

Mechanisms of Rule Acquisition and Rule Following in Inductive Reasoning

Cristiano Crescentini,^{1,2} Shima Seyed-Allaei,¹ Nicola De Pisapia,³ Jorge Jovicich,^{3,4} Daniele Amati,¹ and Tim Shallice^{1,4,5}

¹Cognitive Neuroscience Sector, International School for Advanced Studies, 34136 Trieste, Italy, ²Magnetic Resonance Unit, Azienda Ospedaliero-Universitaria Santa Maria della Misericordia, 33100 Udine, Italy, ³CIMeC - Center for Mind/Brain Sciences and ⁴Department of Cognitive Science and Education, University of Trento, 38068 Rovereto, Italy, and ⁵Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom

Despite the recent interest in the neuroanatomy of inductive reasoning processes, the regional specificity within prefrontal cortex (PFC) for the different mechanisms involved in induction tasks remains to be determined. In this study, we used fMRI to investigate the contribution of PFC regions to rule acquisition (rule search and rule discovery) and rule following. Twenty-six healthy young adult participants were presented with a series of images of cards, each consisting of a set of circles numbered in sequence with one colored blue. Participants had to predict the position of the blue circle on the next card. The rules that had to be acquired pertained to the relationship among succeeding stimuli. Responses given by subjects were categorized in a series of phases either tapping rule acquisition (responses given up to and including rule discovery) or rule following (correct responses after rule acquisition). Mid-dorsolateral PFC (mid-DLPFC) was active during rule search and remained active until successful rule acquisition. By contrast, rule following was associated with activation in temporal, motor, and medial/anterior prefrontal cortex. Moreover, frontopolar cortex (FPC) was active throughout the rule acquisition and rule following phases before a rule became familiar. We attributed activation in mid-DLPFC to hypothesis generation and in FPC to integration of multiple separate inferences. The present study provides evidence that brain activation during inductive reasoning involves a complex network of frontal processes and that different subregions respond during rule acquisition and rule following phases.

Introduction

Inductive reasoning is central to human learning; it takes specific instances and creates a general rule (Rips, 1999). Such inferences can be considered to occur in a series of phases. First, the instances are collected and held in memory. Second, a regularity is identified and a hypothesis generated about relationships between the instances. Third, current inferences may be integrated with older ones. Finally, generalizations are confirmed through additional observations.

Although it is thought that rule-guided behavior depends on a network of frontal, parietal, and temporal brain regions (Bunge, 2004; Bunge and Wallis, 2008), the regional specificity for the different mechanisms involved in induction tasks, such as rule learning and rule following, remains underspecified. Thus, some studies, including neurophysiological recordings in nonhuman primates (White and Wise, 1999; Goel and Dolan, 2000; Seger et al., 2000; Strange et al., 2001; Bunge, 2004) have attributed a role for the dorsolateral prefrontal cortex (DLPFC) (particularly the

left DLPFC) in the application and maintenance of rules, rather than in their initial learning. Nonetheless, recent evidence suggests that DLPFC plays a role in hypothesis generation during rule learning independently of working memory (Boettiger and D'Esposito, 2005; Reverberi et al., 2005b; Specht et al., 2009).

Frontopolar cortex (FPC) has also been proposed to contribute to rule learning. However, it is not clear whether it is important in hypothesis generation in general (Strange et al., 2001) or whether it operates in more specific conditions such as when it is necessary to integrate the outcomes of more than one inference (Ramnani and Owen, 2004; Koechlin and Hyafil, 2007). Yet another possibility is that FPC is involved in maintaining a rule active before it becomes familiar; this would fit with its known role in intention maintenance (Burgess et al., 2000, 2001) and branching (Koechlin et al., 1999).

In this study, we adapted a rule attainment task, the Brixton Spatial Anticipation Test (Burgess and Shallice, 1996), to a functional imaging context with the goal of testing the contribution of different prefrontal cortex (PFC) and temporo-parietal regions to rule acquisition and rule following. In this test, participants are presented with cards, each consisting of circles numbered sequentially of which only one is colored blue. The subjects must predict the position of the blue circle on the next card. Generally, while searching for the correct rule, participants make a series of incorrect responses before coming up with a correct hypothesis for the current rule. Once correctly acquired, a rule is typically followed perfectly until a new one comes into operation. If DLPFC is involved in hypothesis generation, regardless of the

Received Sept. 1, 2010; revised March 22, 2011; accepted March 26, 2011.

Author contributions: C.C., S.S.-A., N.D.P., J.J., D.A., and T.S. designed research; C.C., S.S.-A., and N.D.P. performed research; C.C. and S.S.-A. analyzed data; C.C. wrote the paper.

This work was partially supported by postdoctoral fellowships awarded to C.C. [International School for Advanced Studies (SISSA)] and N.D.P. (Società Scienze Mente Cervello/Fondazione Cassa di Risparmio di Trento e Rovereto), and a doctoral fellowship awarded to S.S.-A. (SISSA).

Correspondence should be addressed to Cristiano Crescentini, International School for Advanced Studies, Cognitive Neuroscience Sector, Via Bonomea, 265, 34136 Trieste, Italy. E-mail: crescent@sisssa.it.

DOI:10.1523/JNEUROSCI.4579-10.2011

Copyright © 2011 the authors 0270-6474/11/317763-12\$15.00/0

working memory demands, then this region should be activated consistently until successful rule acquisition. Moreover, we aim to test the possibility that the FPC should be active during rule acquisition but that the activation would continue during rule following until any uncertainty concerning the correctness of the rule is resolved.

Materials and Methods

Participants. Twenty-six right-handed healthy volunteers (nine females; 26.1 ± 4.9 years) participated in the study. All participants had no existing neurological or psychiatric illness. All participants gave written informed consent, and the study was approved by the independent Ethics Committee of the University of Trento. A total of six subjects were excluded from the behavioral and neuroimaging analysis: for one subject, technical problems in recording vocal responses precluded the analysis of his performance; for the other five excluded participants, the task performance was too poor (<50% of the rules were acquired, with only 23% of the rules being obtained on average). By contrast, all the final sample of 20 subjects, on which we based all the reported analyses, gave clear vocal responses and acquired >50% of the rules, obtaining on average 72.3% of them.

Stimuli and design. The Brixton Spatial Rule Attainment Task is a rule acquisition and rule following test. In the standard version (Burgess and Shallice, 1996), participants are presented one at time with a series of cards each containing a 2×5 display of circles numbered sequentially. One circle only is colored blue, the rest being presented in outline only. Subjects must predict which circle will be blue on the next card. The rules that have to be determined pertain to the relationship between successive stimuli. Typical examples of rules used in the test are the +1 rule (e.g., from 3 to 4, then to 5, and so on) and alternating between circles 5 and 10 (for additional examples, see Table 1). In the standard version, each rule is in operation from three to eight trials, and then it changes without warning.

In the present study, we adapted the standard version of the Brixton test to a functional magnetic resonance imaging (fMRI) context, the main difference being that each card contained a 2×6 display of circles (Figs. 1, 2A). The number of circles in each card was increased from 10 to 12 to increase the variability as well as the complexity of the rules. Moreover, in the current version, a card without any filled circles was presented between each pair of colored cards. This produced a brief rest period between the responses given by participants.

The experimental procedure involved 30 different rules, which ranged from very easy ones such as the +1 rule (see above) to more complex ones, such as adding 1, then adding 2, then adding 3, and so on. The rules were organized into six fMRI runs of five rules each; in each run, the rules were presented one after the other. The design was self-paced (for more details, see below, Procedures), so each run had a different duration. On average, a run lasted for 6.24 min (± 0.65 ; range, 5.26–9.05 min) and the total time of fMRI scanning was on average 37.4 min (± 3.1 ; range, 33.18–44.95 min).

In the Brixton test, Search, Discovery, and Following phases can generally be separated. In fact, the first response to a new rule is typically a guess; this may be followed by a series of incorrect responses, before the rule is correctly discovered by subjects; then the rule is followed until the next rule change. In the present study, a rule was judged as correctly acquired if and only if at least the final two responses were correct (a successful sequence). In other words, a sequence of correct responses could not be followed by an incorrect response on the last or next-to-last cards.

To assess whether a separation between Search, Discovery, and Following phases also applied to the rule set used in this study, we considered all rules on which the participants were correct on the last two responses. In a high proportion of them, namely 74%, the subjects were consistently wrong and then consistently right. In the remaining 26% of the rules, the subjects produced a correct response in the Search phase (i.e., a correct response followed by an incorrect one before the successful sequence). Most of these correct responses were guesses after a rule change. This

justifies separating Acquisition (Search and Discovery phases) from Following parts of the task.

Within the Search phase, the response given immediately after a rule change (SearchPhase1) in which the subject has no useful information clearly needs to be distinguished from the following responses (SearchPhase2). Moreover, with respect to the Following phase, the subjects may remain uncertain immediately after rule discovery that they have correctly attained the rule, but this uncertainty will fade over the correct trials that follow. For this reason, we separated the second correct response in the successful sequence (FollowingPhase1) from the following correct responses (FollowingPhase2). In sum, the key aspect of the study concerned the categorization of responses in correctly acquired rules into five different phases.

The five identified phases were as follows: Acquisition part: (1) SearchPhase1: the first response with a new rule; (2) SearchPhase2: all the responses after SearchPhase1 that precede the successful sequence of correct responses; (3) Discovery: the first correct response in the successful sequence; Following part: (4) FollowingPhase1: the second correct response in the successful sequence; (5) FollowingPhase2: the rest of the correct responses (i.e., from the third) in the successful sequence until the rule change.

This categorization of phases allowed us to contrast the Acquisition and the Following parts of the task, and subsequently to look more in detail for activation differences within the Acquisition part (SearchPhase1 vs SearchPhase2 vs Discovery) and within the Following part (FollowingPhase1 vs FollowingPhase2).

An important issue relates to the type of rules used. As mentioned above, the total set of 30 rules contained both easy and complex instances. The difficulty level of rules was evaluated according to minimum description length (MDL) (Solomonoff, 1964; Rissanen, 1978). The idea of MDL is based on Kolmogorov complexity (Li and Vitányi, 1997) that measures the amount of information needed to specify an object. For example, consider the following two strings of length 14: 1 2 1 2 1 2 1 2 1 2 1 2 and 5 8 1 3 2 5 5 6 1 2 2 8 3 6. The first string can simply be described in English as “1 2 seven times,” whereas the second string is generated randomly and the simplest description is the string itself. More formally, the Kolmogorov complexity of an object is the length of the shortest program in any computer programming language that can generate that object.

The application of Kolmogorov complexity to cognitive science, with the aim of measuring difficulty in cognitive tasks, requires that some assumption is made about the types of mental representations involved in the task. We assume that there are two kinds of possible representations that can be used to solve the Brixton test. Either or both arithmetic and geometric operations can be used. The arithmetic operations are addition and subtraction, whereas in a geometric operation the current blue circle would move to one of the five neighboring positions. Based on the kind of operations considered, different MDLs can be assigned to each rule. In general, the difficulty level of each rule was defined as the minimum of the arithmetic and geometric MDLs. As an example, Figure 1 represents two rules that we used in the task together with their MDLs.

The 30 rules we used were divided equally into easy and difficult ones. The MDLs varied from 1 to 3 for easy rules and from 3 to 6 for difficult ones. In the case of two MDLs of 3, the rule with the shorter mean arithmetic and geometric MDLs was considered as easier. Table 1 gives all the 30 rules used and, for each, the arithmetic and geometric descriptions, the MDL, the difficulty level, the period, and the average discovery trial. The period of a given rule denotes the number of cards before the rule starts repeating (excluding the first card). The average discovery trial specifies the trial at which rule acquisition occurs. This corresponds to the Discovery phase, namely when the first correct response occurs in a successful sequence of correct responses. It naturally follows that, on average, easy rules have a shorter period than difficult rules (1.66 vs 3 trials; $t_{(27)} = -5.53$; $p < 0.01$) and are acquired more quickly (after 3.59 vs 5.10 trials; $t_{(28)} = -4.21$; $p < 0.01$). As shown in the table, both types of rules are typically acquired after the completion of the corresponding period.

Distinguishing between easy and difficult rules was important since, in a subsidiary investigation, we examined whether there were brain regions

Table 1. The 30 rules used during fMRI

Rule ID	Rule	Example	Arithmetic description	Geometric description	ADL	GML	Per	MDL	ADT	DL
1	Same	4-4-4-4-4	Same	Same	1	1	1	1	2.72	E
2	Add 2	8-10-12-2-46	+2	→→	3	2	1	2	2.89	E
3	Subtract 2	8-6-4-2-12-10	-2	←←	3	2	1	2	2.72	E
4	+1, 0	6-7-7-8-8	+1 0	→ .	3	2	2	2	3.82	E
5	-1, 0	6-5-5-4-4-3-3	-1 0	← .	3	2	2	2	4.21	E
6	Top down	1-8-2-9-3-10	+7 -6	↓ ↑	8	2	2	2	2.71	E
7	Bottom up	11-4-10-3-9-2	-7 +6	↑ ↓	8	2	2	2	3.33	E
8	Diagonal	2-9-4-11-6-7	+7 -5	↘ ↗	8	2	2	2	3.57	E
9	Add 3	3-6-9-12-3	+3	→→→	3	3	1	3	2.94	E
10	Subtract 3	2-11-8-5-2	-3	←←←	3	3	1	3	3	E
11	+2, 0	6-8-8-10-10-12-12	+2 0	→→ .	4	3	2	3	3.89	E
12	-2, 0	8-6-6-4-4-2-2	-2 0	←← .	4	3	2	3	4.76	E
13	1; same	1-6-2-6-3-6-4-6	$x = \text{current pos, pos}$ $x + 1$	$x = \text{current pos, pos}$ $x \rightarrow$	3	3	2	3	5.37	E
14	+2, +1	9-11-12-2-3-5-6	+2 +1	→→ →	5	3	2	3	3.82	E
15	-2, -1	9-7-6-4-3-2-1	-2 -1	←← ←	5	3	2	3	4.14	E
16	+2, -1	5-7-6-8-7-9-8	+2 -1	→→ ←	5	3	2	3	4.66	D
17	-2, +1	4-2-3-1-2-12	-2 +1	←← →	5	3	2	3	3.91	D
18	Vertical	12-6-10-4-8-2-12	-6 +4	↑ ↓←	8	3	2	3	4.91	D
19	Parallel diagonal	5-12-4-11-3-10	+7 -8	↘ ↗	9	3	2	3	3.94	D
20	Left loop diagonal	6-5-12-5-4-11-4	-1 +7	← ↓	10	3	3	3	6.5	D
21	Right loop diagonal	1-2-9-2-3-10-3-4	+1 +7	→ ↓	10	3	3	3	5.90	D
22	1(2), 2	6-7-8-10-11-12-2	+1(2) +2	↗ →	6	4	3	4	6.16	D
23	Greek fret	11-5-4-10-9-3	+6 +1	↑ → ↓ →	6	4	4	4	3.93	D
24	Extremes	12-6-1-7-12-6-1-7	-6 -5	↑ → ↓ → ↓ →	8	4	4	4	3.89	D
25	Central square	9-3-4-10-9-3-4-10	-6 +1 -6 -1	↑ → ↓ ←	12	4	4	4	4.37	D
26	Series on two rows	1-6-2-7-3-8-4-9-5-10-6	+5 -4	← (only once) → (only once) ↓ ↗	8	4	3	4	4.78	D
27	1; diagonal	8-9-4-5-12-7	+1 -5 +1 +7	→ ↗ → ↓	12	4	4	4	6.69	D
28	+1, +2(2)	3-4-6-8-9-11-13	+1 +2 repeat	→ →→ repeat	6	4	3	4	5.92	D
29	3, 1(2)	7-10-11-12-3-4-5-8	+3 +1 repeat	→→→ → →	6	5	3	5	6.91	D
30	Progression	7-8-10-1-5-10-4	Load $x, + x$ $x = x + 1$	—	6	45	—	6	4.00	D

The numbers in the example column refer to the position of the blue circle in the 2 row × 6 column array. ADL, Arithmetic description length; GML, geometric description length; Per, period (number of cards, excluded the first, before rule starts repeating); MDL, minimum description length; ADT, average discovery trial (trial at which the rule is acquired); DL, difficulty level (easy/difficult). Despite equal ADL, GML, Per, and MDL, rules 16 and 17 were considered to be more difficult than rules 14 and 15 because they involve two different arithmetic operations.

modulated by rule difficulty in the Acquisition versus the Following part of the task. With respect to this, we first tested for the main effect of difficult minus easy rules, considering all rule phases together. We then applied an exclusive masking procedure, using the main effect as the masking contrast, to identify voxels in which any potential rule difficulty effect is not shared between the two contrasts of interest: (1) difficult minus easy rules computed for the rule acquisition part and (2) the corresponding contrast performed for the rule following part. The categorization of responses in different phases, as explained above for the overall set of rules, remained the same for easy and difficult rules. However, there was an exception: responses given after a rule change (i.e., Search-Phase1) were not analyzed separately for the two types of rules since all rules can be considered equivalent at this stage.

Procedures. Figure 2A shows a schematic representation of the experimental paradigm during fMRI. This comprised 30 rules organized in six runs. Each run consisted of five random rules balanced for difficulty and number of cards within runs. Each rule occurred only once in the experiment and consisted of a variable number of colored cards. A rule was in operation for a minimum of 6 colored cards and a maximum of 10 colored cards; easier rules consisted of fewer cards than more difficult rules (on average 6.8 cards vs 9.2 cards, respectively).

Each colored card was presented for a maximum of 7 s. Participants were instructed to use this time to think about the circle that will appear colored blue on the next card. Subjects were asked to press a button when ready to give a response. At the button press, the colored card was replaced by an empty card and subjects were required to say aloud the number where the next blue circle would appear as defined by the current rule. Subjects were told to consider the button press as a way to switch on a microphone. Vocal responses were collected for a maximum of 4 s after a button press. After this time and a jitter of 1–3 s, the next colored card was presented. Subjects were told that a rule could change after a certain number of colored cards and that they would then need to detect the change and infer the new rule as rapidly as they could. They were discouraged from trying to predict when the rule would change but to follow the learned rule as long as it was in operation. Fixation blocks of 15 s each were added both at the beginning and at the end of each run.

During fMRI, overt verbal responses were acquired using an MR-compatible fiber optic microphone that was attached to the head coil (Fibersound; model FOM1-MR); each participant's response was recorded as a .wav file. The responses were transcribed and checked for accuracy. For all subjects included in the analysis, all responses were recorded clearly.

Before fMRI scanning, participants practiced the tasks for ~7 min. Subjects were first explained the test and then presented with five rules using the same procedure to the one used in the scanner. The rules used during practice were different from those used during fMRI. Moreover, during practice, subjects were instructed on how to produce spoken responses while moving the face muscles as little as possible.

fMRI methods: acquisition and analyses. Images were acquired using a 4 T MRI scanner (Bruker Medical) using a birdcage transmit, eight-channel receive head coil (USA Instruments) and the point spread function distortion-corrected echoplanar imaging (Zaitsev et al., 2004). Head movement was minimized by mild restraint and cushioning. Thirty-seven slices of functional MR images were acquired using blood oxygenation level-dependent imaging (3×3 mm; 3 mm thick; repetition time, 2.2 s; time echo, 21 ms), covering the entire cortex. At the beginning of

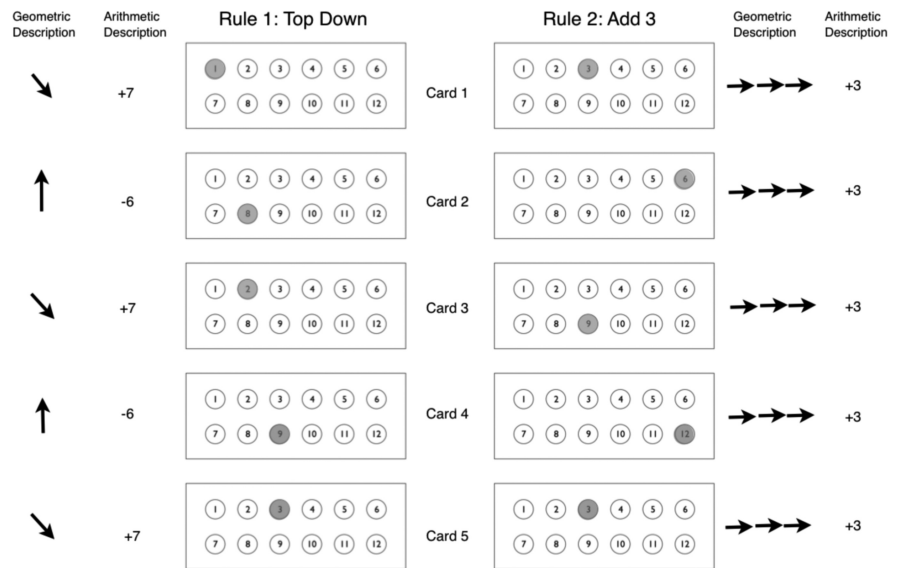


Figure 1. Examples of two rules in the Brixton test. Five cards are presented for each rule. Each rule can be described either in geometric or arithmetic terms. Geometrically, the first rule (rule 1: top down) (see also rule ID 6 in Table 1) can be described as “right down diagonal,” then “up.” If we consider moving to each neighbor of a position to be a geometric operation of length 1, then the geometric MDL is 2 for this rule. Arithmetically, this rule can be described as “+7,” then “–6.” To allow the arithmetic description length to consider smaller operands like +1 easier than larger operands like +7, the description length of operands is given by their length in the binary system. The description length of each arithmetic operator is also assumed to be 1. In the example, both “+7” and “–6” have an arithmetic length of $3 + 1 = 4$, and the total arithmetic MDL for the rule is $4 + 4 = 8$. Thus, the minimum of arithmetic and geometric MDLs is 2 for this rule. The second rule (rule 2: add 3) (see also rule ID 9 in Table 1) can be described geometrically as “right right right” and arithmetically as “+3,” which has MDL of 3 in both cases.

the scanning session, anatomical scans were also acquired for each participant, using a T1-weighted MP-RAGE.

The experimental task was presented using ASF (“A Simple Framework”) (available at <http://code.google.com/p/asf/>), based on Psychtoolbox for MATLAB (Brainard, 1997). SPM8 (Wellcome Department of Cognitive Neurology; see also <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>) was used for data preprocessing and statistical analyses. For all participants, we acquired on average 1020 volumes (170 volumes on average for each fMRI run); the first 4 volumes were discarded for each run to allow the magnetization to reach steady state. Slice acquisition delays were corrected using the middle slice as reference. All images were then corrected for head movement. All images were normalized to the standard SPM EPI template and spatially smoothed using an 8 mm FWHM Gaussian filter. The high-pass filter was set to the cutoff value of 128 s.

All subsequent analyses of the functional images were performed using the general linear model implemented in SPM8. First, for each subject, the data were fitted at every voxel using a combination of effects of interest. For correctly acquired rules, we modeled the onset of each colored card separately for all the five phases (i.e., SearchPhases1&2, Discovery, FollowingPhases1&2) and convolved with the hemodynamic response function (HRF). Each phase was modeled using a boxcar with a variable duration, so that the resultant β values represented an estimate of the neural response per unit time spent thinking about the future position of the blue circle. The duration of each boxcar function corresponded to the latency of the response (button press) generated by the subject in the corresponding phase.

On average, the block duration of the different phases was as follows: 2.98 s (SearchPhase1), 2.76 s (SearchPhase2), 2.60 s (Discovery3), 1.49 s (FollowingPhase1), and 1.26 s (FollowingPhase2) (see also Fig. 2B). Although there was no difference between the block duration for the first three phases (all $p > 0.09$) (see Results, Behavioral data), these blocks lasted significantly longer than the blocks in FollowingPhases1&2. Moreover, the block duration of FollowingPhase1 was longer than that of FollowingPhase2 ($p < 0.001$). Considering these block durations, for each subject and for each of the five phases we acquired on average 17.71

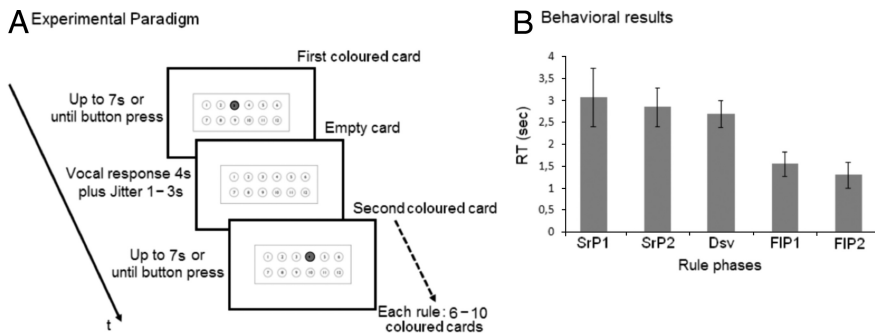


Figure 2. *A*, Experimental paradigm. The timing of the succession of the events is reported. Each given rule consists of a variable number of colored cards (in which a circle is colored blue) and empty cards in which all circles are presented white. When presented with colored cards, subjects are required to press a button when ready to predict which circle will be colored blue in the next colored card. Immediately after button press, colored cards are replaced with empty cards. During the presentation of empty cards, participants are asked to say aloud the number of the corresponding circle. The next colored card follows an empty card after a variable time (1–3 s). *B*, Behavioral results during fMRI. The graph reports mean RT of responses (button press) given in the five phases identified in correctly acquired rules. Phase connotation is as follows: (1) SearchPhase1 (SrP1) denotes the first response with a new rule, (2) SearchPhase2 (SrP2) identifies all the responses after SearchPhase1 that precede a successful sequence of correct responses, (3) Discovery (Dsv) corresponds to the first correct response in the successful sequence, (4) FollowingPhase1 (FIP1) denotes the second correct response in the successful sequence, and (5) FollowingPhase2 (FIP2) corresponds to the rest of correct responses (from the third) in the successful sequence until the rule change. Error bars indicate SD.

volumes for SearchPhase1, 62.75 for SearchPhase2, 24.98 for Discovery, 13.74 for FollowingPhase1, and 29.28 for FollowingPhase2.

In addition, the model comprised the onset of each fixation period (block duration, 15 s) and also, as two separated regressors, the onsets of each colored card presented for the rules that were not acquired [block duration dependent on mean reaction time (RT); mean duration equal to 2.53 s], plus the onset of the vocal responses (block duration, 3 s), convolved with the HRF. The latter two regressors were excluded from subsequent group-level analyses. Finally, the first-level analyses also included the parameters of the realignment (motion correction) as covariates of no interest.

Next, we obtained five contrast images per participant, corresponding to the five conditions of interest (the five phases) versus the baseline period (i.e., fixation) and pooling across the six runs. These five contrasts were then submitted to a 1×5 full factorial ANOVA, for group-level random effects statistical inference. Of importance, to test for the average effects for contrasts between rule acquisition and rule following that are unbalanced (i.e., SearchPhases1&2 and Discovery minus FollowingPhases1&2 and reversed), the contrast vectors were balanced by being defined as $[1/3 \ 1/3 \ 1/3 \ -1/2 \ -1/2]$ and $[-1/3 \ -1/3 \ -1/3 \ 1/2 \ 1/2]$, respectively.

Moreover, in a separate ANOVA, we tested for effects of rule difficulty. This was done by specifying a model similar to that described above with the only exception being that this additional model included easy and difficult correctly acquired rules, separately. Correction for nonsphericity (Friston et al., 2002) was used in both ANOVAs to account for possible differences in error variance across conditions and any nonindependent error terms for the repeated measures. Statistical threshold was set to $p_{\text{corr}} = 0.05$ corrected for multiple comparisons at the cluster level (cluster size estimated at $p_{\text{unc}} = 0.001$), considering the whole brain as the volume of interest. Furthermore, with respect to the ANOVA performed to assess rule difficulty effects, the SPM constituting the exclusive mask was thresholded at the default $p < 0.05$, whereas the contrasts to be masked were thresholded at $p < 0.001$ and then corrected for multiple comparisons at the cluster level ($p_{\text{corr}} = 0.05$).

Finally, the whole-brain analyses aimed to assess the effects of rule acquisition versus rule following were supplemented with region of interest (ROI) analyses. An ROI approach was used to confirm effects found in the whole-brain corrected data and, more critically, to test for theoretically relevant dissociations in activation between regions and conditions (i.e., the five rule phases) (Henson, 2005). Location of ROIs was unbiased with respect to the tests of interest since the coordinates were taken from two previous imaging studies of abstract rule learning

and rule use (Donohue et al., 2005; Badre et al., 2010). Given our hypotheses and the findings of whole-brain analysis, the first two ROIs were defined in left mid-DLPFC ($-50 \ 26 \ 24$) and left FPC ($-36 \ 50 \ 6$). These regions were derived from the study by Badre et al. (2010). These regions were also anatomically close to those reported in other functional imaging studies of reasoning [Hampshire et al. (2011), their Table 4]. A third ROI was centered in the superior parietal lobule (SPL) ($-27 \ -52 \ 57$) and was again based on the study by Badre et al. (2010). A fourth ROI was defined in the posterior middle temporal gyrus (postMTG) ($-56 \ -40 \ 2$); as no activation in this region was reported in the study by Badre et al. (2010), although this region has been associated with rule storage and rule use in past research (Bunge, 2004), the coordinates were derived from the study by Donohue et al. (2005), a key paper with respect to the role of the posterior temporal cortex in rule use. Since the whole-brain analysis did not produce lateralization effects, we also used the homologues regions in the right hemisphere [for the right SPL, the coordinates were $24 \ -56 \ 55$, and were derived from the study by Badre et al., (2010)].

The ROIs were constructed by creating a sphere with a radius of 6 mm around the centers defined by the sets of coordinates reported above. The mean ROI data for each event type (five phases plus fixation baseline blocks) were obtained from the design matrix of each subject. Using the Marsbar toolbox (Brett et al., 2002), the percentage signal change of all phases, including fixation baseline blocks, for each subject's ROI were computed for an event duration of 0 s and the absolute maximum function was used to calculate the signal change from each canonical event. The resulting data for all six runs were averaged over runs and subjects. All ROI data were used in repeated-measures ANOVAs.

Results

Behavioral data

Participants acquired on average 21.7 rules (± 4.19 ; range, 16–29). Subjects also acquired more easy than difficult rules (11.9 ± 2.17 vs 9.5 ± 2.56 ; $t_{(19)} = 4.90$; $p < 0.001$).

The RT data for each of the five phases identified in correctly acquired rules are shown in Figure 2*B*. A 1×5 repeated-measure ANOVA with the factor of rule phase (from SearchPhase1 to FollowingPhase2) revealed differences in RT between the different phases ($F_{(1,19)} = 40.53$; $p < 0.001$). Pairwise comparisons assessed the significance of the magnitude of the differences. These showed that rule acquisition, namely SearchPhases1&2 and Discovery, was associated with slower RT than rule following (FollowingPhases1&2). In fact, the responses in all three of the acquisition phases were slower than those produced during the two rule following phases (all $F > 32.91$; all $p < 0.001$). Of interest, although there were no differences between the three rule acquisition phases (all $F < 3.08$; all $p > 0.09$), the second given responses in a successful sequence of correct responses (i.e., FollowingPhase1) were slower than those produced subsequently (i.e., FollowingPhase2; $p < 0.001$).

Neuroimaging data

The main aim of the fMRI analyses was to investigate whether there exist regions preferentially active during rule acquisition and also others contributing more to rule following. Accordingly, we first report the results relative to the contrasts aimed to highlight brain regions involved in acquiring rules, and next we will

Table 2. Rule acquisition minus rule following

Anatomical localization	MNI coordinates				p_{corr}	Z value	Voxels per cluster
	~BA	x	y	z			
Main effect: regions more active for SearchPhases1&2 and Discovery than FollowingPhases1&2							
L middle frontal gyrus	6	-40	4	40	<0.001	6.20	2795
L middle frontal gyrus	9	-46	28	36		4.72	
L middle frontal gyrus	46	-44	22	26		4.55	
L superior frontal gyrus	6/8	-28	0	68		3.60	
L inferior frontal gyrus	45/46	-54	30	18		3.37	
R middle frontal gyrus	6	30	-2	54	<0.001	5.15	1923
R middle frontal gyrus	9	50	24	30		4.56	
R superior parietal lobe	7	34	-46	48	<0.001	4.21	789
R inferior parietal lobe	40	42	-36	48		3.90	
R precuneus	19	30	-66	42		3.74	
L superior parietal lobe	7	-30	-48	48	<0.02	3.77	405
L precuneus	7	-26	-68	36		3.33	
Conjunction: regions more active for SearchPhase1 minus FollowingPhases1&2 \cap SearchPhase2 minus FollowingPhases1&2 \cap Discovery minus FollowingPhases1&2							
L middle frontal gyrus	6	-40	2	42	<0.001	5.02	1267
L middle frontal gyrus	9	-48	26	30		3.80	
L middle frontal gyrus	8	-50	20	38		3.60	
R middle frontal gyrus	6	28	-4	56	<0.002	4.49	655
R middle frontal gyrus	9	54	22	30		3.75	

Stereotactic MNI coordinates for significant clusters (random effects, cluster-level $p < 0.05$, corrected, estimated at $p < 0.001$, uncorrected) given in millimeters with effect sizes (z scores) and cluster extent. In the voxels per cluster column, cluster extent is reported in correspondence of the main peak. Subpeaks were selected dividing each cluster into Brodmann areas (BA) and then selecting peaks within each area. L, Left; R, right.

turn to patterns of activation related to rule following. For rule acquisition, we first focus on the main effect of SearchPhases1&2 and Discovery minus FollowingPhases1&2 (Table 2, first contrast; Fig. 3). For rule following, we focus on the main effect of FollowingPhases1&2 minus SearchPhases1&2 and Discovery (Table 3, Fig. 4). Moreover, repeated-measures ANOVAs were performed on ROI data to test for significant interactions between region and condition (i.e., the five phases identified in correctly acquired rules).

As already mentioned, we also investigated whether rule difficulty modulated activity in any brain region, either during rule acquisition or rule following. Accordingly, we first report the main effect of difficult minus easy rules (Table 4, first contrast), and then we focus on the two corresponding simple main effects, which were computed separately for the acquisition phases and the following phases using the main effect as an exclusive mask (Table 4, second contrast).

Rule acquisition versus rule following

For rule acquisition minus rule following, we compared activation during SearchPhases1&2 and Discovery versus FollowingPhases1&2. This contrast revealed activation for rule acquisition > rule following in four clusters (Table 2, first contrast). The largest cluster was a swathe of cortex in the left hemisphere from the mid-DLPFC through the inferior frontal junction area (IFJ) (Brass et al., 2005). This latter region was very close to what is termed the dorsal anterior premotor cortex (pre-Pmd) in the studies by Badre and D'Esposito (2007) and Badre et al. (2010). The left superior frontal gyrus was

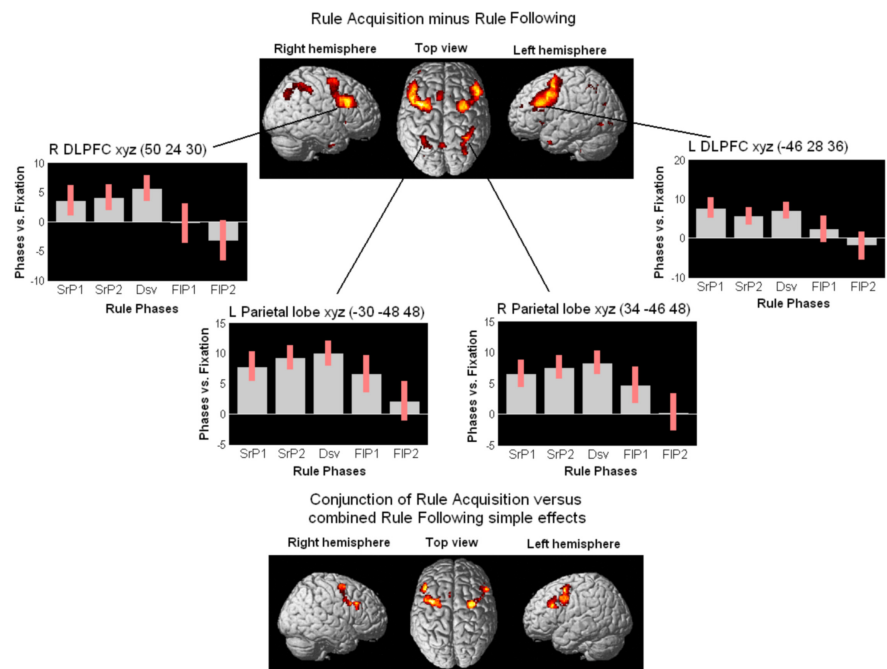


Figure 3. Effect of rule acquisition minus rule following (i.e., SearchPhases1&2 and Discovery minus FollowingPhases1&2). Brain activation in both hemispheres is reported in the top part of the figure. Signal plots for each rule phase are reported for both the right and left DLPFC and the right and left parietal lobe. Plots depict activity in experimental conditions relative to fixation [in arbitrary units (a.u.); $\pm 90\%$ confidence interval]; plots report the pattern of activity at the peaks of activation (i.e., single voxels) selected from the whole-brain contrast SPM maps. The bottom part of the figure reports brain activation in both hemispheres for the conjunction of rule acquisition versus combined rule following simple effects. The analysis was performed to test for any common effect of rule acquisition > rule following, regardless of the individual acquisition phases. The conjunction is given by the following: SearchPhase1 minus FollowingPhases1&2 \cap SearchPhase2 minus FollowingPhases1&2 \cap Discovery minus FollowingPhases1&2. SrP1, SrP2, Dsv, FIP1, and FIP2 refer to SearchPhase1, SearchPhase2, Discovery, FollowingPhase1, and FollowingPhase2, respectively.

also activated in this cluster. A similar pattern of activation in the corresponding regions of the right hemisphere was found in the second largest cluster; activation in this cluster extended slightly more posteriorly in the dorsal premotor cortex (Pmd) (Badre et al., 2010). Moreover, the superior parietal lobe and the precuneus were activated in both hemispheres as part of

Table 3. Rule following minus rule acquisition

Anatomical localization	MNI coordinates				p_{corr}	Z value	Voxels per cluster	
	~BA	x	y	z				
Main effect: regions more active for FollowingPhases1&2 than SearchPhases1&2 and Discovery								
L superior temporal gyrus	41	−56	−26	10	<0.001	6.54	11,770	
L superior temporal gyrus	22	−56	−12	6		6.27		
L precentral gyrus	4	−58	−4	18		6.20		
L precentral gyrus	6	−50	−14	32		5.66		
L postcentral gyrus	3	−24	−28	62		5.59		
R precentral gyrus	4	18	−24	66		5.42		
L parahippocampal gyrus	//	−26	−2	−14		5.39		
L cingulate gyrus	24	−4	0	44		5.34		
L insula	13	−34	0	12		5.17		
L putamen	//	−28	−14	2		4.86		
L inferior parietal lobe	40	−64	−26	26		4.37		
L postcentral gyrus	5	−20	−40	68		4.23		
R superior temporal gyrus	22	52	−2	0	<0.001	6.51	6187	
R superior temporal gyrus	21	64	−14	−2		6.36		
R superior temporal gyrus	41	50	−28	10		6.12		
R precentral gyrus	6	64	2	10		6.09		
R parahippocampal gyrus	//	26	−2	−14		4.70		
R putamen	//	28	−10	−2		4.23		
R inferior parietal lobe	40	66	−28	28		3.28		
R inferior frontal gyrus	47	58	22	0		3.16		
R postcentral gyrus	2	68	−22	32		3.14		
R cerebellum	//	18	−58	−22	<0.001	6.50		3637
L cerebellum	//	−18	−62	−22		4.80		
L cuneus	19	−10	−84	30		4.63		
R cuneus	19	8	−80	28		4.57		
R lingual gyrus	18	16	−62	−2		4.30		
L posterior cingulate	31	−2	−68	16		3.35		
R posterior cingulate	31	2	−70	14		3.28		
R medial frontal gyrus	10	10	56	8	<0.001	5.21	1130	
L medial frontal gyrus	10	−10	52	10		4.66		
R anterior cingulate	32	6	38	−4		3.89		
L anterior cingulate	32	−6	38	0		3.64		

Stereotactic MNI coordinates for significant clusters (random effects, cluster-level $p < 0.05$, corrected, estimated at $p < 0.001$, uncorrected) given in millimeters with effect sizes (z scores) and cluster extent. In the voxels per cluster column, cluster extent is reported in correspondence of the main peak. Subpeaks were selected dividing each cluster into Brodmann areas (BA) and then selecting peaks within each area.

two separate clusters. The activation also extended to the inferior parietal lobe in the right hemisphere. Of interest, the FPC did not show an effect of rule acquisition minus rule following. Figure 3 shows the areas that were activated for the main effect, plus the signal plots for the left and right mid-DLPFC and parietal cortex, including all five phases. The plots demonstrate an effect of rule acquisition minus rule following (SrP1, SrP2, and Dsv minus FIP1 and FIP2 in the plots) in all four regions. Nevertheless, whereas the mid-DLPFC appeared to be consistently more activated in both hemispheres during SearchPhases1&2 and Discovery than FollowingPhases1&2, the decrease in activation observed in FollowingPhase1&2 was more gradual for the parietal lobe. Consistent with this, Discovery and FollowingPhase1 trials differed in terms of mid-DLPFC activation (xyz , $-42\ 22\ 24$, $z = 4.43$, $p < 0.001$; xyz , $46\ 24\ 26$, $z = 4.53$, $p < 0.001$) but not in terms of parietal activation.

Interestingly, the left mid-DLPFC (xyz , $-52\ 20\ 28$; $z = 4.00$; $p < 0.001$) and the left IFJ (xyz , $-40\ 2\ 42$; $z = 5.77$; $p < 0.001$) were the only active regions, in a single cluster of 1197 voxels extent, when we tested for the same effect of rule acquisition > rule following, in an ANOVA containing RT as a covariate for each trial. This indicates a modulation of these regions over-and-above the effect of RT in our study. This cluster even includes (xyz , $-52\ 12\ 32$) the peak of the region found by Goel and Dolan (2000) showing an effect of rule application > perceptual baseline.

Moreover, because we expected a similar effect of rule acquisition > rule following in the DLPFC regardless of the different phases, we formally assessed this prediction by performing a conjunction analysis (Nichols et al., 2005) of SearchPhase1 minus FollowingPhases1&2 \cap SearchPhase2 minus FollowingPhases1&2 \cap Discovery minus FollowingPhases1&2. The bottom part of Figure 3 shows that the only regions that were significantly and consistently more active during each phase of rule acquisition than during rule following were the mid-DLPFC and a more posterior region that includes the IFJ/pre-PMd area. Both of these regions were activated bilaterally. The conjunction analysis confirmed the pattern of activation observed in the signal plots relative to the right and left mid-DLPFC. Consistent with this, no activation was found in the mid-DLPFC (or in the parietal lobe, the IFJ/pre-PMd area, and the FPC) when the three rule acquisition phases were directly contrasted against each other.

Summarizing, the imaging results for the contrasts involving rule acquisition minus rule following phases highlighted the role of a network including frontal and parietal brain regions in rule acquisition. More specifically, the mid-DLPFC and IFJ/pre-PMd regions were consistently activated in both hemispheres in all phases of rule acquisition and rapidly declined during rule following phases. Unlike the mid-DLPFC, the parietal lobes appeared to remain active during FollowingPhase1 trials. Finally, the FPC was not part of the regions showing a main effect of rule acquisition minus rule following.

Rule following versus rule acquisition

In our analyses, we also investigated whether there were regions contributing more to rule following than rule acquisition. With respect to this, we tested for the main effect of rule following minus rule acquisition, namely FollowingPhases1&2 minus SearchPhases1&2 and Discovery. This contrast revealed activation in four large clusters (Table 3). The largest had its peak of activation in the left superior temporal gyrus. Activation in this cluster also extended to the left precentral and post-central gyri as well as to the cingulate gyrus and to the most lateral part of the inferior parietal lobe. Activation in sub-cortical regions such as the left putamen was also found in this cluster. Moreover, a second large cluster involved very similar regions in the right hemisphere. A third cluster was centered in the cerebellum and extended to other posterior brain regions such as the cuneus and the lingual gyrus. A fourth cluster involved the anterior, ventral part of the medial frontal gyrus and extended to the anterior cingulate cortex. Figure 4 shows the areas that activated for the main effect of rule following minus rule acquisition, plus the signal plots for the right and left superior temporal gyri and for the medial frontal gyrus including all five phases. The plots demonstrate that there are consistently greater activations in these regions during FollowingPhases1&2 than the rule acquisition phases. Of interest, the plot concerning the medial frontal gyrus shows that SearchPhases1&2 and Discovery activated this region less than fixation. Remarkably, we also found that, during Discovery trials, there was a positive correlation between the mean activation in the cluster involving the medial PFC (computed using the Marsbar toolbox) and RT in those trials ($r = +0.49$; $p < 0.03$). This correlation indicated that the less active the areas involved in this cluster were during the Discovery phase, the faster participants responded in this phase.

The main effect of rule following minus rule acquisition was also tested in an ANOVA involving RT as a covariate for each trial. In a similar fashion to the basic ANOVA results just described, this main effect revealed bilateral activation in both the precentral gyrus and the superior temporal gyrus (peak activation in the right hemisphere cluster: xyz 58 -8 2; $z = 6.80$; $p < 0.001$; cluster extent, 2726 voxels; peak activation in the left hemisphere cluster: xyz -56 -28 10; $z = 6.77$; $p < 0.001$; cluster extent, 3728 voxels). Finally, we also assessed whether there were differences between the two rule following phases. The only region reporting a significant effect was the left FPC (-46 48 0, $Z = 4.80$, $p_{FWE-corr} = 0.021$), which was more active for FollowingPhase1 than FollowingPhase2 trials.

Analysis of ROI data

An ROI approach was adopted to further characterize the activation profiles observed in prefrontal, temporal, and parietal regions during rule acquisition and rule following phases. As already mentioned, we defined ROIs in mid-DLPFC (-50 26 24; 50 26 24), FPC (-36 50 6; 36 50 6), SPL (-27 -52 57; 24 -56 55), and postMTG (-56 -40 2; 56 -40 2). Critically, this ROI approach allowed one to test for dissociations between these regions (i.e., region by condition interaction; the factor condition being given by the five phases identified in our test). First, it should be

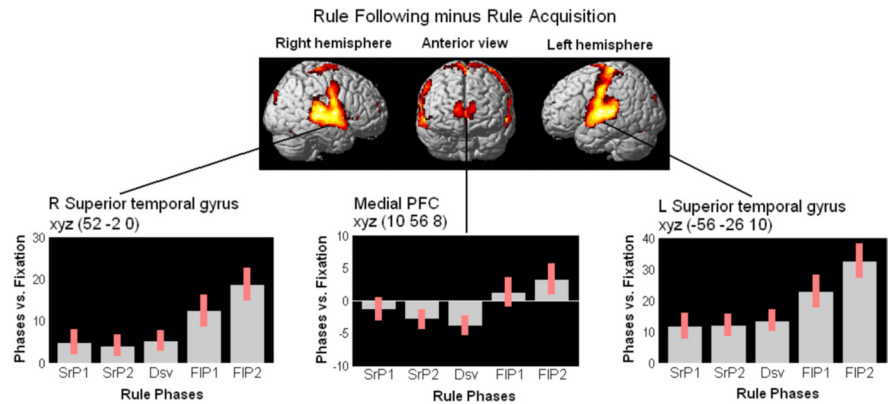


Figure 4. Effect of rule following minus rule acquisition (i.e., FollowingPhases1&2 minus SearchPhases1&2 and Discovery). Brain activation in both hemispheres is reported in the top part of the figure. Signal plots for each rule phase are reported for both the right and left superior temporal gyrus and the medial prefrontal cortex. Plots depict activity in experimental conditions relative to fixation [in arbitrary units (a.u.); $\pm 90\%$ confidence interval]; plots report the pattern of activity at the peaks of activation (i.e., single voxels) selected from the whole-brain contrast SPM maps. SrP1, SrP2, Dsv, FIP1, and FIP2 refer to SearchPhase1, SearchPhase2, Discovery, FollowingPhase1, and FollowingPhase2, respectively.

noted that all ROIs, except the right mid-DLPFC one, showed a significant effect of all phases (i.e., from SearchPhase1 to FollowingPhase2) minus fixation (all $p < 0.05$; $p = 0.18$ for the right mid-DLPFC ROI). This was consistent with the whole-brain data, which showed the same effect in voxels within 6 mm of the center of each ROI, this time also in the right mid-DLPFC ROI.

The ROI data were first submitted to a 2 (hemisphere: left and right) \times 4 (region: mid-DLPFC, FPC, SPL, and postMTG) \times 5 (condition: from SearchPhase1 to FollowingPhase2) repeated-measures ANOVA. The analysis showed the significant main effect of region ($F_{(3,57)} = 10.20$; $p < 0.001$) but not the effects of hemisphere ($F_{(1,19)} = 2.96$; $p = 0.101$) or condition ($F_{(4,76)} = 2.15$; $p = 0.082$). *Post hoc* tests (Bonferroni correction applied) performed for the region factor showed that the postMTG was more active than the other three regions (all $p < 0.03$) with no difference among the three (all $p > 0.11$). The analysis also showed the two-factor interactions between hemisphere and region ($F_{(3,57)} = 7.46$; $p < 0.001$) and hemisphere and condition ($F_{(4,76)} = 4.50$; $p < 0.004$). *Post hoc* tests (Bonferroni correction applied) executed for the first of the two interactions showed that the mid-DLPFC was more active in the left than the right hemisphere ($p < 0.018$) with no difference between the two hemispheres being present for the other three regions (all $p > 0.28$). *Post hoc* tests (Bonferroni correction applied) performed for the hemisphere by condition interaction revealed a difference between the two hemispheres (left $>$ right) for the discovery phase ($p < 0.015$) but not for all other phases (all $p > 0.17$). Critically, the ANOVA also showed a significant interaction between region and condition ($F_{(12,228)} = 11.04$; $p < 0.001$). This interaction was not modulated by the hemisphere factor (i.e., hemisphere by region by condition interaction: $F_{(12,228)} = 1.02$, $p = 0.43$). In view of this latter result, we averaged the signal across the right and left hemispheres in each ROI and used this averaged signal in all subsequent analyses concerning the mid-DLPFC, FPC, SPL, and postMTG ROIs. The averaged signal change for the four resulting ROIs is shown in Figure 5 for each phase identified in our test and for the fixation baseline blocks.

To examine where the original region by condition interaction arose, we performed six 2 (region) \times 5 (condition) ANOVAs using a Bonferroni correction criterion of $p < 0.0083$. In four of the ANOVAs, the region by condition interaction was significant, including all the interactions involving the postMTG (postMTG

Table 4. Effect of rule difficulty

Anatomical localization	~BA	MNI coordinates			p_{corr}	Z value	Voxels per cluster	
		x	y	z				
Main effect: regions more active for difficult rules than easy rules								
R superior temporal gyrus	22	68	−22	2	<0.001	6.99	7624	
R precentral gyrus	6	64	6	10		6.57		
R superior temporal gyrus	21	62	−16	−2		6.48		
R precentral gyrus	4	54	−4	46		5.83		
R precentral gyrus	44	60	14	4		5.60		
R inferior frontal gyrus	45	48	28	4		4.56		
R superior temporal gyrus	42	68	−30	14		4.23		
R inferior temporal gyrus	37	62	−54	−4		3.96		
R putamen	//	30	−12	−6		3.95		
R inferior frontal gyrus	47	44	36	−2		3.67		
L precentral gyrus	4	−44	−14	36	<0.001	7.00		12,422
L superior temporal gyrus	22	−48	−22	4		6.79		
L superior temporal gyrus	42	−60	−26	10		6.53		
L precentral gyrus	6	−40	−18	66		5.64		
L medial frontal gyrus	32	−8	10	42		5.28		
L postcentral gyrus	3	−22	−26	68		5.05		
L insula	13	−40	4	−2		4.90		
R medial frontal gyrus	32	8	10	44		4.82		
R cingulate gyrus	24	−4	0	46		4.74		
L putamen	//	−26	−10	4		4.56		
L inferior parietal lobe	40	−50	−32	58		4.52		
R cerebellum	//	16	−58	−24	<0.001	5.77	1663	
L cerebellum	//	−46	−62	−32		4.48		
R inferior parietal lobe	40	52	−46	50	<0.03	4.37		
L fusiform gyrus	18	−22	−86	−18	<0.04	3.63	220	
Regions more active for difficult rules than easy rules (in rule acquisition phases)								
masked by the main effect of difficult rules minus easy rules								
R superior parietal lobe	7	20	−64	54	<0.001	4.39	1199	
L superior parietal lobe	7	−18	−64	54		3.93		
R precuneus	7	30	−72	50		3.92		
R inferior parietal lobe	40	36	−50	50		3.77		
L precuneus	7	−26	−70	52		3.52		
L inferior parietal lobe	40	−40	−56	46		3.46		
L superior frontal gyrus	6	−22	4	68	<0.003	4.37		441
L middle frontal gyrus	9	−46	26	34	<0.001	3.98		541
L middle frontal gyrus	6	−48	8	44		3.92		

Stereotactic MNI coordinates for significant clusters (random effects, cluster-level $p < 0.05$ corrected, estimated at $p < 0.001$ uncorrected) given in millimeter with effect sizes (z scores) and cluster extent. In the Voxels per cluster column, cluster extent is reported in correspondence of the main peak. Subpeaks were selected dividing each cluster into Brodmann areas (BA) and then selecting peaks within each area.

vs mid-DLPFC: $F_{(4,76)} = 26.91$, $p < 0.001$; postMTG vs SPL: $F_{(4,76)} = 16.08$, $p < 0.001$; and postMTG vs FPC: $F_{(4,76)} = 7.01$, $p < 0.001$). Critically, however, the region by condition interaction was also significant for the mid-DLPFC versus FPC comparison ($F_{(4,76)} = 10.16$; $p < 0.001$). There was also a trend for a significant region by condition interaction between mid-DLPFC and SPL ($F_{(4,76)} = 2.96$; $p = 0.025$), but not for the FPC versus SPL comparison ($F_{(4,76)} = 1.82$; $p = 0.13$). Returning to the critical region by condition interaction for the mid-DLPFC versus FPC comparison, *post hoc* tests (Bonferroni correction applied) showed that the two regions differed during rule following trials ($p < 0.01$ for both FollowingPhase1 and FollowingPhase2 trials) but not during rule acquisition trials (all $p > 0.6$). *Post hoc* tests also showed a trend of FollowingPhase1 > FollowingPhase2 in the FPC ($p = 0.07$); this was in line with the same effect found in a similar region of the left hemisphere in the whole-brain corrected data. Finally, *post hoc* tests (Bonferroni correction applied) performed for each of the three two-factor interactions involving the postMTG showed that this region was generally more active than the other three regions specifically during rule following phases, particularly during FollowingPhase2 trials.

Finally, it should be noted that an analogous pattern of results for the two-factor region by condition interactions re-

ported above was obtained when only the ROIs in the left hemisphere were considered in the analyses.

Analysis of the effect of rule difficulty

Finally, in the whole-brain data, we investigated whether there were brain regions modulated by rule difficulty. We first tested for the main effect of difficult rules minus easy rules, considering all the four phases (from SearchPhase2 to FollowingPhase2) for which the two types of rule were analyzed separately. This contrast revealed activation in several clusters (Table 4, first contrast). The largest one had its peak of activation in the left precentral gyrus and extended to the left superior temporal gyrus as well as to regions in the superior portion of the medial frontal gyrus. Similar regions were also activated in the right hemisphere as part of a second large cluster. Activation in this cluster extended to the right inferior frontal gyrus. Moreover, both the cerebellum and the lateral part of the inferior parietal lobe were also more active (bilaterally) for difficult than easy rules.

Of interest, some of the regions globally modulated by rule difficulty were similar to the regions contributing more to rule following than rule acquisition (i.e., superior temporal gyrus, cingulate gyrus, cerebellum, putamen, and regions in the posterior/lateral part of the precentral gyrus) (see above, Rule follow-

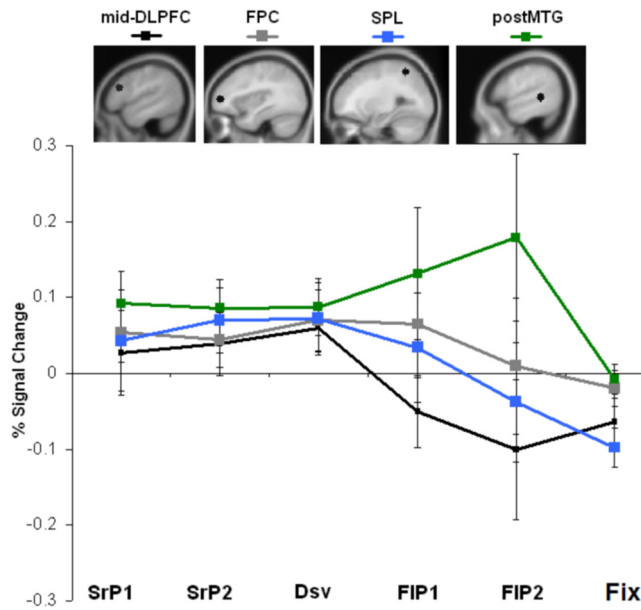


Figure 5. Activity in mid-DLPFC, FPC, SPL, and postMTG ROIs for each phase identified in the test and for fixation baseline blocks (Fix). Activity is measured by percentage signal change. The vertical bars denote SD of the means. The location of each ROI is reported in top part of the figure (the left hemisphere is shown). The signal was averaged across the right and left hemispheres in each ROI. SrP1, SrP2, Dsv, FIP1, and FIP2 refer to SearchPhase1, SearchPhase2, Discovery, FollowingPhase1, and FollowingPhase2, respectively.

ing versus rule acquisition) (Tables 3, 4). It is important to note, however, that an effect of rule following minus rule acquisition was found in these regions even when difficult and easy rules were considered separately (contrasts not shown). In contrast, regions in the middle frontal gyrus (bilaterally) and in the left inferior frontal gyrus, which showed increased activation for rule acquisition relative to rule following in the basic analysis (Table 2), were not overall more active for difficult than easy rules. Nonetheless, the regions in the parietal cortex, especially in the right hemisphere, which showed an effect of rule acquisition > rule following, were reasonably close to the right inferior parietal lobe region showing a general effect of rule difficulty.

Next, we tested for the two simple main effects of difficult minus easy rules separately for the rule acquisition phases (i.e., SearchPhase2 and Discovery) and the rule following phases (i.e., FollowingPhase1 and FollowingPhase2). We used the main effect contrast just described as an exclusive mask for both the two contrasts. The first contrast concerned rule acquisition and showed activation in three clusters (Table 4, second contrast). One was centered in the right superior parietal lobe and involved the same region in the left hemisphere as well as the inferior parietal lobe bilaterally. A second cluster involved the left superior frontal gyrus, whereas the third was centered in the left (but not the right) middle frontal gyrus and involved the mid-DLPFC and the IFJ/pre-PMD regions. Of importance, these latter regions as well as the parietal ones, were very similar to those showing a main effect of rule acquisition minus rule following in the main analysis (see above, Rule acquisition versus rule following) (Tables 2, 4). The second simple main effect concerned the rule following phases and did not reveal any significant activation.

Finally, it should be noted that the FPC appeared not to be modulated by rule difficulty overall. However, this region is particular in not showing an effect of rule acquisition minus rule following in the main analysis. Thus, it is insensitive to the contrasts examined in the previous paragraph. In sum, the additional

analyses related to rule difficulty show a greater activation of temporal and motor cortices for difficult than easy rules. More generally, the data indicate that when a region is active, it is more active for difficult than easy rules. Thus, we also found increased activation of the left mid-DLPFC and of the left and right parietal cortex for difficult rules during the acquisition phases. However, this generalization does not apply to the right mid-DLPFC or the medial/anterior PFC, which were activated to a similar extent by easy and difficult rules during the rule acquisition and rule following phases, respectively (see also later in Discussion).

Discussion

In the present study, we investigated mechanisms of rule acquisition and rule following in inductive reasoning. We used the Brixton test (Burgess and Shallice, 1996) in which subjects attain and then follow 30 different rules. Five distinct phases could be clearly distinguished during performance on each successfully attained rule. Whereas the first three phases tapped rule acquisition, the remaining two involved rule following. Comparing between the Acquisition part and the Following part (SearchPhases1&2 and Discovery minus FollowingPhases1&2) revealed bilateral activation in frontal (mid-DLPFC and IFG/pre-PMD) and parietal regions. These regions, except for the right mid-DLPFC, were modulated by rule difficulty, being more active for difficult than easy rules during rule acquisition. The opposite comparison (Following minus Acquisition) produced bilateral activation in the temporal and precentral gyri, and in the medial/anterior PFC. Orthogonal ROI analyses confirmed this pattern of results, showing in addition a dissociation between FPC and mid-DLPFC. This was attributable to the prolonged activation of the FPC from rule acquisition trials to initial rule following trials (i.e., FollowingPhase1 trials). Moreover, the selective response of the postMTG during rule following dissociated this region from the mid-DLPFC, FPC, and SPL ones.

Activation of a network of frontoparietal regions in the Brixton test is consistent with results of previous fMRI studies on reasoning (Hampshire et al., 2011) and spatial working memory (Smith and Jonides, 1997; D'Esposito et al., 1998), as well as with the hypothesis that this network constitutes an adaptable system that is recruited in novel situations whenever the general level of difficulty and the need for executive control increases (Hampshire et al., 2008). Nevertheless, our results extend those of previous studies by showing that frontal, parietal, and temporal regions differed in their patterns of activation during the five phases in which we divided the test.

The pattern of activation found in the mid-DLPFC fits with a role for this region in rule acquisition rather than rule following. During the Acquisition part of the Brixton test, participants tend to generate at least one possible hypothesis for each card presented; on this view the mid-DLPFC appears to be critical for detecting regularities across stimuli and for using this information to generate appropriate hypotheses. The increased activation of this area (in the left hemisphere) more for difficult than easy rules indicates that abstracting an organizing rule is particularly demanding in difficult cases. A number of alternative hypotheses must often be generated before this type of rule can be correctly induced.

This generalization is in line with previous research (Reverberi et al., 2005a,b; Specht et al., 2009), but differs in emphasis from conclusions from other studies, including those using neurophysiological recordings in nonhuman primates. Thus, although some studies (Wang et al., 2000; Bussey et al., 2001) have shown that lesioning ventral PFC after learning does not affect subse-

quent rule use in monkeys, other studies, in which recording was from regions in the principal sulcus extending to DLPFC (BA 9/46), have supported the idea that this latter region controls the guidance of behavior according to well learned rules (White and Wise, 1999; Asaad et al., 2000; Mansouri et al., 2006). From this perspective, mid-DLPFC, as well as more posterior brain regions such as the IFJ and PMd, would mediate rule implementation by selecting and maintaining the appropriate response contingencies associated with the currently relevant task rules (Bunge, 2004; Brass et al., 2005; Bunge et al., 2005).

An alternative possibility, however, that cannot be ruled out is that the mid-DLPFC could have played a general monitoring role in our task. Monitoring would be required during the acquisition phases to overcome the tendency to repeat inappropriate rules/hypotheses (Reverberi et al., 2005a).

The pattern of activation that we found for the FPC indicates that, unlike the mid-DLPFC, this region contributes to inductive reasoning after hypothesis generation. We suggest that, in this task, FPC maintained alternative hypotheses in a pending state during rule acquisition, until the positive feedback received after the Discovery and FollowingPhase1 phases, unambiguously indicates the validity of one hypothesis. Take, for instance, the central square rule reported in Table 1 (rule 25); a typical situation for the first correct response is when the blue circle is predicted to be in position 10. On this card, the participant should consider at least two alternative rules producing number 9 (correct) and then number 3, in the first case, and numbers 11 and 5, in the second. The FPC would still be activated when the blue circle is in position 9 (FollowingPhase1 trial).

The present account for the FPC activation is in line with other proposals concerning the region. Thus, Burgess et al. (2000, 2001) argued that the FPC is critical for the setting up and realization of intentions. Boorman et al. (2009) recently strengthened this idea by showing that this region is important in representing a pending or alternative potential course of action to determine when to switch to that behavior. On this account, the parietal lobes would then be responsible for implementing the switches. It is striking that, in our study, the parietal lobes too remained relatively active during FollowingPhase1 trials.

Recently, there has been a growing consensus that PFC may be functionally organized to support progressively more abstract representations along a rostrocaudal axis from PMd to FPC cortex, passing through pre-PMd and mid-DLPFC (Badre and D'Esposito, 2007). Overall, our findings of sustained activation of the pre-PMd and mid-DLPFC during rule acquisition and the protracted activation of the FPC during FollowingPhase1 trials are in agreement with the proposed hierarchical organization of the PFC.

In this context, in a recent study, Badre et al. (2010) studied the neural mechanisms of abstract rule learning and again found a rostrocaudal pattern of activation in the PFC. However, unlike our study, in that by Badre et al., abstract rule learning was mediated by processing in the pre-PMd region and activity in the PMd (and in the mid-DLPFC and FPC regions) did not distinguish between the hierarchical (abstract) rule set and the flat rule set they used. It is possible that the difference between our study and that by Badre et al. in the pattern of results reflects differences in the type of rules and how they are learned, namely by inductive reasoning and reinforcement learning, respectively. From this perspective, rules in the Brixton test would be sufficiently abstract to require the involvement of the mid-DLPFC and the FPC cortex. On this account, the latter area would be at the apex of our ability to process abstract representations such as those involved

in the Brixton test when one needs to represent the outcomes of multiple separate inferences simultaneously (Ramnani and Owen, 2004; Bunge et al., 2009).

We now turn to the fMRI analysis that dealt with activity during rule following minus rule acquisition. Here, we found activation in regions such as temporal and motor cortex that appear as likely candidates for rule storage and motor planning/execution, respectively (Bunge, 2004; Donohue et al., 2005). The activation found in the medial/anterior PFC may reflect, however, default-mode network (DMN) activity (Raichle, 2010). A linear decrease in activation of the midline DMN regions has been observed with increasing task difficulty and working memory demands (Esposito et al., 2006; Sambataro et al., 2010). Our study supports the position that the deactivation observed in this region is task induced; there is a linear decrease in activation from SearchPhase1 up to Discovery, a period over which the number of instances and hypotheses grows; there is also a correlation between activation in this region and RT during Discovery trials. Similarly, the medial/anterior PFC is the one region (together with the right mid-DLPFC) that breaks the following rule: if a region is positively activated in a phase, then its activation is increased for the more difficult rules. This does not apply to the medial/anterior PFC in FollowingPhase1 and FollowingPhase2 trials. For other regions, greater relevant resource requirements (e.g., hypothesis generation, rule storage) lead to greater activations.

In conclusion, the present study provides evidence that the network of frontal, parietal, and temporal regions underlying inductive reasoning fractionates with different subregions responding to rule acquisition and rule following phases. Mid-DLPFC was consistently activated during rule acquisition; FPC contributed to rule acquisition but also to the initial phases of rule following. Parietal cortex showed a pattern similar to FPC. By contrast, motor and temporal cortex contributed particularly to rule following.

References

- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–459.
- Badre D, D'Esposito M (2007) Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J Cogn Neurosci* 19:2082–2099.
- Badre D, Kayser AS, D'Esposito M (2010) Frontal cortex and the discovery of abstract action rules. *Neuron* 66:315–326.
- Boettiger CA, D'Esposito M (2005) Frontal networks for learning and executing arbitrary stimulus-response associations. *J Neurosci* 25:2723–2732.
- Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62:733–743.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Brass M, Derrfuss J, Forstmann B, von Cramon DY (2005) The role of the inferior frontal junction area in cognitive control. *Trends Cogn Sci* 9:314–316.
- Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox. Paper presented at Eighth International Conference on Functional Mapping of the Human Brain, Sendai, Japan, June.
- Bunge SA (2004) How we use rules to select actions: a review of evidence from cognitive neuroscience. *Cogn Affect Behav Neurosci* 4:564–579.
- Bunge SA, Wallis JD (2008) *Neuroscience of rule-guided behavior*. Oxford: Oxford UP.
- Bunge SA, Wallis JD, Parker A, Brass M, Crone EA, Hoshi E, Sakai K (2005) Neural circuitry underlying rule use in humans and nonhuman primates. *J Neurosci* 25:10347–10350.
- Bunge SA, Helskog EH, Wendelken C (2009) Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46:338–342.
- Burgess PW, Shallice T (1996) Bizarre responses, rule detection and frontal lobe lesions. *Cortex* 32:241–259.

- Burgess PW, Veitch E, de Lacy Costello A, Shallice T (2000) The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia* 38:848–863.
- Burgess PW, Quayle A, Frith CD (2001) Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia* 39:545–555.
- Bussey TJ, Wise SP, Murray EA (2001) The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (*Macaca mulatta*). *Behav Neurosci* 115:971–982.
- D'Esposito M, Aguirre GK, Zarahn E, Ballard D, Shin RK, Lease J (1998) Functional MRI studies of spatial and nonspatial working memory. *Brain Res Cogn Brain Res* 7:1–13.
- Donohue SE, Wendelken C, Crone EA, Bunge SA (2005) Retrieving rules for behavior from long-term memory. *Neuroimage* 26:1140–1149.
- Esposito F, Bertolino A, Scarabino T, Latorre V, Blasi G, Popolizio T, Tedeschi G, Cirillo S, Goebel R, Di Salle F (2006) Independent component model of the default-mode brain function: assessing the impact of active thinking. *Brain Res Bull* 70:263–269.
- Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J (2002) Classical and Bayesian inference in neuroimaging: applications. *Neuroimage* 16:484–512.
- Goel V, Dolan RJ (2000) Anatomical segregation of component processes in an inductive inference task. *J Cogn Neurosci* 12:110–119.
- Hampshire A, Thompson R, Duncan J, Owen AM (2008) The target selective neural response—similarity, ambiguity, and learning effects. *PLoS One* 3:e2520.
- Hampshire A, Thompson R, Duncan J, Owen AM (2011) Lateral prefrontal cortex subregions make dissociable contributions during fluid reasoning. *Cereb Cortex* 21:1–10.
- Henson R (2005) What can functional neuroimaging tell the experimental psychologist? *Q J Exp Psychol A* 58:193–233.
- Koechlin E, Hyafil A (2007) Anterior prefrontal function and the limits of human decision-making. *Science* 318:594–598.
- Koechlin E, Basso G, Pietrini P, Panzer S, Grafman J (1999) The role of the anterior prefrontal cortex in human cognition. *Nature* 399:148–151.
- Li M, Vitányi P (1997) An introduction to Kolmogorov complexity and its applications. New York: Springer.
- Mansouri FA, Matsumoto K, Tanaka K (2006) Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *J Neurosci* 26:2745–2756.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Raichle ME (2010) Two views of brain function. *Trends Cogn Sci* 14:180–190.
- Ramnani N, Owen AM (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci* 5:184–194.
- Reverberi C, Lavaroni A, Gigli GL, Skrap M, Shallice T (2005a) Specific impairments of rule induction in different frontal lobe subgroups. *Neuropsychologia* 43:460–472.
- Reverberi C, D'Agostini S, Skrap M, Shallice T (2005b) Generation and recognition of abstract rules in different frontal lobe subgroups. *Neuropsychologia* 43:1924–1937.
- Rips LJ (1999) Deductive reasoning. In: *The MIT encyclopedia of the cognitive sciences* (Wilson RA, Keil FC, eds), pp 225–226. Cambridge, MA: MIT.
- Rissanen J (1978) Modeling by shortest data description. *Automatica* 14:465–471.
- Sambataro F, Murty VP, Callicott JH, Tan HY, Das S, Weinberger DR, Mattay VS (2010) Age-related alterations in default mode network: impact on working memory performance. *Neurobiol Aging* 31:839–852.
- Seger CA, Poldrack RA, Prabhakaran V, Zhao M, Glover GH, Gabrieli JD (2000) Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia* 38:1316–1324.
- Smith EE, Jonides J (1997) Working memory: a view from neuroimaging. *Cogn Psychol* 33:5–42.
- Solomonoff R (1964) A formal theory of inductive inference, part I. *Inform Control* 7:1–22.
- Specht K, Lie CH, Shah NJ, Fink GR (2009) Disentangling the prefrontal network for rule selection by means of a non-verbal variant of the Wisconsin Card Sorting Test. *Hum Brain Mapp* 30:1734–1743.
- Strange BA, Henson RN, Friston KJ, Dolan RJ (2001) Anterior prefrontal cortex mediates rule learning in humans. *Cereb Cortex* 11:1040–1046.
- Wang M, Zhang H, Li BM (2000) Deficit in conditional visuomotor learning by local infusion of bicuculline into the ventral prefrontal cortex in monkeys. *Eur J Neurosci* 12:3787–3796.
- White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126:315–335.
- Zaitsev M, Hennig J, Speck O (2004) Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. *Magn Reson Med* 52:1156–1166.