Group are underlined; *$P < 0.05$. Averages with S.D. in parentheses are reported. MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal; CTL, Controls.

Table 7
Number of correct responses on Brixton WM test (out of 10), indices for Digit Span Forward, Digit Span Backward for each group

	MED	LL	RL	Patients overall	CTL
Brixton WM test	8.33* (2.63)	7.91* (2.70)	9.27 (1.49)	8.46* (2.43)	9.61 (0.78)
Digit Span FWD	5.35* (0.88)	5.40* (1.26)	5.73 (1.56)	5.45* (1.15)	6.04 (1.24)
Digit Span BKW	4.43 (1.28)	4.10* (1.04)	4.36 (0.88)	4.34* (1.15)	5.08 (1.47)
Corsi test	5.32 (1.13)	5.78 (1.30)	5.67 (1.00)	5.50 (1.13)	

Values significantly different from Control Group are underlined; $^{*}P < 0.05$. Averages with S.D. in parentheses are reported. MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal; CTL, controls.

$P < 0.1$] while Left Lateral patients were not significantly impaired relative to control participants [$F(1,34) = 1.431$, $P > 0.1$].

3.2.1. Error type analysis

The frontal patients overall did not show a significant increase in any of the error types considered (Table 6): either for PRe [$F(1,69) = 0.013$, $P > 0.1$], PPRu [Mann–Whitney, $z = 1.092$, $P > 0.1$], SR [$F(1,69) = 2.537$, $P > 0.05$], or move errors [$F(1,69) = 1.250$, $P > 0.1$]. At the subgroup level, the picture was the same for the Left Lateral [PRe: $F(1,34) = 1.164$, $P > 0.1$; PPRu: Mann–Whitney, $z = 0.998$, $P > 0.1$; SR: $F(1,34) = 0.077$, $P > 0.1$; move errors: $F(1,33) = 0.095$, $P > 0.1$] and Right Lateral subgroups [PRe: $F(1,34) = 0.017$, $P > 0.1$; PPRu: $F(1,34) = 0.356$, $P > 0.1$; SR: $F(1,34) = 0.261$, $P > 0.1$; move errors: Mann–Whitney, $z = 1.411$, $P > 0.1$]. The Medial frontal subgroup showed a significant increase in SR errors [$F(1,47) = 4.959$, $P < 0.05$], with the other error types not being significantly different from controls [PRe: $F(1,47) = 0.585$, $P > 0.1$; PPRu: Mann–Whitney, $z = 0.868$, $P > 0.1$; move errors: $F(1,47) = 0.821$, $P > 0.1$].

3.3. Memory tests

As in our previous study (Reverberi, Lavaroni, et al., 2005), the Brixton Working Memory test (Table 7) was straightforward for controls, all but one of whom had a score equal to or greater than 8 out of 10 (the outlier had a score of 7). The Frontal Group, by contrast, produced a significantly higher error rate compared to the Control Group [Mann–Whitney, $z = 1.944$, $P < 0.05$]. A deficit is not found among all the lesion subgroups: only the LL [Mann–Whitney, $z = 1.730$, $P < 0.05$], and the Medial [Mann–Whitney, $z = 2.002$, $P < 0.05$] subgroups performed significantly worse than healthy controls.

We explored the correlation of the Brixton WM test with other standard short-term memory tests both verbal and spatial (Table 8). A series of regression analyses was run on the results of the Frontal Group with Brixton WM test score as the dependent variable (in the Control Group the variance of the Brixton WM score is too low to carry out a powerful enough correlation analyses). The percentage of the variance (R^2) explained by each of these variables was evaluated after having partialled out the effects of age and years of education. The Digit Span Forward and

Table 8
Correlation (r) between four indices of short-term memory in the frontal patients group

	Brixton WM test	Digit Span FWD	Digit Span BKW
Digit Span FWD	0.228		
Digit Span BKW	0.319*	0.377*	
Corsi test	0.279	0.171	0.164

Values significantly different from Control Group are underlined; $^{*}P < 0.05$. The effects of age and education have been partialled out.

Corsi scores were not good predictors of the Brixton WM test performance [$R^2 = 0.045$, $F(1,40) = 2.01$, $P > 0.1$ two-tailed; $R^2 = 0.07$, $F(1,36) = 3.05$, $P > 0.05$ two-tailed, respectively]. The effect of Digit Span Backward was significant, even though rather small [$R^2 = 0.09$, $F(1,40) = 4.53$, $P < 0.05$ two-tailed].

3.3.1. The effect of WM capacity on Recognition of the Rule test

We split the lesion subgroups into patients who scored in the normal range on the Brixton WM test (WM+) and patients with a score below the normal range (WM−). The WM− frontal patients overall had a RecRule score significantly lower than the Control Group [$F(1,35) = 14.966$, $P < 0.001$]; moreover, the same pattern replicates for the two lesion subgroups with more than two WM− participants (Fig. 5), namely the Left Lateral [$F(1,28) = 6.165$, $P < 0.01$] and Medial [$F(1,29) = 7.559$, $P < 0.01$]. By contrast none of the WM+

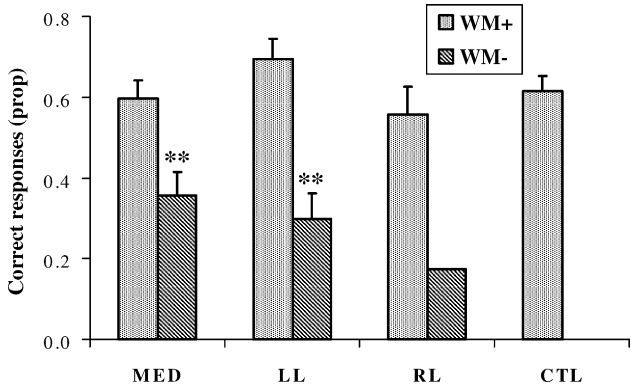

Fig. 5. Recognition of the Rule test: performance of patient subgroups according to whether they scored in (WM+) or below (WM−) the normal range on the Brixton WM test. MED, Medial frontal; LL, Left Lateral frontal; RL, Right Lateral frontal; CTL, Control Group.

groups showed an impairment, either for frontal patients overall [$F(1,57) = 0.403$, $P > 0.1$] or for any of the subgroups [LL: $F(1,29) = 0.270$, $P > 0.1$; RL: $F(1,33) = 1.170$, $P > 0.1$; MED: $F(1,41) = 0.138$, $P > 0.1$].

Finally, even after having partialled out the demographic factors, the Brixton WM score had a significant effect on RecRule score [$R^2 = 0.11$, $F(1,42) = 12.185 < 0.01$ two-tailed] as did the Digit Span Backward [$R^2 = 0.09$, $F(1,40) = 9.206 < 0.01$ two-tailed] but not Digit Span Forward or Corsi's test. A regression model with the two significant memory tests and the demographic factor as predictors can explain, in the frontal group, 66% of the RecRule variance (that corresponds to an $r = 0.811$); in the model all factors are significant. This proportion rises to 96% ($r = 0.979$) if we consider only the Left Lateral patients, by contrast it remains stable to 66% ($r = 0.812$) when selecting only the Medial patients.

3.4. Relationship between Generation and Recognition tests

The proportion of recognised sequences in the subset of rules already produced in the preceding Generation of Hypotheses test (Table 6) did not differ significantly from the controls either for the frontal patients overall [$F(1,69) = 2.709$, $P > 0.1$ two-tailed] or for any of the frontal subgroups [LL: $F(1,34) = 1.074$, $P > 0.1$ two-tailed; RL: $F(1,34) = 0.954$, $P > 0.1$ two-tailed; MED: $F(1,47) = 2.480$, $P > 0.1$ two-tailed]. The pattern is clearly different if we evaluate the same variable in the two functional subgroups—within or outside the normal range for Brixton WM test score. The WM− patients recognised only 37.5% of the rules they have generated, significantly less than controls (79%) [$F(1,35) = 8.735$, $P < 0.01$ two-tailed]; by contrast the WM+ patients recognised 72.2% of the generated rules, not significantly different from controls [$F(1,57) = 0.899$, $P > 0.1$ two-tailed].

4. Discussion

Induction is basic to human thought processes. However, where in the brain the relevant processes take place is not well understood. The Brixton Rule Attainment Task is one of the main clinical neuropsychological tests, which involve inductive processes; it is sensitive to lesions to the frontal cortex (Burgess & Shallice, 1996). In our preceding work (Reverberi, Lavaroni, et al., 2005), we presented evidence that the most probable cause of the impairment on the Brixton test in a specific frontal subgroup – the Left Lateral – is an inductive reasoning deficit. Furthermore, we argued that the Left Lateral frontal cortex is a necessary part of the neural network involved in carrying out inductive inferences.

The detailed pattern of errors and certain qualitative aspects of the behaviour of patients while they carried out the Brixton task (Reverberi, Lavaroni, et al., 2005) allowed us to produce a more specific hypothesis on the causes of the Left Lateral impairment: namely that the deficit is specific to the stage of generation of new rules rather than that of the recognition and implementation of a rule which had already been formulated.

The principal aim of the present work was to evaluate this possibility by means of two new tasks: the Generation of Hypotheses test and the Recognition of the Rule test. In particular, if our hypothesis holds, we predict that:

(i) The Left Lateral subgroup would fail in the Generation of Hypotheses test but would show spared performance on the Recognition of the Rule test, unless another concurrent deficit is present, in particular, a reduced Working Memory span.
(ii) The other frontal subgroups would perform normally on both tests, unless, again, there is a reduced Working Memory capacity, a deficit that can affect the performance of the Recognition of the Rule test.

In addition, we aimed to corroborate certain of the findings of our preceding work, i.e. the absence, in the Left Lateral group of perseverative errors and move errors (Reverberi, Lavaroni, et al., 2005). Finally, the relationship between the Working Memory test developed specifically to assess the working memory requirements of the Brixton task (hereafter "Brixton WM test") and standard short-term memory tests was explored.

We first examine the predictions for the different lesion groups.

4.1. Predictions for the Left Lateral frontal subgroup

Our predictions were well corroborated in the case of the Left Lateral subgroup. On the Generation of Hypotheses task this patient subgroup produced 27% less rules than the Control Group. Moreover, we did not find evidence favouring any of the alternative explanations that might be given for a failure on the test. In particular, the impairment of the Left Lateral patients cannot be accounted by:

(i) A lack of time for the crucial generation phase of the test due to a lengthening of the time taken to demonstrate the rules, which, in fact, did not differ significantly from the controls. The average time available for generation was virtually identical in the control and patient groups (Table 5).
(ii) A working memory deficit. In this case, patients would have generated less new rules either because of forgetting rules which had been produced or because of a difficulty in showing the rule that they had generated (e.g. they can forget one or both of the preceding chosen positions or the generated rule itself). In the former case, the higher number of repetitions that would follow, and the consequent waste of time, would explain the observed reduction for the new rules. However, in our sample Left Lateral patients did not produce more repetitions than

controls. In the second case, we would have found an increase of incomprehensible or inadmissible rules, but this is not the case (Table 5).

(iii) "Bizarre" behaviour which could lead frontal patients to create incomprehensible or inadmissible rules. This would cause waste of time and could result in a drop in the number of new admissible rules. However, the Left Lateral subgroup, on average, produced *less* inadmissible and incomprehensible rules than the Control Group.

In contrast to the Generation of Hypotheses test, the Left Lateral patients were not impaired on the Recognition of the Rule test. Moreover, if only the Left Lateral patients with spared Working Memory span are considered, the average performance of the subgroup is (insignificantly) *above* the one of the Control Group (Fig. 5).

Taken together with our previous work on Brixton test (Reverberi, Lavaroni, et al., 2005), the present findings strongly support the hypothesis that the Left Lateral frontal cortex is crucial for inductive reasoning. More specifically, the ability to generate alternative hypotheses seems to be the stage that is affected by damage to this frontal region.

4.2. Predictions for the Right Lateral frontal subgroup

For the Right Lateral group we predicted no impairments either in the Generation of Hypotheses test or in the Recognition of the Rule test. The predictions are supported; the Right Lateral patients neither generated nor recognised significantly fewer rules than did the control participants. In the case of the Recognition test, the pattern is even clearer if only patients without a working memory deficit are considered.

4.3. Predictions for the Medial frontal subgroup

The predictions for the Medial frontal subgroup were the same as for the Right Lateral one: namely no impairments should be found on either test. As expected, the Medial patients without a working memory deficit were not significantly impaired on the Recognition of the Rule test. However, the Medial subgroup was significantly impaired on the Generation of Hypotheses test. Moreover, none of the alternative explanatory hypotheses considered for the Left Lateral subgroup (the lengthening of the demonstration of the rule phase, a working memory deficit or an excess of "bizarre" responses) apply for the Medial group for the same reasons (Table 5). Finally, the performance of the Medial and Left Lateral Groups was also very similar if the number of generated rules was evaluated along the whole temporal course of the production process (Fig. 7, supplementary material).

These findings are somewhat difficult to reconcile with our preceding results on the Brixton test (Reverberi, Lavaroni, et al., 2005). In that study, patients, in either the Inferior Medial or the Superior Medial subgroups, without a reduced Working Memory span did not show any significant impairment on the Brixton test. Yet as for the generation task here, the generation phase is crucial for that task too. How is one to explain this conflict between the results in the two studies?

The negative result on the Brixton test itself might be attributable to the smaller size of the sample as, in the earlier study, the Inferior Medial and Superior Medial subgroups were analysed separately. To test this possibility, we reanalysed the data from the earlier Brixton test study collapsing the two subgroups into one ($n = 14$). Despite the increased size of the combined Medial group, the difference from the Control Group remained insignificant.

It might be thought that the Brixton test is less sensitive than the Generation of Hypotheses test to a deficit specific to the generation stage. In this case, one could argue that since the rules required in the Brixton test are at the most "simple/prototypical" end of the spectrum of all possible rules, and a generation deficit would have more impact at the difficult end, a subgroup with a "milder" deficit could have normal performance on the Brixton but be impaired on the Generation of Hypotheses test.

However, if it is assumed that the simpler/more prototypical a rule is, the more often it would be generated by a healthy samples, it would be apparent (see Table 3) that only a minority of the rules we used in the Brixton test (add/subtract 1, top down, alternation, extremes, same) are in the simple range.

This leaves two main possibilities. First, the inconsistencies with the preceding work on the Brixton task should be explained by conjecturing a different precise distribution of lesions in the two samples, with a crucial medial structure being less frequently damaged in the earlier study than in the present one. In correspondence with this possibility the Superior Medial patients (8/14) were less strongly represented in the previous study than in the current one (20/24). Under this assumption, this medial structure could be viewed as having a crucial role in generating new hypotheses like the Left Lateral cortex. A more specific anatomo-functional hypothesis is consistent with this possibility. This is one recently put forward by Stuss, Binns, Murphy, and Alexander (2002). They proposed that the superior medial cortex is involved in some general activation of the response mechanisms, so that when it is damaged, a similar pattern of impairments would be found to that occurring after lesions to whichever lateral cortex was more critical for the relevant task. In the present case, the crucial cortex is presumably the Left Lateral one, so the Medial group would have been expected to have a similar performance to the Left Lateral group, as is found. However, on this approach one would have expected a deficit in the Medial group on the previous study; it is possible that the 8/14 Superior Medial patients were not sufficient to make any such effect significant.

A second different possibility presupposes that the spared performance of the Medial patient groups on the Brixton task

(Reverberi, Lavaroni, et al., 2005) is not merely a result of too few patients of the appropriate type to obtain significance. This means that there is no involvement (direct or indirect) of the medial frontal cortex on induction processes. Instead it could be hypothesized that the impairment of the Medial group on the Rule Generation task is due to a deficit to a cognitive function different from that of the generating new hypotheses, but which is nonetheless required for the test. Given that the Generation task, by contrast with the Brixton test, is time limited, one possible candidate could be an initiation deficit, known to especially affect patients with Medial lesions (Godefroy, Lhullier-Lamy, & Rousseaux, 2002; Reverberi, Capitani, & Laiacona, 2005; Stuss et al., 2005). Further investigations are needed to decide between these possibilities.

4.4. The comparison between the Recognition of the Rule and the Brixton tests

Apart from the initial training, the Recognition of the Rule task has a very similar structure to that of the Brixton test; the only differences are that it has more rules and so involves more cards. This allowed us to carry out again some of the error analyses used in our previous work (but see also Burgess & Shallice, 1996), which were critical for rejecting hypotheses for the Left Lateral deficit other than an impairment of induction processes per se. The present study confirms that, during a "Brixton type" test, the frontal patients do not make either more perseverative errors or more move errors than do the Control Group (Table 8). These two findings are even more critical than the preceding ones since, in the Recognition of the Rule task, patients attain rules faster than in the Brixton test, so stronger activations of the representations of each rule would be produced (the participants receive positive feedback more frequently). In addition, there are a larger number of cards on which a move error can be made. The Recognition of the Rule task is, in other words, more sensitive to both of these error types. Therefore, our findings confirm that "Brixton type" tests do not elicit, in frontal patients, perseverative behaviour, and do not pose severe problem of rule application or implementation for them once the rule has been induced. This supports the overall interpretation of the Brixton task as a relatively pure test of inductive reasoning (Reverberi, Lavaroni, et al., 2005).

The strong correlation we found between the Recognition of the Rule test and the linearly combined score of two Working Memory tests (the Brixton WM test and the Digit Span Backward) in the Left Lateral group is also consistent with the induction interpretation. The score of the two memory test, along with demographic factors, was able to explain almost all (96%) the variance of the Recognition of the Rule test; it would appear that for the Left Lateral group the Brixton task reduces to a Working Memory test when the inductive component is removed (as is in the Recognition of the Rule test). This does not mean that the inductive and the working memory factors are the only two relevant ones in the Brixton test, but rather that they are the two most relevant ones for performance of the Brixton test in the Left Lateral group.

4.5. Working Memory

The findings on the Working Memory test confirm the anatomo-functional pattern found in our preceding work (Reverberi, Lavaroni, et al., 2005): both the Left Lateral and the Medial (the Inferior Medial in Reverberi, Lavaroni, et al., 2005) frontal group were significantly impaired, while the Right Lateral is spared (Table 5).

The correlations within the patient group of our special purpose Working Memory test and three other standard short-term memory tests (Digit Span Forward, Digit Span Backward and the Corsi Span test) are in general quite low, the only significant one being with Digit Span Backward. Neither of the low correlations obtained with the scores on the Corsi test and the Digit Span Forward are surprising. During Brixton WM test administration, it is clearly apparent that the preferred encoding strategy by most of the participants involves the use of verbal-numerical representations (position 1, position 2 and so on) as well as or rather than spatial ones; thus, a possible impairment on visuo-spatial short-term memory, to which the Corsi test is sensitive, should not have major consequences on the performance of the Brixton Working Memory test, as we found. Moreover, since the number of digits (corresponding to the positions of the blue circle), that a participant has to retain during the test is quite low (i.e. three, see above in Section 2), even a moderate deficit of the verbal short-term memory stores should not impact negatively on performance of the Brixton WM test. Finally, the significant correlation of the Brixton WM test with the Digit Span Backward fits with the need, in both tests, to select the appropriate information in short-term memory. This fits with the position of Rowe, Toni, Josephs, Frackowiak, and Passingham (2000) that dorsolateral prefrontal cortex is critical for operations on the contents of working memory stores rather than for storage per se.

4.6. Lesion classification methodology

Observation of the overlay lesion plots of both lateral subgroups (Fig. 1) may give rise to the question of whether the functional patterns we observed in these two subgroups are an effect of damage to the *lateral* frontal cortex or whether they should be interpreted as resulting from damage to one of the frontal lobes as a whole. More specifically, we can ask if a lesion either to the *medial* frontal cortex or to the white matter between lateral and medial surface is crucial in order to produce the behaviour observed in the Left Lateral frontal subgroup. The former possibility can be safely rejected since excluding from the analyses the only two lateral patients who may have a minor involvement of medial cortex (see cases LL-6 and LL-7, Fig. 6a, supplementary online material) does not change the main results. The latter possibility cannot be

ruled out solely on the basis of the present lesion data. However, it should be noted that:

(i) In our lateral groups the majority of the lesions arise from meningiomas – specifically 8/11 for Left Laterals – and not from conditions where white matter damage is inevitable. The meningiomas of all the patients in the lateral groups arise from the meninges of the lateral convexity; this means it is the grey matter of the lateral frontal cortex which will be damaged most.
(ii) Several dissociations between "lateral" and "medial" patients classified using the same methodology as the present study have been reported (Reverberi, Lavaroni, et al., 2005; Stuss et al., 1998, 2000, 2005). Thus, in our preceding work on Brixton task (Reverberi, Lavaroni, et al., 2005) the Left Lateral group (impaired) dissociated from the Medial group (see this paper Section 4.3).

Overall given both that the lesion classification procedure adopted in the present paper is essentially that developed by Stuss et al. and already adopted in a series of papers (see Stuss et al., 2005 for discussion) and the nature of our patient sample, it seems appropriate to label two subgroups as "Left Lateral" and "Right Lateral". Nonetheless it has to be underlined that it cannot be excluded that damage to frontal white matter contributed to the pattern of impairments found.

4.7. Conclusions

In this study, we evaluated the predictions derived from the hypothesis that the ability to generate abstract rules depends on the left prefrontal cortex. The findings on the Left Lateral frontal subgroup are consistent with our predictions. The study, therefore, supports the position that the Left Lateral frontal cortex is one of the crucial neural substrates for inductive reasoning and more specifically for the generation of hypotheses stage. The proposal to localise (at least some of the) inductive reasoning processes to the Left Lateral frontal cortex is also compatible with the findings of some of the recent imaging studies on the topic (Duncan et al., 2000; Goel & Dolan, 2004; Goel, Gold, Kapur, & Houle, 1997; Osherson et al., 1998; Parsons & Osherson, 2001; Strange, Henson, Friston, & Dolan, 2001).

The findings on the Medial subgroup do not fit our predictions: the group would have been expected to have had spared performance on the Generation of Hypotheses test, which it did not. The Medial pattern can be interpreted as indicating either that some medial structure is also crucial for generating new rules or that our Generation test also taps other non-reasoning processes, such as initiation, that have their neural basis in the medial frontal region. From the current findings alone one cannot determine the respective roles of medial and lateral surfaces of the left frontal lobe in induction. However, taken in conjunction with the findings on the frontal structures involved in performance of the Brixton task per se (Reverberi, Lavaroni, et al., 2005) the lateral area appears to be much more critical.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at 10.1016/j.neuropsychologia.2005.03.004.

References

Burgess, P. W., & Shallice, T. (1996). Bizarre responses, rule detection and frontal lobe lesions. *Cortex*, *32*(2), 241–259.
Cicerone, K. D., Lazar, R. M., & Shapiro, W. R. (1983). Effects of frontal lobe lesions on hypothesis sampling during concept formation. *Neuropsychologia*, *21*(5), 513–524.
Dehaene, S., & Changeux, J. P. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, *1*(1), 62–79.
Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, *17*(2), 255–264.
Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Primate analogue of the Wisconsin Card Sorting Test: Effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behavioral Neuroscience*, *110*(5), 872–886.
Drewe, E. A. (1974). The effect of type and area of brain lesion on Wisconsin card sorting test performance. *Cortex*, *10*(2), 159–170.
Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., et al. (2000). A neural basis for general intelligence. *Science*, *289*(5478), 457–460.
Godefroy, O., Lhullier-Lamy, C., & Rousseaux, M. (2002). SRT lengthening: Role of an alertness deficit in frontal damaged patients. *Neuropsychologia*, *40*(13), 2234–2241.
Goel, V., & Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition*, *93*(3), 109–121.
Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *NeuroReport*, *8*(5), 1305–1310.
Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*(4), 411–428.
Milner, B. (1964). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, *9*, 100–110.
Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., & Fazio, F. (1998). Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia*, *36*(4), 369–376.
Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*, *29*(10), 993–1006.
Parsons, L. M., & Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebral Cortex*, *11*(10), 954–965.
Reverberi, C., Capitani, E., & Laiacona, M. (2005) Qualitative features of semantic fluency performance in mesial and lateral frontal patients. Manuscript in preparation.
Reverberi, C., Lavaroni, A., Gigli, G. L., Skrap, M., & Shallice, T. (2005). Specific impairments of rule induction in different frontal lobe subgroups. *Neuropsychologia*, *43*(3), 460–472.
Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, *10*(3), 284–294.
Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*(4), 191–200.

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S., & Passingham, R. E. (2000). The prefrontal cortex: Response selection or maintenance within working memory? *Science*, *288*(5471), 1656–1660.

Spinnler, H., & Tognoni, G. (1987). Gruppo italiano per lo studio neuropsicologico dell' invecchiamento: Standardizzazione e taratura italiana di test neuropsicologici (Italian standardisation of neuropsychological tests). *The Italian Journal of Neurological Sciences*, *6*(Suppl. 8), 1–120.

Strange, B. A., Henson, R. N., Friston, K. J., & Dolan, R. J. (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex*, *11*(11), 1040–1046.

Stuss, D. T., Alexander, M. P., Hamer, L., Palumbo, C., Dempster, R., Binns, M., et al. (1998). The effects of focal anterior and posterior brain lesions on verbal fluency. *Journal of the International Neuropsychological Society*, *4*(3), 265–278.

Stuss, D. T., Alexander, M. P., Shallice, T., Picton, T. W., Binns, M. A., Macdonald, R., et al. (2005). Multiple frontal systems controlling response speed. *Neuropsychologia*, *43*(3), 396–417.

Stuss, D. T., Binns, M. A., Murphy, K. J., & Alexander, M. P. (2002). Dissociations within the anterior attentional system: Effects of task complexity and irrelevant information on reaction time speed and accuracy. *Neuropsychology*, *16*(4), 500–513.

Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, *53*, 401–433.

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., et al. (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*(4), 388–402.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Warrington, E. K. (2000). Homophone meaning generation: A new test of verbal switching for the detection of frontal lobe dysfunction. *Journal of the International Neuropsychological Society*, *6*(6), 643–648.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). Psychological Corporation.