doi:10.1093/brain/awh577 *Brain* (2005), **128,** 2882–2890

Better without (lateral) frontal cortex? Insight problems solved by frontal patients

Carlo Reverberi,[1,2] Alessio Toraldo,[3] Serena D'Agostini[4] and Miran Skrap[4]

[1]International School for Advanced Studies (SISSA–ISAS), Trieste, Italy, [2]Department of Psychology, Università Milano–Bicocca, Milano, Italy, [3]Department of Psychology, Università degli Studi di Pavia, Pavia, Italy and [4]Azienda Ospedaliera S. Maria della Misericordia, Udine, Italy

Correspondence to: Carlo Reverberi, Department of Psychology, Università degli Studi di Milano–Bicocca,
Piazza Ateneo Nuovo, 1, 20126 Milano, Italy
E-mail: carlo.reverberi@unimib.it

A recently proposed theory on frontal lobe functions claims that the prefrontal cortex, particularly its dorsolateral aspect, is crucial in defining a set of responses suitable for a particular task, and biasing these for selection. This activity is carried out for virtually any kind of non-routine tasks, without distinction of content. The aim of this study is to test the prediction of Frith's 'sculpting the response space' hypothesis by means of an 'insight' problem-solving task, namely the matchstick arithmetic task. Starting from Knoblich *et al.*'s interpretation for the failure of healthy controls to solve the matchstick problem, and Frith's theory on the role of dorsolateral frontal cortex, we derived the counterintuitive prediction that patients with focal damage to the lateral frontal cortex should perform better than a group of healthy participants on this rather difficult task. We administered the matchstick task to 35 patients (aged 26–65 years) with a single focal brain lesion as determined by a CT or an MRI scan, and to 23 healthy participants (aged 34–62 years). The findings seemed in line with theoretical predictions. While only 43% of healthy participants could solve the most difficult matchstick problems ('type C'), 82% of lateral frontal patients did so (Fisher's exact test, *P* < 0.05). In conclusion, the combination of Frith's and Knoblich *et al.*'s theories was corroborated.

Keywords: executive functions; frontal cortex; insight, problem solving; thinking

Abbreviations: CTL = control group; LAT = lateral frontal; MED = medial frontal

Received March 2, 2005. Revised May 2, 2005. Accepted May 22, 2005. Advance Access publication June 23, 2005

Introduction

The prefrontal cortex is known for being involved in the successful execution of a wide variety of tasks. For example, it has been claimed, both in the neuroimaging and in the neuropsychological literature, that frontal lobes are involved in episodic memory (Stuss *et al.*, 1994; Tulving *et al.*, 1994), semantic memory (Henry and Crawford, 2004; Thompson-Schill *et al.*, 1997), planning (Shallice, 1982), attentional switching (Nagahama *et al.*, 1996; Stuss *et al.*, 2000), and reasoning (Goel and Dolan, 2004; Reverberi *et al.*, 2005).

One parsimonious approach to cope with such a variety of functions is that of hypothesizing a function for the prefrontal cortex that is sufficiently abstract to contribute to all the tasks listed above (Duncan, 2001). A recently proposed theory adopts this strategy: it claims that the prefrontal cortex, particularly its dorso-lateral aspect, is crucial for defining a set of responses suitable for a particular task, and biasing these for selection ('sculpting the response space'; Frith, 2000). This activity could be carried out for virtually any kind of non-routine tasks, regardless of content. For example, if a lay person is asked to plan her/his dream house, she/he will likely not consider the possibility of building the roof with caramel, cork and ice. In Frith's framework this reduction in the number of alternative hypotheses considered, and the building of the *ad hoc* category of 'suitable material for roof making' is ascribed to frontal lobes. To date, the corroborating evidence mainly consists of neuroimaging studies that either directly refer to Frith's theory (Fletcher *et al.*, 2000; Nathaniel-James and Frith, 2002) or belong to different but compatible theoretical views (e.g. Duncan *et al.*, 2000; Kerns *et al.*, 2004; Koechlin *et al.*, 2003; Thompson-Schill *et al.*, 1997). Just a few neuropsychological studies have considered the issue so far, and only within the domains of semantic memory and language processing (Metzler, 2001; Thompson-Schill *et al.*, 1998).

The aim of the present study is to test the prediction of Frith's 'sculpting the response space' hypothesis by means of a problem-solving task of the 'insight' type (Sternberg and Davidson, 1995), namely the matchstick arithmetic task (Knoblich *et al.*, 1999).

In this task, a false arithmetic statement, written using roman numerals (e.g. 'I', 'II' and 'IV'), operations ('+' and '−') and an equal sign ('=') all composed of matchsticks, can be transformed into a true statement by moving only one stick from one position to another within the pattern. It has been shown that some types of matchstick arithmetic problems are quite difficult for healthy participants, particularly those in which the only solution involves changing the operators (+ and −), producing a tautological statement (e.g. IV = IV = IV), or both (Knoblich *et al.*, 1999). Knoblich and collaborators interpreted this pattern of results as suggesting that the initial representation of those kinds of problem was biased by two strong constraints, the 'operator constraint' and the 'tautology constraint', that prevented participants from considering and evaluating the correct solutions. In Frith's terminology, the response space that was sculpted by the dorsolateral frontal lobes of the healthy participants excluded the correct responses. Thus the search for them was unsuccessful, at least until the response space was revised. Knoblich *et al.*'s (1999) hypothesis was corroborated in another study (Knoblich *et al.*, 2001) in which eye movements of participants solving the same problems were recorded. They were able to show that participants, during the initial phase, attended more frequently to the result and the operands of the equations than to the operator and the equal sign (e.g. for 'V = III − II' healthy participants attended more to 'V', 'III' and 'II' than to '=' and '−'). They also found that solution finding corresponded to an increased number of eye movements towards the previously neglected but crucial element of the equation. Knoblich *et al.* argued that in this second phase participants revised the initial misleading representation of the problem, thus eliminating ('relaxing') the inappropriate constraints, and could access the solution.

By combining both Knoblich *et al.*'s (1999) interpretation for the failure of healthy controls on the matchstick problem, and Frith's theory on the role of dorsolateral frontal cortex, we derived the counterintuitive prediction that a group of patients with focal damage to the lateral frontal cortex should perform better than a group of healthy participants on this rather difficult task. Indeed, if both theories held true (A, Knoblich *et al.*'s; B, Frith's), the cognitive factors causing the inadequate performance of healthy participants (the two detrimental constraints, 'operators' and 'tautology', following theory A) would no longer be present in a group of lateral frontal patients (following theory B). Hence the prediction of lateral frontal patients performing *better* than a control group on the matchstick task.

In this study we administered a revised version of the matchstick arithmetic task (Knoblich *et al.*, 1999) to a group of patients with focal brain damage to the frontal lobes. Since Frith's theoretical claim only regards the lateral surface of the frontal lobes, we divided our series into lateral and medial damaged patients. The crucial test for the theory will be that on lateral patients; data from the medial subgroup will be useful, at most, for estimating the anatomical specificity of the phenomenon.

The prediction of patients outperforming healthy participants provides a methodological advantage. Since in only very rare instances a lesion to a cognitive structure produces an improvement in performance (e.g. Warrington and Davidoff, 2000), it can be reasonably assumed that all processes and representations occurring before the crucial stage in the information flow (in the present case the 'response space sculpting' function) are relatively spared, without being forced to test this hypothesis with additional manipulation or diagnostic tools. This advantage applies equally well to all studies driven by similar predictions.

Materials and methods

Participants

Thirty-five 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 in the study; the study was approved by the ethical committee of SISSA-ISAS (International School for Advanced Studies, Trieste). The aetiology varied across patients: stroke, neoplasm and arachnoid cyst (Table 1). Exclusion criteria were (i) a clinical history of psychiatric disorders, substance abuse or previous neurological disease, (ii) neuroradiological evidence of diffuse brain damage, (iii) age <18 or >70 and (iv) educational level <8 years. Time after lesion onset (Table 2) ranged between 7 and 1507 days (the onset date considered in the case of neoplasm was that of surgery). Twenty-three normal control volunteers also participated in the study. Controls were matched to patients for age and educational level. There were no significant differences between frontal patients and controls both for age [$t(56) = 0.388$, $P > 0.1$] and education [$t(56) = 0.403$, $P > 0.1$].

Neuroradiological assessment

For all patients we obtained a CT or an MRI scan. Patients were assigned to two anatomically defined subgroups depending on their lesion site: medial (MED), in which the lesion involved the orbital and/or the medial surface on one or both frontal lobes, and lateral (LAT), showing unilateral damage to the frontal lobe convexity

Table 1 Aetiology for each lesion group

	LAT	MED	Patients overall
Arachnoid cyst		1	1
Glioma high grade	2	1	3
Glioma low grade	3	6	9
Meningioma	10	9	19
Stroke	2	1	3

Absolute frequencies of patients included in the study.
LAT, lateral frontal; MED, medial frontal.

Table 2 Demographic and clinical variables for each lesion group and for control participants

	LAT	MED	Patients overall	CTL
N	17	18	35	23
Age [mean (SD)]	47 (13)	47 (11)	47 (11)	46 (9)
Education [mean (SD)]	10.59 (2.83)	10.44 (2.64)	10.51 (2.69)	10.22 (2.83)
Sex [female proportion (%)]	50	71	60	56
Lesion volume (cc) [mean (SD)]	50 (47)	44 (32)	47 (40)	
Days since lesion [median (range)]	653 (7–1314)	360 (7–1507)	619 (7–1507)	

LAT, lateral frontal; MED, medial frontal; CTL, control group.

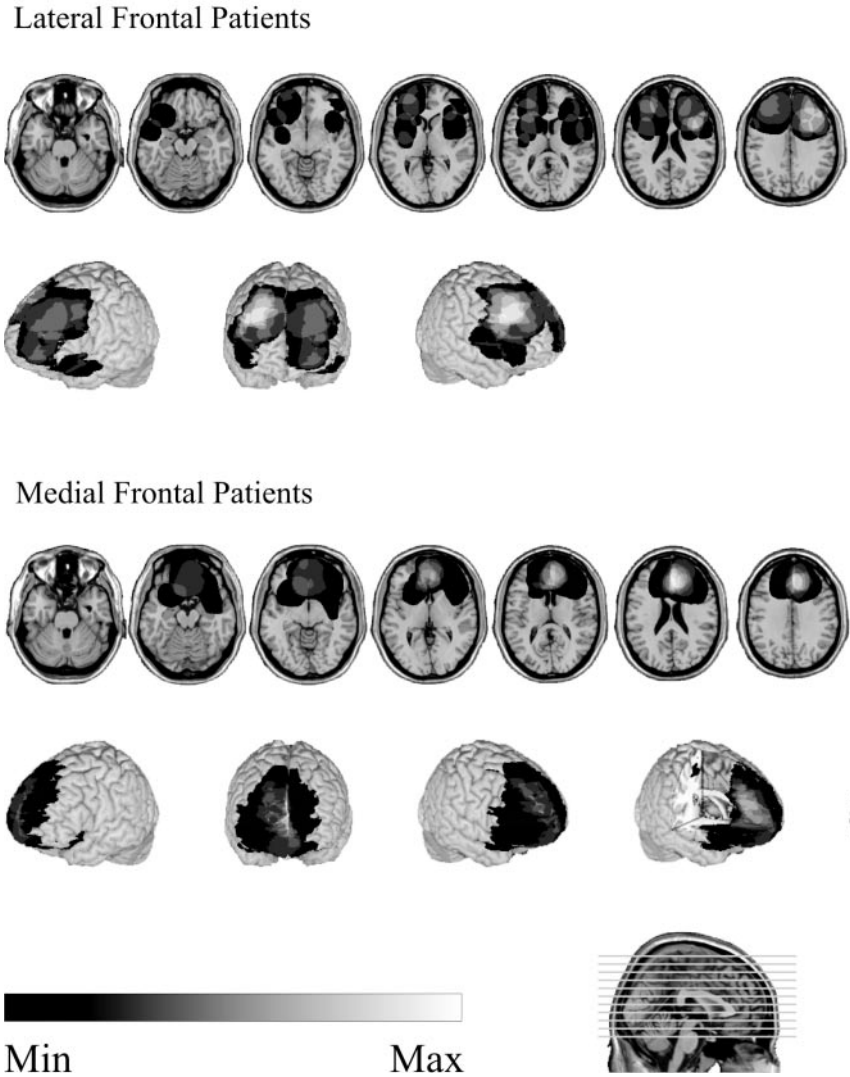

Fig. 1 Overlay lesion plots for the two lesion subgroups. The number of overlapping lesions in each voxel is illustrated on a grey scale: the lighter a voxel, the higher the number of patients with damage to that. The grey scale is devised so that voxels that were damaged with maximal frequency within a patient subgroup are shown in white. Thus, white areas were damaged in 6 out of 17 lateral frontal patients, and in 10 out of 18 medial frontal patients. Talairach z-coordinates (Talairach and Tournoux, 1988) of each transverse slice are 45, 55, 65, 75, 85, 95, 105, 115, 125, 135, 145.

(Fig. 1). This classification was carried out by a senior neuroradiologist (S.D.A.) blind to the behavioural results. All lesions were also mapped using the free MRIcro (www.mricro.com) software distribution (Rorden and Brett, 2000) and were drawn manually 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 oriented to approximately match the Talairach space (Talairach and Tournoux, 1988) and is distributed with MRIcro. As a final result, 17 patients were classified as lateral and 18 as medial. Lesion volume was also obtained by means of the automatic routines of MRIcro.

Materials and procedure

A matchstick arithmetic problem (Knoblich *et al.*, 1999) consists of a false arithmetic statement written with roman numerals (I, II, III, etc.), arithmetic operations (+, −), and equal signs constructed out of matchsticks. For example in

$$\mathrm{IV} = \mathrm{III} + \mathrm{III} \qquad (1)$$

the participant is required to move a single stick in such a way that the initial false statement is transformed into a true arithmetic statement. A move consists of grasping a single stick and moving it, rotating it or sliding it. The rules are that (i) only one stick can be moved, (ii) a stick cannot be discarded and (iii) the result must be a correct arithmetic statement. Two additional rules are that (iv) an isolated slanted stick cannot be interpreted as I (one) and that (v) a V symbol must always be composed of two slanted sticks.

As an example, consider the false equation (1) above: it can be transformed into the true equation by moving the left-most stick of number 'IV' to the immediate right of the 'V':

$$\mathrm{VI} = \mathrm{III} + \mathrm{III} \qquad (2)$$

All matchstick arithmetic problems are composed by three roman numerals separated by two arithmetic signs and have a unique solution, consisting of a single move. Hence, differences in difficulty are solely a function of how hard it is to think of the correct move.

Three different classes of problems can be identified on the basis of the kind of move necessary to achieve the solution. These classes are:

Type A

This type of problem is solved by moving a matchstick that is part of a numeral, to another numeral. For example, the problem 'II = III + I' is solved by moving one of the matchsticks of the 'III' to the 'II' in head position.

Type B

In this case it is necessary to move a matchstick from the equal sign to the minus sign, in order to change it into an equal sign. Thus, e.g. the false equation 'IV = III − I', should be transformed in the true 'IV − III = I'.

Type C

In this last problem type, a plus sign has to be changed, by rotating its vertical matchstick through 90°, into an equal sign. Crucially, this action transforms the starting equation into a tautology; e.g. 'VI = VI + VI' becomes 'VI = VI = VI'.

Four blocks of three problems were administered (Table 3). In each of the four blocks, an equation of each type was presented in pseudo-random order. The equations were built by the experimenter on the table in front of the participants by using real matchsticks. Participants were allowed to touch and move the matchsticks while they were looking for the solution. Three minutes were given to solve each problem.

If a participant failed to solve a problem on the first block, the solution was *not* shown, and the next problem was presented. By contrast, if the participant failed in the succeeding blocks, a cueing procedure followed each unsuccessful trial, with 'first-level' and 'second-level' cues (Table 3). Both levels consisted of suggesting to the participant to avoid moving or changing some of the components of the equation. The second-level cue was more informative than the first-level one, and included the information given in it. If the participant failed after the first-level cue, she/he was given the second-level cue. For example, for the problem 'IV = III = I', unsuccessful participants were informed that they should leave unchanged:

(i) (first-level cue) 'II': V = III – **II** i.e. participants were informed that they would not find the solution by changing the roman numeral 'II' (bold and underlined in the example);

(ii) (second-level cue) 'V', 'III' and 'II' : **V** = **III** – **II** i.e. participants were informed that they would not find the solution by changing the roman numerals 'V', 'III' or 'II' (bold and underlined in the example).

After each cue, a further minute was granted to look for the solution. To make the cues clear and readily available throughout the whole thinking period, the elements that were not to be changed were composed of black sticks of the same size as the original ones.

Accuracy and solution time were collected for each problem.

Variables

We analysed the following variables.

(i) Index of success ('success score' henceforth). This index estimated the participants' ability to solve the problems without assistance, i.e. before the delivery of cues. Therefore, the index was derived from the performance on the first two attempts to solve a problem type (first two blocks). The index was dichotomous: if the participant could work out the correct answer in at least one of those two attempts, she/he was given a 'pass' score (1); a 'fail' score (0) was given otherwise. Since the index was

Table 3 Matchstick problems in the experimental presentation order

Block	Type	Problem	Solution	Cue 1	Cue 2
1	B	IV = III − I	IV − III = I	/	/
	A	VI = VII + I	VII = VI + I	/	/
	C	III = III + III	III = III = III	/	/
2	A	IV = III + III	VI = III + III	IV = **III** + III	IV = **III** + **III**
	B	V = III − II	V − III = II	V = III − **II**	**V** = **III** − **II**
	C	VI = VI + VI	VI = VI = VI	VI = **VI** + VI	**VI** = **VI** + **VI**
3	B	VIII = VI − II	VIII − VI = II	VIII = **VI** − II	**VIII** = **VI** − **II**
	C	IV = IV + IV	IV = IV = IV	**IV** = IV + IV	**IV** = **IV** + **IV**
	A	II = III + I	III = II + I	II = III + **I**	II = III **+** **I**
4	C	VII = VII + VII	VII = VII = VII	VII = VII + **VII**	**VII** = **VII** + **VII**
	A	VII = II + III	VI = III + III	VII = II + **III**	VII = II **+** **III**
	B	VI = IV − II	VI − IV = II	VI = IV − **II**	**VI** = **IV** − **II**

specific of the problem type, each participant received three dichotomous scores: 'success A', 'success B', 'success C'.

(ii) Index of accuracy after relaxation of constraint ('relax score'). When a subject solves a problem for the first time, either autonomously or helped by the cueing procedures, the solution-preventing constraint is removed ('relaxed'). Therefore, the performance on that problem type after the first success expresses the participant's ability to solve it in the absence of constraints. This ability was estimated by means of the 'relax' score, i.e. the proportion of times a participant was able to find the solution (without cues) after having solved that specific problem type (with or without cues) for the first time. For instance, if a participant obtained the following scores for type A problems [1st block: 0; 2nd block: 0 (after cue, **1**); 3rd block: 1; 4th block: 0 (after cue, 1)], the participant was given a type A relax index of 0.5. Indeed, after the first success (reported in bold), she/he obtained one pass out of two uncued attempts (underlined). Three relax indices were thus obtained, one for each problem type ('relax A', 'relax B', 'relax C').

Appropriateness of the task for testing theoretical predictions

Replication of Knoblich et al.'s (1999) results

Since our control and patient samples tapped a different and wider range of age and education than that of the original study (Knoblich *et al.*, 1999), a first step was to see whether or not we were able to replicate its main findings. Therefore, normal participants should show a significantly better performance on low-constraint problems (type A) than on high-constraint problems (types B and C).

Specific form of the main prediction

The main prediction of this study is that patients should perform better than controls on high-constraint problems (types B and C). That is, they should outperform controls on Success B and Success C scores. This should happen if (i) the Frith–Knoblich theory is true and (ii) patients do not have concomitant deficits, i.e. deficits other than the inability to sculpt the response space, affecting their matchstick performance (e.g. processing difficulties; Kershaw and Ohlsson, 2004). Indeed, if such supplementary deficits were present, they would mask the advantage of patients by reducing it, nullifying it, or even reverting it into a disadvantage. Therefore, in order to provide a valid test of the Frith–Knoblich prediction, problem types that were found to contain, for the patients, difficulties other than the constraints, had to be excluded.

Such extra difficulties would be apparent from looking at Relax scores (that are independent of Success scores, on which the crucial prediction was made), for the following reasons. Suppose that the constraints were the only difficulties. In this case, after their relaxation, performance should reach a high level both in patients and in controls, without group differences. Suppose instead that patients had supplementary difficulties. If this were the case, they would still show a disadvantage with respect to controls, even after constraint relaxation. Therefore, problem types that produced a significant disadvantage for patients versus controls after relaxation (i.e. on relax scores) had to be excluded.

By considering these remarks, the Frith–Knoblich prediction takes a more specific form: frontal lateral patients should obtain success scores higher than those of normal controls, on problem types that were solved with high probability and equally well by both groups after constraint relaxation (relax scores).

Planned statistical analyses

Success scores

Success scores were dichotomous (pass–fail); thus, nominal-scale statistics were appropriate in this case. Fisher's Exact test was used for between-subjects comparisons. The McNemar test and Cochran's Q were used for, respectively, two- and three-level within-subjects comparisons.

The performance of control and patient groups during the first two blocks was compared for each type of problem. Moreover the performance on each high-constraint problem type (B and C) was compared with that on the low-constraint problem type (A). Only the types of problem meeting the criterion discussed above (Section Specific form of the main predicition) were considered. Since we had *a priori* expectations on the direction of the effects, one-tailed *P*-values were used. *P*-values were all estimated from Exact distributions.

Relax scores

Relax scores ranged between 0 and 1, with few intermediate values (0.33, 0.66). Therefore, ordinal-level non-parametric analyses were prudently applied. The Mann–Whitney test was carried out for between-subjects comparisons, while Wilcoxon and Friedman tests were applied for, respectively, two- and three-level within-subjects comparisons. Since we expected patients to show, if anything, a worse Relax index than controls, one-tailed *P*-values were considered. *P*-values were all estimated from Exact distributions.

Results

Effects of clinical variables

A logistic regression analysis was carried out on both the success and the relax scores of the patient group. Predictors were lesion volume (cc) and the logarithm of post-onset time (days). No significant effect of either variable on either score was found.

Appropriateness of the task

Healthy participants

Success indices were significantly different according to problem type (A, B and C) in the control group (Cochran's Q = 10.778, $P < 0.01$; see Fig. 2). As expected, control participants scored significantly better on problem type A than on B (McNemar test, $P < 0.05$) and C (McNemar test, $P < 0.01$). Relax indices (Fig. 3) did not differ significantly for different problem types (Friedman test, $\chi^2 = 3.800$, $P > 0.1$). This pattern of results replicated successfully the main findings of Knoblich *et al.* (1999) in an older and less educated population.

Patient groups

Lateral frontal patients showed a significant effect of problem type (A, B, C) on relax indices (Friedman test, $\chi^2 = 17.429$, $P < 0.001$; see Fig. 3). This finding is attributable to type B problems, which had a significantly lower relax score than

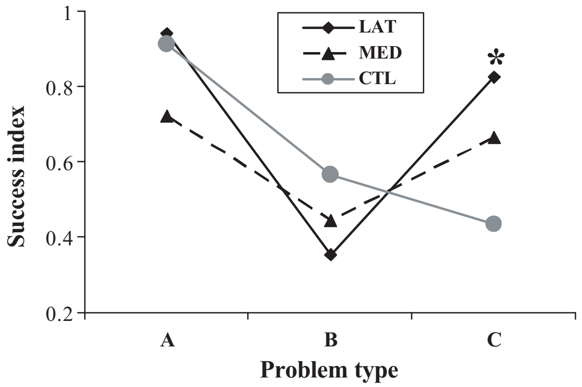

Fig. 2 Success scores of patient and control groups on the matchstick arithmetic task for each problem type. LAT, lateral frontal; MED, medial frontal; CTL, control group. *$P < 0.05$, **$P < 0.01$ for the control group versus patient subgroups comparisons.

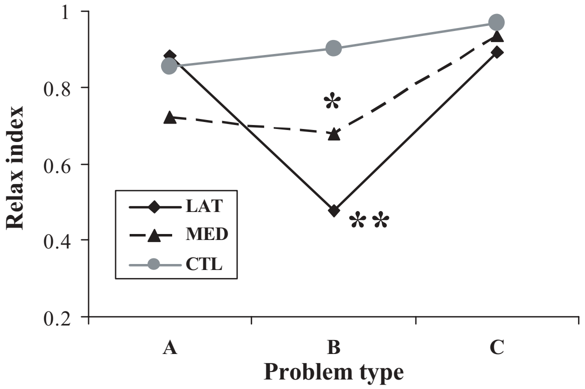

Fig. 3 Relax score of patient and control groups on the matchstick arithmetic task for each problem type. Conventions as in Fig. 2.

both type A (Wilcoxon test, $P < 0.01$) and type C (Wilcoxon test, $P < 0.01$) problems. By contrast, type A and type C problems had similar relax scores (Wilcoxon test, $P > 0.1$). Moreover, relax B index of lateral frontal patients was significantly lower than that of the control group (Mann–Whitney, $z = 3.150$, $P < 0.01$), but relax A (Mann–Whitney, $z = 1.405$, $P > 0.05$) and relax C (Mann–Whitney, $z = 0.878$, $P > 0.1$) were not.

In medial frontal patients relax indices for A, B and C types (Fig. 3) showed a statistical trend in being different from each other (Friedman test, $\chi^2 = 5.448$, $P < 0.1$). This trend is attributable to an advantage of type C over both type B (Wilcoxon test, $P < 0.01$) and type A (Wilcoxon test, $P < 0.01$). Relax B scores of medial frontal patients were significantly lower than those of the control group (Mann–Whitney, $z = 2.343$, $P < 0.05$). This difference was not found for relax A (Mann–Whitney, $z = 1.346$, $P > 0.05$) or relax C (Mann–Whitney, $z = 0.758$, $P > 0.1$).

The relax score of patients (from both anatomical subgroups) on type B problems was significantly lower with respect to both the controls' score on that same type, and the scores of the patients themselves on problem types A and C. Thus, according to the above-mentioned criteria (see the section Specific form of the main prediction) type B was not considered in the following analyses.

Comparison between success scores of control and patient subgroups

The type A success score of the lateral group was not significantly different from that of the control group (Fisher's exact test, $P > 0.1$). In contrast, lateral patients performed significantly better than controls on problem type C (Fisher's exact test, $P < 0.05$). Within the lateral group, A and C success scores were not significantly different from each other (McNemar test, $P > 0.1$).

The success score of the medial group was not significantly different from that of controls both on type A and on type C problems (Fisher's exact test, $P > 0.1$). Within the medial group, A and C success scores were not significantly different from each other (McNemar test, $P > 0.1$).

Is the patients' profile specific of the lateral subgroup?

The success profile of medial patients was not significantly different from both controls' and lateral patients' profiles (Fig. 2). Given this lack of significance it is not possible to rule out either of the extreme possibilities: that the real medial patients' profile matches closely the controls' or the lateral patients' pattern. One possible suggestion is that medial patients' profile fell in between the other two groups' profiles (see especially type C problems, Fig. 2), because some of the medial patients had a minor involvement of the lateral cortex surface. If this were the case, by excluding patients with such minor involvement from the medial group, the pattern of residual patients should match that of controls. We selected, in a further analysis, the subset of the medial group that had pure medial or orbito-frontal damage, that is patients with no involvement whatsoever of the frontal lobe convexity ($n = 9$; Fig. 1, Supplementary material). Interestingly, their mean success score on type A was only 56%, significantly lower than both lateral patients' (Fisher's exact test, $P < 0.05$) and controls' (Fisher's exact test, $P < 0.05$) scores. It is thus clear that pure medial patients cannot be assumed to have an entirely normal profile. The crucial type C success score was also 56%, not significantly different from both controls' and lateral patients' scores. Overall, the success profile of pure medial group was essentially flat (the effect of problem type was not significant; Cochran's Q = 1.143, $P > 0.05$) and quite far from optimal (type A: 56% = 5/9; B: 33% = 3/9; C: 56% = 5/9; see Fig. 2 and Table 1, Supplementary material).

Discussion

It is widely acknowledged that frontal cortex is crucial in order to cope with problems that are novel and 'difficult'. In this study, we tested a prediction that was deduced from the combination of Frith's theory (Frith, 2000) on lateral frontal cortex functions with Knoblich and collaborators' observational theory on the matchstick arithmetic task (Knoblich *et al*., 1999). The counterintuitive prediction was that lateral frontal patients should perform better than healthy controls on the most difficult trials of a novel task.

Two requisites were necessary for our experiment to be an adequate test of the proposed prediction. First, our revised version of Knoblich *et al*.'s task should have produced, from healthy participants, the same results pattern as that of the original study, in spite of age and education differences. Second, problem types should have been excluded that encompassed, for lateral patients, difficulties other than the triggering of inappropriate constraints, i.e. problems that were still performed below normal level after 'relaxation' of those constraints. In this way, the possibility that concomitant deficits at stages other than the one of interest (the constraint implementation stage) masked the predicted supra-normal performance by lateral patients would have been ruled out.

Both of these prerequisites were met. Although older and less educated, our control subjects showed a progressively declining performance from type A to B to C (91, 57, 43%, respectively, see Fig. 2), as did Knoblich *et al*.'s participants. Furthermore, problem type B was excluded because relaxation of constraints, for lateral patients, did not lead to normal performance. Therefore, by analysing only problem types A and C we could test the prediction generated by the Frith–Knoblich model.

The pattern of results closely matched the prediction. Lateral frontal patients were as successful as healthy controls in solving type A problems, which have weak constraints, but significantly more successful than controls on type C problems, which have strong constraints. Besides statistical significance, the difference was far from negligible (82% of lateral patients solved those problems, while only 43% of controls did so). The hypothesis that lateral frontal patients constrain the response space less than controls do, is thus corroborated. According to this view, it is as if lateral frontal patients faced a problem with a trial-and-error approach without a prior assessment of the likely fruitfulness or appropriateness of a strategy. They would simply explore the whole of the response space (see appendix I). In the artificial situation of the matchstick arithmetic task, this procedure can be an advantage, but in real life situations, in which a preliminary downgrade of the possibilities to be considered is a necessary step in order to make problems tractable, this produces less efficient and more disorganized behaviour (Shallice and Burgess, 1991).

The hypothesis that lateral patients apply a trial-and-error approach could also account for their impairment on type B problems. By evaluating problems, without any constraint whatsoever apart from respecting the composition rules for roman numerals and the rules of equation writing it can easily be observed that type B problems have about twice as many possible matchstick moves (on average, 10.2) as types A and C problems (respectively 5.7 and 5.5). Thus, it is possible to speculate that lateral patients' advantage in having no constraints is not enough to compensate for the higher computational load induced by problems with a larger response space (Kershaw and Ohlsson, 2004) (see appendix II).

Processing contextual hints

One might wonder whether the inability to implement constraints is the only possible explanation for the present findings from the matchstick arithmetic task. Another explanation might be that in this task the 'impasse' arises in healthy participants, because they inappropriately represent the problems as if they came from the superficially similar field of algebra (Knoblich *et al*., 1999). For instance:

(i) Healthy participants might initially think that the solution should respect the general prototypical form of an equation, with at least an equal sign and an operator (+ or −). Thus tautologies (e.g. III = III = III) are not considered appropriate solutions (type C problems);
(ii) Healthy participants might think that the only variable parts of an equation are numbers (type B and C problems).

In this view, the behaviour of lateral frontal patients could be explained by proposing that they do not take into consideration contextual hints, as proposed by a number of theories of frontal lobe functions (e.g. Braver and Cohen, 2000; Kerns *et al*., 2004; Metzler, 2001). In the present case, the superficial similarities of the matchstick problems to simple equations could constitute the relevant contextual hint. Frontal patients may not be able to take advantage—or disadvantage as in the present case—of tricks that apply to the more familiar field of simple mathematical problems (see the two examples above). More generally, lateral frontal patients would not 'enrich' their representation of the problematic situation (more specifically, of the 'goal state' in cognitivistic terms) with past experience, as healthy controls do. This explanation would link the present findings to evidence suggesting a crucial role of lateral frontal cortex in the efficient encoding of novel stimuli (particularly in the memory domain), e.g. through chunking (Bor *et al*., 2003).

According to this explanatory hypothesis, lateral frontal patients would still be able to implement constraints on the response space but they would not do so because they fail to envisage or generate possible candidates to the role of constraint. The present study does not provide enough evidence to decide whether our findings are related to the inability to 'sculpt the response space' by implementing constraints, or to the inability to generate candidates to the role of constraint before implementation.

Medial frontal patients

The general pattern of success scores of medial patients seemed quite similar to that of lateral patients, although less marked (see Fig. 2). Nevertheless, from the statistical viewpoint, this medial profile was not distinguishable also from that of the controls. Given the present data, two alternative accounts may be proposed.

(i) Frith's claim about the anatomical structure subserving constraints implementation (lateral frontal) is correct. In the present medial group, some patients had also the lateral cortical surface involved, which would explain the partial resemblance of the medial group's to the lateral group's success profile.

(ii) The lateral frontal cortex is not the only region involved in constraints implementation, and perhaps some medial structure is also crucial in these processes (such as the anterior cingulate, as Duncan and collaborators would argue, see Duncan and Owen, 2000).

The prediction of the first hypothesis (i) is that in the pure medial subgroup, patients with a lesion of the medial or orbitofrontal cortex but without any involvement whatsoever of the frontal lobe convexity, should show, as controls do, a clear effect of constraint level. Type C success score should be lower than type A score (type B is to be excluded as discussed in the section Specific form of the main prediction). By contrast, if hypothesis (ii) held true, pure medial patients should show, as lateral patients do, no effect of constraint level. They should show the same performance on types A and C and a clear advantage on problem type C with respect to the control group. The pattern of results (see Fig. 2 and Table 1, Supplementary material) is more compatible with hypothesis (ii). Pure medial patients obtained the same score (56%) on both type A and type C problems. However, unlike lateral patients' scores, their performance was clearly suboptimal. For instance, pure medial patients obtained a significantly lower type A score (56%) with respect to both lateral (94%) and control (91%) groups. An additional cognitive deficit (e.g. of initiation) is to be assumed in order to explain why pure medial patients did not take full advantage, as lateral patients did, from the absence of constraints.

In summary, pure medial patients' performance may be interpreted as the effect of a combination of an aspecific cognitive deficit (e.g. a general slowing down), that accounts for the suboptimal level of performance, and a lack of constraints on response space, that accounts for the similarity of the success index across problem types. In any case, the pure medial sample size ($n = 9$) makes conclusions premature in this respect.

Conclusions

This study provides strong evidence in favour of a role of lateral—and perhaps medial—frontal structures in biasing response space. Further evidence, allowing one to carry out a single-case estimation of biasing ability, would be useful in order to build more fine-grained anatomo-functional maps.

One final remark regards the advantage of testing theoretically-driven predictions of a better performance in a neurological population with respect to a control population. Brain lesions produce deficits, i.e. performances that are lower than in healthy controls, much more frequently than improvements. Therefore, models of the cognitive architecture have been proposed in cognitive neuropsychology that when 'damaged' predict deficits. As a consequence if one wants to drive inferences about stage X, she/he has to guarantee that there are no deficits both downstream and upstream from X in the information flow. In a trivial example, both blindness and plegia have to be excluded if one wants to study visual depth perception in a task in which the patient has to respond by pressing a key. If the prediction of a cognitive theory is that of a paradoxical improvement of performance after a lesion of stage X, there is no need to test for integrity all the other processing stages. If these latter were damaged, this would, if anything, predict a deficit in the final performance, not an improvement. The advantage is, therefore, in terms of much fewer tests to be administered to each participant. Furthermore, theories that predict counterintuitive phenomena, like improvement after brain damage, tend to have relatively fewer rival theories. Thus, whenever a counterintuitive prediction is confirmed, it represents a strong corroboration in favour of the originating theory (Lakatos, 1978).

Supplementary material

The Supplementary material cited in this article is available at *Brain* on-line.

Acknowledgements

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). We are grateful to Günther Knoblich, Tim Shallice and Claudio Luzzatti for their thoughtful comments on earlier drafts of the paper. We are also indebted to Ms Sarah J. Healy for her kind help in revising the English.

Appendices

Appendix 1

In this context an unpublished finding from a former study by our lab (Reverberi *et al.*, 2005) is worth mentioning. In this work, the authors administered the Brixton Spatial Rule Attainment task to a sample of frontal patients. On this task, participants were presented with a card containing a 2 × 5 display of circles one only of which was coloured in blue. The blue circle moved from one card to the next following five rules. Participants had to predict which circle should be blue on the next card. One of the rules was 'stay the same', that is the blue circle remained in the same position across six cards. This rule was quite difficult for the control group (the rule was attained only by 32% of the healthy participants, see Fig. 3, Supplementary material), but not for the right lateral damaged patients (70%, significantly

>32%). If we interpret the difficulty of the control group as arising because of an *a priori* constraint—not specified in the instructions—such as 'the blue circle should move in some direction', the advantage of right lateral patients could be explained by appealing to Frith's theory. In the present study, the advantage observed in lateral frontal patients tended to be higher in right than in left unilateral patients (Right: $N = 9$; success type C = 0.89; significantly higher than the control group's one; Left: $N = 8$; success type C = 0.75; not significantly higher than the control group's one).

Appendix II

Some examples of computation of the number of possible moves follow:

(i) Problem: IV = IV + IV. Four possible moves: VI = IV + IV; IV = VI + IV; IV = IV + VI; IV = IV = IV.
(ii) Problem: V = III − II. Ten possible moves: VI = II − II; IV = II − II; VI = III − I; IV = III − I; V = II − III; V = III + I; V = II + II; V = II = II; V = III = I; V − III = II.

The stimuli used in the present work are a subset of Knoblich *et al.*'s (1999) problems. At an early stage of their work (G. Knoblich, personal communication) the authors found no effect of number of possible moves in a set of healthy participants. Hence, this variable was not taken into account both in their published study and in the present work. Further investigation in which problem type and number of possible moves are independently manipulated could help clarify the relevance of the latter factor in a patient population.

References

Bor D, Duncan J, Wiseman RJ, Owen AM. Encoding strategies dissociate prefrontal activity from working memory demand. Neuron 2003; 37: 361–7.

Braver TS, Cohen JD. On the control of control. The role of dopamine in regulating prefrontal function and working memory. In Monsell S, Driver J, editors. Attention and performance XVIII: control of cognitive performance. Cambridge, MA:MIT Press; 2000. p. 713–37.

Duncan J. An adaptive coding model of neural function in prefrontal cortex. Nat Rev Neurosci 2001; 2: 820–9.

Duncan J, Owen AM. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci 2000; 23: 475–83.

Duncan J, Seitz RJ, Kolodny J, Bor D, Herzog H, Ahmed A, et al. A neural basis for general intelligence. Science 2000; 289: 457–60.

Fletcher PC, Shallice T, Dolan RJ. "Sculpting the response space"—an account of left prefrontal activation at encoding. Neuroimage 2000; 12: 404–17.

Frith CD. The role of the dorsolateral prefrontal cortex in the selection of action as revealed by functional imaging. In: Monsell S, Driver J, editors. Control of cognitive processes: attention and performance XVIII. Cambridge, MA: MIT Press; 2000. p. 549–65.

Goel V, Dolan RJ. Differential involvement of left prefrontal cortex in inductive and deductive reasoning. Cognition 2004; 93: B109–21.

Henry JD, Crawford JR. A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology 2004; 18: 284–95.

Kerns JG, Cohen JD, Stenger VA, Carter CS. Prefrontal cortex guides context-appropriate responding during language production. Neuron 2004; 43: 283–91.

Kershaw TC, Ohlsson S. Multiple causes of difficulty in insight: the case of the nine-dot problem. J Exp Psychol Learn Mem Cogn 2004; 30: 3–13.

Knoblich G, Ohlsson S, Haider H, Rhenius D. Constraint relaxation and chunk decomposition in insight problem solving. J Exp Psychol Learn Mem Cogn 1999; 25: 1534–55.

Knoblich G, Ohlsson S, Raney GE. An eye movement study of insight problem solving. Mem Cognit 2001; 29: 1000–9.

Koechlin E, Ody C, Kouneiher F. The architecture of cognitive control in the human prefrontal cortex. Science 2003; 302: 1181–5.

Lakatos I. The methodology of scientific research programmes. Cambridge (UK): Cambridge University Press; 1978.

Metzler C. Effects of left frontal lesions on the selection of context-appropriate meanings. Neuropsychology 2001; 15: 315–28.

Nagahama Y, Fukuyama H, Yamauchi H, Matsuzaki S, Konishi J, Shibasaki H, et al. Cerebral activation during performance of a card sorting test. Brain 1996; 119: 1667–75.

Nathaniel-James DA, Frith CD. The role of the dorsolateral prefrontal cortex: evidence from the effects of contextual constraint in a sentence completion task. Neuroimage 2002; 16: 1094–102.

Reverberi C, Lavaroni A, Gigli GL, Skrap M, Shallice T. Specific impairments of rule induction in different frontal lobe subgroups. Neuropsychologia 2005; 43: 460–72.

Rorden C, Brett M. Stereotaxic display of brain lesions. Behav Neurol 2000; 12: 191–200.

Shallice T. Specific impairments of planning. Philos Trans R Soc Lond B Biol Sci 1982; 298: 199–209.

Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. Brain 1991; 114: 727–41.

Sternberg RJ, Davidson JE. The nature of insight. Cambridge, MA: MIT Press; 1995.

Stuss DT, Alexander MP, Palumbo CL, Buckle L, et al. Organizational strategies with unilateral or bilateral frontal lobe injury in word learning tasks. Neuropsychology 1994; 8: 355–373.

Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, et al. 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 2000; 38: 388–402.

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

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci USA 1997; 94: 14792–7.

Thompson-Schill SL, Swick D, Farah MJ, D'Esposito M, Kan IP, Knight RT. Verb generation in patients with focal frontal lesions: a neuropsychological test of neuroimaging findings. Proc Natl Acad Sci USA 1998; 95: 15855–60.

Tulving E, Kapur S, Craik FI, Moscovitch M, Houle S. Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. Proc Natl Acad Sci USA 1994; 91: 2016–20.

Warrington EK, Davidoff J. Failure at object identification improves mirror image matching. Neuropsychologia 2000; 38: 1229–34.