Anterior Prefrontal Cortex Mediates Rule Learning in Humans

B.A. Strange[1], R.N.A. Henson[1,2], K.J. Friston[1] and R.J. Dolan[1,3]

[1]Wellcome Department of Cognitive Neurology, Institute of Neurology, 12 Queen Square, London WC1N 3BG, [2]Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR and [3]Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3, UK

Despite a need for rule learning in everyday life, the brain regions involved in explicit rule induction remain undetermined. Here we use event-related functional magnetic resonance imaging to measure learning-dependent neuronal responses during an explicit categorization task. Subjects made category decisions, with feedback, to exemplar letter strings for which the rule governing category membership was periodically changed. Bilateral fronto-polar prefrontal cortices were selectively engaged following rule change. This activation pattern declined with improving task performance reflecting rule acquisition. The vocabulary of letters comprising the exemplars was also periodically changed, independently of rule changes. This exemplar change modulated activation in left anterior hippocampus. Our finding that fronto-polar cortex mediates rule learning supports a functional contribution of this region to generic reasoning and problem-solving behaviours.

Introduction

The psychological processes through which humans learn to categorize stimuli have been studied extensively (Smith *et al.*, 1998). Considerable interest surrounds the proposal that people abstract the rules that define category membership unconsciously, through simple exposure to exemplars of the categories (Reber, 1967). This proposal remains controversial however (Shanks, 1995). Firstly, much of the evidence that claims to demonstrate abstract rule learning can equally be explained in terms of categorization on the basis of superficial similarity, either between whole exemplars [instance-based categorization (Nosofsky, 1986)] or exemplar parts [fragment-based categorization (Perruchet and Pacteau, 1990)]. Secondly, the situations that provide the most robust evidence for abstract rule induction are those that involve explicit (conscious) hypothesis testing rather than passive stimulus exposure (Shanks and St John, 1994).

We have previously attempted to determine the neuroanatomical correlates of category learning by measuring haemodynamic responses during a modified artificial grammar (AG) learning paradigm (Fletcher *et al.*, 1999; Strange *et al.*, 1999). An AG is a set of rules governing the concatenation of symbols into strings. In our previous studies however, the extent to which learning was implicit or explicit, or based on similarity- or rule-based mechanisms, was unclear. Contrary to previous AG studies, learning was intentional rather than incidental, with the grammatical status of exemplars indicated with trial-by-trial feedback, which may have encouraged explicit rule induction. Nonetheless, the learning may also have involved implicit or explicit similarity-based comparisons, given that the exemplars were presented repeatedly and the vocabulary of the grammar (the symbols comprising the exemplars) was constant over the experiment.

The critical test of abstract rule-based learning is whether categorization performance transfers to exemplars drawn from a new vocabulary [for which similarity-based mechanisms cannot operate (Smith *et al.*, 1992)]. Though Fletcher *et al.* demonstrated some transfer of categorization performance from one set of exemplars to another, these exemplars were drawn from the same vocabulary, hence transfer could equally have been based on similarity-based processes (Fletcher *et al.*, 1999).

In the present study we address the neural correlates of explicit abstract rule induction. Subjects were required to categorize letter strings as 'grammatical' or 'ungrammatical' according to a currently relevant rule, with trial-by-trial feedback. The rule, which was based on the position of a repeated letter in four-letter strings, was simple enough for people to learn over the course of 20 trials (see Fig. 1). The rule was changed periodically to enable detection of neuroanatomical regions transiently engaged by rule induction. Furthermore, the letters that comprised exemplars (the vocabulary) were also changed periodically (independently of rule changes). This enabled us to determine whether performance transferred across exemplar changes, and so establish whether subjects had successfully abstracted the rules.

To measure rule learning-dependent responses, we used event-related functional magnetic resonance imaging (fMRI) to test for responses, to correct trials alone, that correlated with each subject's performance over time. Thus, the predicted rate of adaptation of neuronal responses was tailored to individual learning rates, but was independent of trial-specific feedback. We also tested for a more general response adaptation, independent of subjects' performance, associated specifically with adaptation following exemplar changes. On the basis of previous neuroimaging (Berman *et al.*, 1995; Nagahama *et al.*, 1996; Goldberg *et al.*, 1998; Fletcher *et al.*, 1999; Rogers *et al.*, 2000) and human lesion studies (Milner, 1963; Stuss *et al.*, 2000), we hypothesized that rule learning would be frontally mediated. By contrast, on the basis of our previous data, we predicted that exemplar change would engage the hippocampus, consistent with our proposal of an automatic response to perceptual and exemplar novelty in this region (Strange *et al.*, 1999).

Materials and Methods

Subjects

Informed consent was obtained from 10 right-handed subjects (six male, four female; age range 22–37 years; mean age 27.4; recruited by advertisement). Data from two subjects (one male, one female) were excluded from the analysis due to poor task performance. Ethics approval was obtained from the National Hospital for Neurology and Neurosurgery Joints Ethics Committee.

Psychological Task

During scanning, subjects were presented visually with strings of four letters in upper case, at a rate of one every 4 s. Subjects were required to make a push-button response with the right hand to indicate whether each string was correct or incorrect according to a pre-specified abstract

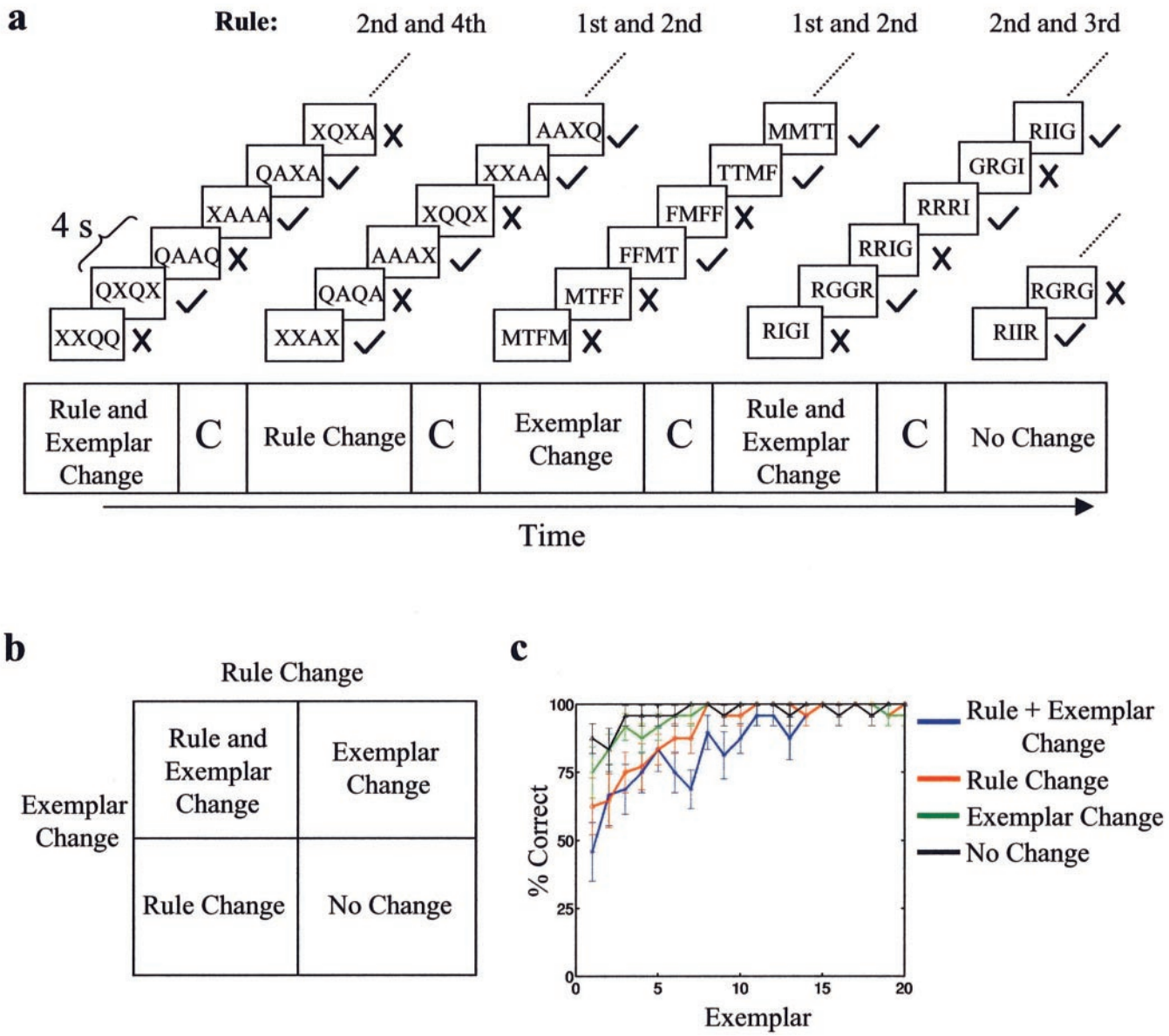

Figure 1. Experimental design and behavioural performance. (*a*) Experiment time line showing 5 of the 12 activation epochs, each followed by a control epoch (C). For each activation epoch, sample stimuli are shown (of the 20 that were presented) along with a tick or cross indicating whether the string conforms to or violates the current rule. This rule is stated above the relevant sample strings and refers to the presence of a repeated letter in the first to fourth position of each string. (*b*) The 2 × 2 factorial design. (*c*) The average performance of the eight subjects for each of the four conditions is plotted (± SE) for the 20 exemplars presented during each activation epoch. Here, and in all subsequent figures, the response following both rule and exemplar change (RC+EC) is shown in blue; rule change (RC) in red; exemplar change (EC) in green; and no change (No) in black.

rule. Prior to scanning, subjects were instructed that rules were based on repeated letters within the string. Subjects were told that a possible rule was 'If the first and last letter are the same, the item is correct. For example, XBFX and BFXB would both be correct, but XFXB would be incorrect'. Twenty strings were presented across individual activation epochs, with no string being presented more than once. Trial-by-trial feedback, indicating whether subjects' responses were right or wrong, was provided to enable subjects to induct the rule over trials. The strings presented in the next activation epoch were constrained by a 2 × 2 factorial design, with rule change as one factor and letter set (exemplar) change as another factor (see Fig. 1). Thus, both the rule and the letters making up the exemplars changed (RC+EC), or the rule changed and the exemplars stayed the same (RC), or the exemplars changed and the rule stayed the same (EC) or both the rule and exemplars were the same as in the previous activation epoch (No). The two subjects that were excluded from the analysis performed poorly in the no change condition. The order of conditions was random and each condition was repeated three times. Each activation epoch was followed by a control epoch during which the strings LLLL or RRRR were presented (five of each), requiring a left (index finger) or right (middle finger) key press respectively. Prior to scanning, subjects were trained on 10 stimuli of each of the four cells in the 2 × 2 factorial design. Note that our rule-learning task is distinct from standard artificial grammar learning paradigms (Reber, 1967), as the latter do not provide feedback and are based on complex rules that subjects may (or may not) abstract during passive exemplar exposure.

Data Acquisition

A Siemens VISION system (Siemens, Erlangen, Germany), operating at 2 T, was used to acquire both T_1-weighted anatomical images and gradient-echo echo-planar T_2*-weighted MRI image volumes with blood oxygenation level dependent (BOLD) contrast. A total of 480 volumes were acquired per subject plus five 'dummy' volumes, subsequently

discarded, to allow for T_1 equilibration effects. Volumes were acquired continuously every 3000 ms. Each volume comprised thirty 3 mm axial slices, with an in-plane resolution of 3 × 3 mm, positioned to cover the whole cerebrum. The imaging time series was realigned to correct for interscan movement and normalized into a standard anatomical space (Talairach and Tournoux, 1988) to allow group analyses. The data were then smoothed with a Gaussian kernel of 8 mm full-width half-maximum to account for residual intersubject differences (Friston *et al*., 1995).

Data Analysis

Data were analysed using Statistical Parametric Mapping (SPM99) employing an event-related model (Josephs *et al*., 1997). The data were first filtered to remove low frequency drifts in signal (cut-off 174 s). In the analysis testing for the effects of rule change, we specified four distinct effects of interest: the event train following change in rule and exemplar (RC+EC), change in rule alone (RC), change in exemplar alone (EC) and no change in rule or exemplar (No). The presentation of each letter string was modelled by convolving a delta function at each event onset with a canonical haemodynamic response. Correct and incorrect responses were modelled separately. To measure rule learning-dependent activation, performance of the *i*th subject was averaged across the four conditions and fitted by the exponential function $1 - \exp(-k_i t)$ using nonlinear techniques implemented in Matlab (The Mathworks, Inc., Natick, MA). The function $\exp(-k_i t)$ was then used to modulate the event train in each activation epoch for both correct and incorrect responses (given that learning-related activation would be inversely related to performance).

In summary, for each subject, four effects were modelled for each of the four conditions: separate regressors for correct and incorrect responses plus a regressor modelling modulation of both by the exponential decay function. The regressors modelling event-related responses that were constant throughout each 80 s activation epoch (epoch responses) embody mean changes in brain activity, following change in either rule or exemplar. The regressors modelling the exponential decay embody subject-specific learning-dependent responses within an epoch. Only contrasts involving correct responses were used in formal statistical analyses (there were too few incorrect responses for these regressors to be tested). Movement parameters, determined during realignment, were entered as covariates of no interest, to remove possible movement-related residual effects.

Subject-specific parameter estimates pertaining to each regressor were calculated for each voxel. Contrasts, confined to the adaptation effects, for the main effect of rule change were specified over subjects and tested with the *t* statistic (i.e. a fixed effects model across subjects). We report all rule learning-related effects at a height threshold of $P < 0.0001$ (uncorrected) and a spatial extent threshold of 5 voxels.

A similar analysis was conducted to test for response adaptation following exemplar change. The purpose of this second analysis was to focus on a region of interest, the anterior extent of the hippocampus (used here to refer to the dentate gyrus, CA subfields and subiculum), which we have previously implicated in detecting novel stimuli that are both task relevant and irrelevant (Strange *et al*., 1999). This previous result suggested an automatic anterior hippocampal response to exemplar novelty, which would not necessarily be correlated with subject-specific behaviour (the behavioural effects of exemplar change in the current paradigm were, in any case, not significant; see Results). Hence, in this analysis, instead of modelling a subject-specific performance-related exponential decay, we chose an arbitrary exponential function to model adaptation to exemplar novelty. The same function modelled novelty-dependent responses in all subjects. For this analysis, the whole-brain SPM was thresholded at $P < 0.05$ (uncorrected) and the anterior hippocampal region previously engaged by perceptual and exemplar novelty (Strange *et al*., 1999) was examined for evidence of exemplar change-induced activation. The uncorrected threshold of $P < 0.05$ was adopted given the strong prediction that exemplar change would engage anterior hippocampus.

Results

Behaviour

Figure 1*c* demonstrates that performance fell following a rule change, as subjects were initially forced to guess the rule, but then improved over trials, reaching 100% by the end of each rule change epoch. As predicted, a repeated measures 2 × 2 × 20 ANOVA demonstrated a significant Rule change [(RC+EC + RC) – (EC + No)] × Time interaction [$F(3.7,26.2) = 2.588$, $P < 0.05$; one-tailed; Greenhouse–Geisser corrected for non-sphericity of time effects]. There was, however, no significant Exemplar change [(RC+EC + EC) – (RC + No)] × Time interaction [$F(3.9,27.5) = 0.580$, $P > 0.3$], suggesting that subjects were able to reach maximal performance more rapidly following an exemplar change than following rule change, nor three-way interaction between Rule change, Exemplar change and Time [$F(4.1,28.6) = 0.570$, $P > 0.3$]. Nonetheless, performance also fell transiently following the introduction of new exemplars, and following no change, despite the rule remaining constant {significant at $P < 0.05$ in a one-sample *t*-test comparing average performance for the first exemplar presented [average(EC and No)$_{1st}$] against 100% performance}. This probably reflects subjects pre-empting a rule change. Critically, however, the fall in performance following exemplar change or no change was less than that following a rule change {significant at $P < 0.05$, one-tailed, in a paired *t*-test of the differences between performance for the first exemplar in the rule change conditions [average(RC+EC and RC)$_{1st}$] versus the exemplar change and no change conditions [average(EC and No)$_{1st}$]}. The presence of an effect of Rule change, but not Exemplar change, in the ANOVA, together with the results of paired *t*-tests, suggest that subjects had learned to categorize on the basis of an abstract rule, rather than a similarity-based process.

Functional Imaging

To determine rule learning-related functional neuroanatomy, we tested for time-dependent changes in neuronal activation following changes in rule where the temporal profile of modelled neuronal responses was tailored to each subject's learning rate. A significant main effect of new rule was observed in bilateral fronto-polar prefrontal cortices (FPPC) (Fig. 2; Table 1). Right FPPC (Fig. 2*a*) was significant ($P < 0.05$ corrected for multiple comparisons), with the left hemisphere homologous area significant at $P < 0.0001$ uncorrected (Fig. 2*b*). A further left FPPC region ($P < 0.0001$ uncorrected), lying in left superior frontal sulcus, also showed a main effect of rule learning (Fig. 2*c*). The parameter estimates and time course of the BOLD response clearly reveal that the exponentially decaying response in right (Fig. 2*a*) and left (Fig. 2*b*,*c*) FPPC was maximal during epochs following a rule change relative to those epochs in which the rule remained the same.

We tested for a hippocampal response to exemplar change in the same left anterior hippocampal region that we had previously found responsive to perceptual and exemplar novelty (Strange *et al*., 1999). Figure 3 demonstrates the SPM of the main effect of exemplar change-evoked exponential adaptation. As predicted, exemplar change evoked significant time-dependent changes in activation in left anterior hippocampus. The BOLD response and the parameter estimates for the epoch-related responses in this region show, however, that all four conditions produce a transient decrement in hippocampal activation. This decrease in hippocampal activation is alleviated by exemplar change. One possibility is that exemplar change-evoked activation in anterior hippocampus is superimposed on a transient task-related decrease in activation.

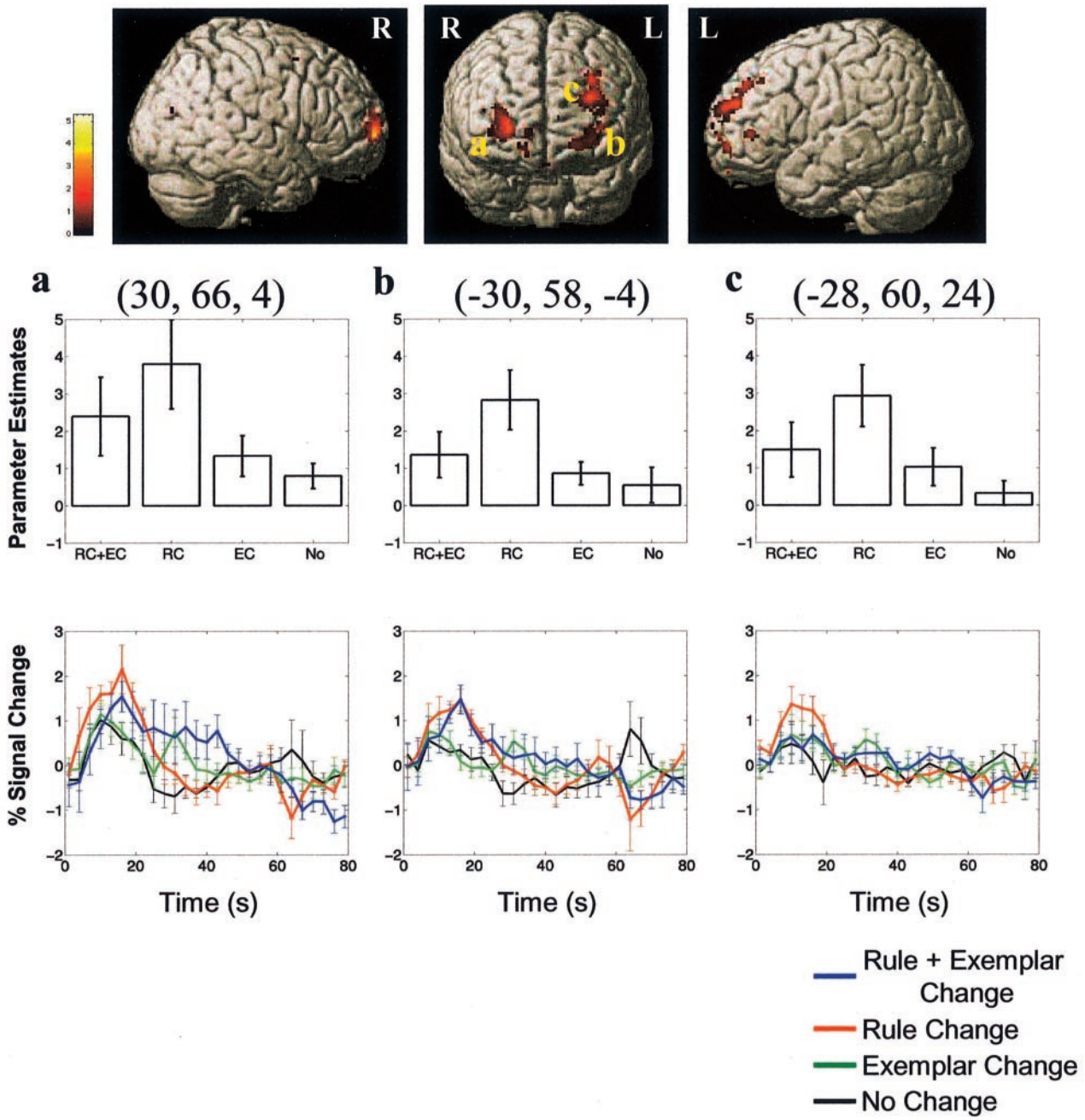

Figure 2. Main effect of rule change. The SPM (threshold $P < 0.001$) has been rendered onto a canonical T_1 structural image and shows activation of bilateral FPPC in response to change of rule. The coloured bar denotes the T value of the activation. Below are plotted the parameter estimates and the time course of the BOLD response (± SE of the mean across the eight subjects) for the four conditions relative to the control task in (*a*) right FPPC, (*b*) left FPPC and (*c*) left superior frontal sulcus. The parameter estimates pertain to the regressors modelling exponential decay of within-epoch activations for correct responses only (units are arbitrary). The BOLD response (expressed as % signal change) has been collapsed for each subject across the three replications of each condition and averaged across the eight subjects.

Discussion

Different psychological mechanisms have been proposed to account for the human ability to categorize stimuli. The brain regions responsible for these categorization processes have not been fully characterized. Our behavioural data provide evidence of transfer of categorization performance to perceptually novel exemplars, confirming that subjects learned to categorize letter strings on the basis of abstract rules and not merely on the basis of similarities between exemplars. Our imaging data show that the learning of an abstract rule selectively engages FPPC. Consistent with a rule learning response profile, the FPPC demonstrated a time-by-condition interaction following rule change, with the temporal profile of neuronal adaptation reflecting each subject's learning rate. A previous study measuring neuronal responses to rule changes, in the absence of awareness that the task was indeed rule-governed (Berns *et al.*, 1997), did not demonstrate activity in anterior prefrontal regions. This suggests that the FPPC role in rule learning reflects processes engaged during explicit requirements to find abstract structure (Shanks and St John, 1994; Dominey *et al.*, 1998), involving the

generation of hypotheses concerning relationships among stimuli (Shanks, 1995).

The precise functional roles of the fronto-polar region in man are not well characterized. Neuropsychological studies of patients with lesions to FPPC are to some degree confounded by an inability to control for the caudal extent of prefrontal lesions (Stuss and Benson, 1986) [but see Stuss *et al*. (Stuss *et al.*, 2000)]. Similarly, neurophysiological and lesion studies of non-human primate prefrontal cortex have generally focused on more posterior prefrontal areas (Fuster, 1989; Passingham, 1993) because of difficulty in accessing the frontal polar region without disrupting more caudal prefrontal cortex.

Despite methodological difficulties particular to functional imaging of FPPC [reviewed by Christoff and Gabrieli (Christoff and Gabrieli, 2000)], functional imaging studies have provided preliminary indications concerning the functional roles of this region. Activation of FPPC has been evoked by complex cognitive tasks, in particular reasoning tasks. Despite evoking activation in multiple and heterogeneous brain regions, reasoning tasks such as the Wisconsin Card Sorting Test (WCST) (Berman *et al.*, 1995; Nagahama *et al.*, 1996; Goldberg *et al.*, 1998; Rogers *et al.*, 2000), the Tower of London task (Baker *et al.*, 1996), inductive and probabilistic reasoning tasks (Goel *et al.*, 1997; Osherson *et al.*, 1998), probabilistic classification (Poldrack *et al.*, 1999) and the Raven's progressive matrices test (Prabhakaran *et al.*, 1997) show consistent activation in FPPC. Of these reasoning tasks, our rule change condition shares the greatest similarity with the WCST, a task considered a robust index of prefrontal function (Milner, 1963). The WCST is a series of visual discriminations across multidimensional stimuli, in which the rule governing reinforcement is periodically changed across different dimensions of the stimuli (Grant and Berg, 1948). Hence, like the above reasoning tasks, the WCST is a heterogeneous task, evoking activation in multiple brain regions (Berman *et al.*, 1995; Nagahama *et al.*, 1996; Goldberg *et al.*, 1998; Rogers *et al.*, 2000). However, a previous study has shown that when brain activity associated with sorting new exemplars under a constant rule is removed from that evoked by sorting exemplars following rule change, the rule change condition evokes activation in anterior superior frontal gyrus and FPPC (Rogers *et al.*, 2000).

The interpretation of previous functional imaging experiments of reasoning or rule learning is, however, limited. These studies used PET (Berman *et al.*, 1995; Baker *et al.*, 1996; Nagahama *et al.*, 1996; Goel *et al.*, 1997; Goldberg *et al.*, 1998; Osherson *et al.*, 1998; Smith *et al.*, 1998; Rogers *et al.*, 2000) or fMRI epoch designs (Prabhakaran *et al.*, 1997; Fletcher *et al.*, 1999; Poldrack *et al.*, 1999; Goel and Dolan, 2000) that require averaging of evoked responses, including those to correct and incorrect trials, over extended periods of 30 s or more. The present experiment enables us to make more specific inferences

Table 1
Main effect of rule ($P < 0.0001$ uncorrected)

Brain region	Talairach coordinates(x, y, z)	Z value
Right FPPC (frontal pole; BA 10)	(30, 66, 4)	5.26*
Left superior frontal sulcus (BA 9/10)	(–28, 60, 24)	4.54
Right inferomedial FPPC (frontal pole; BA 10)	(14, 56, –10)	4.30
Left ventrolateral prefrontal cortex (BA 47)	(–36, 40, 4)	4.18
Left FPPC (frontal pole; BA 10)	(–30, 58, –4)	4.03

*$P < 0.05$ corrected.

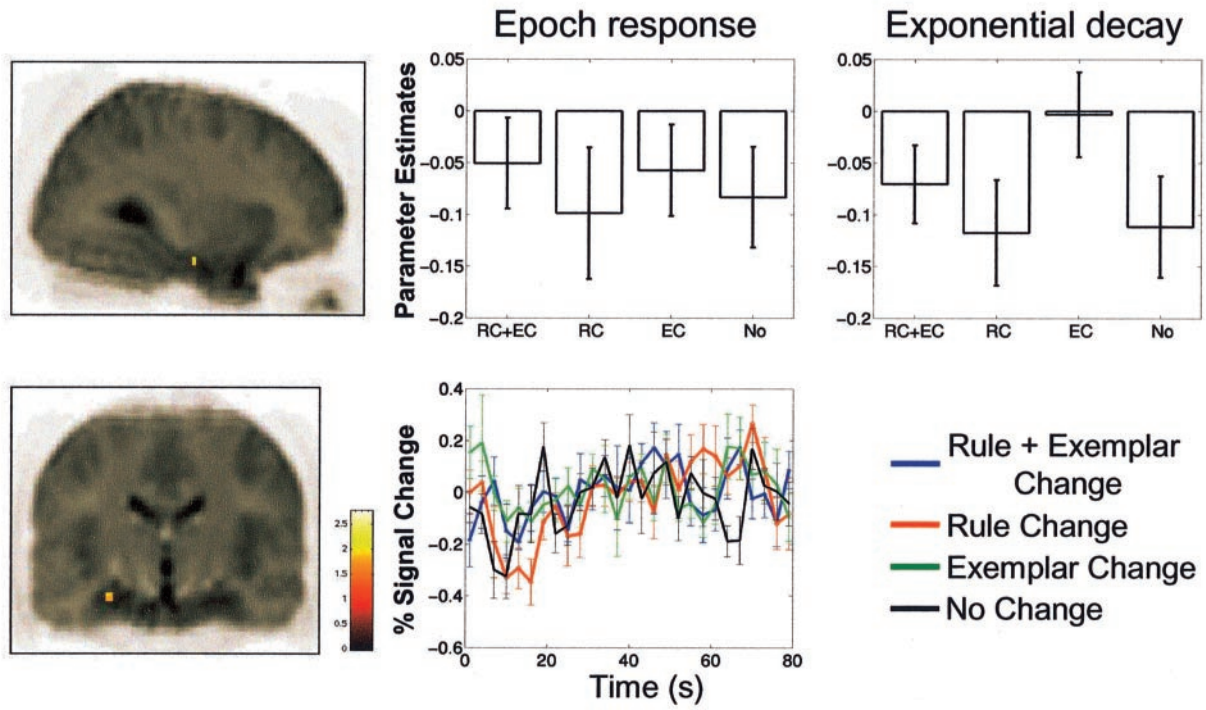

Figure 3. Left anterior hippocampus responds to exemplar change. The SPM (threshold $P < 0.05$) of the main effect of exemplar change-evoked exponential adaptation has been superimposed on a coronal section ($y = -14$) and sagittal section ($x = -30$) of a functional image to demonstrate left anterior hippocampal activation (–30, –14, –20). This image is the mean functional image (produced for each subject during realignment) averaged for the 10 subjects with grey-scale inversion for ease of illustration. Superimposing the SPM on a functional image avoids the issue of distortion in T_1 to T_2* co-registration, which is particularly evident in anterior medial temporal lobe structures, and allows more reliable anatomical identification. For presentation, this SPM has been masked by the main effect of the exemplar change-evoked epoch response. The parameter estimates (pertaining to both the epoch response and exponential decay function) and BOLD response for this activation are shown on the right.

through use of an event-related design that models correct and incorrect trials separately. Furthermore, our design allows us to model neuronal adaptation tailored to each subject's learning rate.

Neuropsychological studies that attempt to dissociate consequences of lesions to different loci of human prefrontal cortex, despite their limitations, lend support to the importance of anterior frontal regions in rule learning. Damage to superior medial frontal areas (including rostral BA 9 and 10) produces impairment in the WCST that is equivalent to that produced by dorsolateral prefrontal cortex (DLPFC) lesions (Stuss *et al*., 2000). In fact, the superior medial frontal group of Stuss *et al*. showed a greater inability to switch sorting category than the DLPFC group, supporting our observation that these regions are critically engaged by rule changes. This finding is in agreement with the suggestion that FPPC mediates switching between different executive processes (Fletcher and Henson, 2001). It should be noted, however, that task switching has been shown to engage other cortical areas besides prefrontal cortex (Kimberg *et al*., 2000; Smith *et al*., 2001).

The patients with DLPFC lesions reported by Stuss *et al*. (Stuss *et al*., 2000) showed more set losses (failures to consistently apply a categorization rule once it is determined) than the superior medial frontal group. This possible DLPFC role in rule application speaks to previous findings (Fletcher *et al*., 1999; Seger *et al*., 2000) of left DLPFC activation with gradual rule acquisition. Neurophysiological recordings in non-human primates demonstrate that prefrontal cortex (dorsal, ventral and dorsolateral) plays a role in guiding behaviour according to previously learned rules (White and Wise, 1999). Taken together with our current finding, we suggest that the FPPC is engaged during intentional or explicit rule induction but once a rule is learnt, more posterior prefrontal areas mediate rule application. We did not find DLPFC to be differentially activated (increasing or decreasing) following change in rule and we suggest that these regions are active in all four conditions (including the no change condition, as this condition also involves rule application).

In the current study, hypothesis generation and testing requires multiple trials to be held in mind. In addition to reasoning tasks, FPPC activation has been evoked during working memory tasks. Critically, FPPC activation is observed when working memory loads approach/exceed people's short-term memory capacity (Grasby *et al*., 1993; Smith *et al*., 1996; Jonides *et al*., 1997; Rypma *et al*., 1999) or when working memory is performed in a dual-task context (Grafton *et al*., 1995; MacLeod *et al*., 1998). Both of these manipulations of working memory are likely to encourage the development of strategies to maintain performance. Koechlin *et al*. attributed activation of FPPC exclusively to 'branching' (Koechlin *et al*., 1999), a process required when tasks involve setting up and maintaining an overall goal while concurrently setting and achieving sub-goals (Fletcher and Henson, 2001). Our rule-learning task did not involve branching, as there was only one goal, rule induction, to be achieved.

In addition to working memory, engaging in episodic memory retrieval consistently activates FPPC [for reviews see Nolde *et al*. and Christoff and Gabrieli (Nolde *et al*., 1998; Christoff and Gabrieli, 2000)]. These activations have been attributed to, amongst other processes, post-retrieval evaluation of the products of the retrieval process [(Shallice *et al*., 1994; Rugg and Wilding, 2000); though see Lepage *et al*. (Lepage *et al*., 2000)]. In the current study, rule learning may require evaluation of the products of recollecting past trials (i.e. the stimulus, response and feedback) to guide subsequent responses. A similar interpretation was given by Reber *et al*. for their observation of FPPC activation during processing of categorical versus noncategorical patterns (Reber *et al*., 1998). An emerging theme, therefore, suggests that activations in FPPC occur in high level tasks that involve planning and executive control of cognitive functions. In particular, many of the tasks require a strategy or evaluative process be applied to information held on-line, for example, to generate and test hypotheses on multiple items during rule learning.

We have previously reported a left anterior hippocampal response to both exemplar and perceptual novelty in the context of a developing rule system (Strange *et al*., 1999). Here we replicate this finding by demonstrating that change in the surface features of exemplars activates left anterior hippocampus, in the same region previously activated. Exemplar change-evoked activation in anterior hippocampus is superimposed on a transient task-related decrease in activation. This response profile is similar to that previously observed (Strange *et al*., 1999), where the enhanced anterior hippocampal response to perceptual novelty was in the context of relative hippocampal deactivation. The WCST (Berman *et al*., 1995) and probabilistic learning (Poldrack *et al*., 1999) also produce a relative decrease in hippocampal activation. These observations suggest that high level cognitive tasks, such as rule learning, that activate frontal regions may also cause relative hippocampal deactivation.

We attribute the left anterior hippocampal response to the perceptual novelty effected by changing the letters subtending the presented stimuli. There are, however, other interpretations. In the current and previous (Strange *et al*., 1999) study, subjects were required to process the relative positions of letters. Relational processing is a hypothesized function of the hippocampus (Cohen *et al*., 1999). Furthermore, it has been argued that subjects can apply multiple categorization strategies simultaneously (Smith *et al*., 1998). Hence, although the emphasis of our task was on explicit rule abstraction, to the extent that similarity-based processes were also operating (perhaps automatically, in parallel), hippocampal activation that tracked exemplar changes could reflect similarity-based categorization. Importantly, medial temporal regions were not engaged by rule changes, which agrees with previous observations that medial temporal lobe lesions do not prevent the acquisition of abstract knowledge in categorization tasks, despite impairing memory for individual items (Knowlton and Squire, 1993).

Our findings suggest that fronto-polar prefrontal cortex selectively mediates rule learning in a categorization task emphasizing explicit rule induction. This suggestion, supported by previous PET and epoch-related fMRI studies of reasoning, implies that the frontal poles are engaged when subjects perform complex problem-solving tasks. Change in surface features during categorization engages left anterior hippocampus, supporting our previous proposal of novelty-evoked activation in this region.

Notes

This work was supported by programme grants from the Wellcome Trust to R.J.D. and K.J.F. B.A.S. is supported by the Astor Foundation Scholarship. R.N.A.H. is supported by Wellcome Trust Grant 060924.

Address correspondence to Bryan A. Strange, Wellcome Department of Cognitive Neurology, Functional Imaging Laboratory, Queen Square, London WC1N 3BG, UK. Email: bstrange@fil.ion.ucl.ac.uk.

References

Baker SC, Rogers RD, Owen AM, Frith CD, Dolan RJ, Frackowiak RSJ,

Robbins TW (1996) Neural systems engaged by planning: a PET study of the Tower of London task. Neuropsychologia 34:515–526.
Berman KF, Ostrem JL, Randolph C, Gold J, Goldberg T, Coppola R, Carson RE, Herscovitch P, Weinberger DR (1995) Physiological activation of a cortical network during performance of the Wisconsin Card Sorting Test: a positron emission tomography study. Neuropsychologia 33:1027–1046.
Berns GS, Cohen JD, Mintun MA (1997) Brain regions responsive to novelty in the absence of awareness. Science 276:1272–1275.
Christoff K, Gabrieli JDE (2000) The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organisation within the human prefrontal cortex. Psychobiology 28:168–186.
Cohen NJ, Ryan J, Hunt C, Romine L, Wszalek T, Nash C (1999) Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. Hippocampus 9:83–98.
Dominey PF, Lelekov T, Ventre-Dominey J, Jeannerod M (1998) Dissociable processes for learning the surface structure and abstract structure of sensorimotor sequences. J Cogn Neurosci 10:734–751.
Fletcher P, Henson RNA (2001) Frontal lobes and human memory. Insights from functional neuroimaging. Brain 124:849–881.
Fletcher P, Buchel C, Josephs O, Friston K, Dolan R (1999) Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. Cereb Cortex 9:168–178.
Friston KJ, Holmes A, Worsley K, Poline J, Frith C, Heather J, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.
Fuster JM (1989) The prefrontal cortex. Anatomy, physiology and neuropsychology of the frontal lobe. New York: Raven Press.
Goel V, Dolan RJ (2000) Anatomical segregation of component processes in an inductive inference task. J Cogn Neurosci 12:110–119.
Goel V, Gold B, Kapur S, Houle S (1997) The seats of reason? An imaging study of deductive and inductive reasoning. NeuroReport 8: 1305–1310.
Goldberg TE, Berman KF, Fleming K, Ostrem J, Van Horn JD, Esposito G, Mattay VS, Gold JM, Weinberger DR (1998) Uncoupling cognitive workload and prefrontal cortical physiology: a PET rCBF study. NeuroImage 7:296–303.
Grafton ST, Hazeltine E, Ivry R (1995) Functional mapping of sequence learning in normal humans. J Cogn Neurosci 7:497–510.
Grant AD, Berg A (1948) A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Wiegl-type card-sorting problem. J Exp Psychol 38:404–411.
Grasby PM, Frith CD, Friston KJ, Bench C, Frackowiak RSJ, Dolan RJ (1993) Functional mapping of brain areas implicated in auditory–verbal memory function. Brain 116:1–20.
Jonides J, Schumacher EH, Smith EE, Lauber EJ, Awh E, Minoshima S, Koeppe R (1997) Verbal working memory load affects regional brain activation as measured by PET. J Cogn Neurosci 9:462–475.
Josephs O, Turner R, Friston KJ (1997) Event-related fMRI. Hum Brain Mapp 5:243–248.
Kimberg DY, Aguirre GK, D'Esposito M (2000) Modulation of task-related neural activity in task-switching: an fMRI study. Cogn Brain Res 10: 189–196.
Knowlton BJ, Squire LR (1993) The learning of categories: parallel brain systems for item memory and category knowledge. Science 262:1747–1749.
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.
Lepage M, Ghaffar O, Nyberg L, Tulving E (2000) Prefrontal cortex and episodic retrieval mode. Proc Natl Acad Sci USA 97:506–511.
MacLeod AK, Buckner RL, Miezin FM, Petersen SE, Raichle ME (1998) Right anterior prefrontal cortex activation during semantic monitoring and working memory. NeuroImage 7:41–48
Milner B (1963) Effects of different brain lesions on card sorting. Arch Neurol 28:100–110.
Nagahama Y, Fukuyama H, Yamauchi H, Matsuzaki S, Konishi J, Shibasaki H, Kimura J (1996) Cerebral activation during performance of a card sorting test. Brain 119:1667–1675.
Nolde SF, Johnson MK, Raye CL (1998) The role of prefrontal cortex during tests of episodic memory. Trends Cogn Sci 2:399–406.
Nosofsky RM (1986) Attention, similarity, and the identification–categorization relationship. J Exp Psychol Gen 115:38–57.
Osherson D, Perani D, Cappa S, Schnur T, Grassi F, Fazio F (1998) Distinct brain loci in deductive versus probabilistic reasoning. Neuropsychologia 36:369–376.
Passingham R (1993) The frontal lobes and voluntary action. Oxford: Oxford University Press.
Perruchet P, Pacteau C (1990) Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge. J Exp Psychol Gen 119:264–275.
Poldrack RA, Prabhakaran V, Seger CA, Gabrieli JDE (1999) Striatal activation during acquisition of a cognitive skill. Neuropsychology 13:564–574.
Prabhakaran V, Smith JAL, Desmond JE, Glover GH, Gabrieli JDE (1997) Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's Progressive Matrices test. Cognit Psychol 33:43–63.
Reber AS (1967) Implicit learning of artificial grammars. J Verbal Learn Verbal Behav 5:855–863.
Reber PJ, Stark CEL, Squire LR (1998) Cortical areas supporting category learning identified using functional MRI. Proc Natl Acad Sci USA 95:747–750.
Rogers RD, Andrews TC, Grasby PM, Brooks DJ, Robbins TW (2000) Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. J Cogn Neurosci 12:142–162.
Rugg MD, Wilding EL (2000) Retrieval processing and episodic memory. Trends Cogn Sci 4:108–115.
Rypma B, Prabhakaran V, Desmond JE, Glover GH, Gabrieli JDE (1999) Load-dependent roles of frontal brain regions in the maintenance of working memory. NeuroImage 9:216–226.
Seger CA, Poldrack RA, Prabhakaran V, Zhao M, Glover GH, Gabrieli JDE (2000) Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. Neuropsychologia 38:1316–1324.
Shallice T, Fletcher PC, Frith CD, Grasby P, Frackowiak RSJ, Dolan RJ (1994) Brain regions associated with acquisition and retrieval of verbal episodic memory. Nature 368:633–635.
Shanks DR (1995) The psychology of associative learning. Cambridge: Cambridge University Press.
Shanks DR, St John MF (1994) Characteristics of dissociable human learning systems. Behav Brain Sci 17:367–447.
Smith EE, Langston C, Nisbett RE (1992) The case for rules in reasoning. Cognit Sci 16:1–40.
Smith EE, Jonides J, Koeppe RA (1996) Dissociating verbal and spatial working memory using PET. Cereb Cortex 6:11–20.
Smith EE, Patalano AL, Jonides J (1998) Alternative strategies of categorization. Cognition 65:167–196.
Smith EE, Geva A, Jonides J, Miller A, Reuter-Lorenz P, Koeppe RA (2001) The neural basis of task-switching in working memory: effects of performance and aging. Proc Natl Acad Sci USA 98:2095–2100.
Strange BA, Fletcher PC, Henson RNA, Friston KJ, Dolan RJ (1999) Segregating the functions of human hippocampus. Proc Natl Acad Sci USA 96:4034–4039.
Stuss DT, Benson DF (1986) The frontal lobes. New York: Raven Press.
Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, Murphy KJ, Izukawa D (2000) Wisconsin card sorting test performance in patients with focal frontal and posterior brain damage: effects on lesion location and test structure on separable cognitive processes. Neuropsychologia 38:388–402.
Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.
White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. Exp Brain Res 126:315–335.