Chapter 4

Neuroanatomical Correlates of Rule Learning

4.1 Introduction:

Chapter 3 described a left anterior hippocampal response to both exemplar and perceptual novelty in the context of learning a rule system. The experiment presented in this chapter addresses specific questions raised by these findings. The previous experiment examined responses to novelty during category learning, with category defined as 'grammatical' or 'ungrammatical'. The psychological processes through which humans learn to categorise stimuli have been studied extensively (Smith *et al*., 1998). Considerable interest surrounds the proposal that people abstract rules that define category membership unconsciously, through simple exposure to exemplars of the categories (Reber, 1967). This proposal remains a source of controversy (Shanks, 1995). Firstly, much of the evidence that claims to demonstrate abstract rule learning can equally be explained in terms of categorisation on the basis of superficial similarity, either between whole exemplars (instance-based categorisation; Nosofsky, 1986) or exemplar parts (fragment-based categorisation; 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).

The modified artificial grammar learning paradigm used in chapter 3 was first used by Fletcher *et al*. (1999) to examine neuroanatomical correlates of category learning by measuring haemodynamic responses associated with learning the rules of the grammar. In that study, and the experiment presented in chapter 3, the extent to which category learning was implicit or explicit, or based on similarity- or rule-based mechanisms, was unclear. Contrary to previous artificial grammar studies, learning was intentional rather than incidental, with the grammatical status of exemplars

indicated with trial-by-trial feedback, which encourages explicit rule induction. Nonetheless, learning may also have involved either 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 entire experiment.

The critical test of abstract rule-based learning is whether categorisation performance transfers to exemplars drawn from a new vocabulary (for which similarity-based mechanisms cannot operate; Smith *et al*., 1992). Though Fletcher *et al*. (1999) demonstrated some transfer of categorisation performance from one set of exemplars to another (see figure 3.1 for similar results), these exemplars were drawn from the same vocabulary set hence transfer could also have reflected similarity-based processes.

Because of limitations in interpreting the nature of category learning in the previous experiment, the brain regions associated with rule learning per se were not examined. In the present study, the neural correlates of explicit, abstract rule induction were addressed. Subjects were required to categorise 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 4-letter strings, was simple enough for subjects to learn over the course of 20 trials (see figure 4.1). This 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 allowed a test of whether performance

transferred across exemplar changes, and so establish whether subjects had successfully abstracted the underlying rules. Measuring hippocampal responses following an exemplar change provided a means to replicate the findings presented in chapter 3.

To measure rule learning-dependent responses, event-related fMRI was used to test for responses, to correct categorisation 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. I 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), the specific hypothesis that could be tested was that rule learning is frontally-mediated. By contrast, on the basis of the data presented in chapter 3, it was predicted that exemplar change would engage the hippocampus, consistent with the proposal of an automatic response to perceptual and exemplar novelty in this region (chapter 3). Finally, the current paradigm also enabled a test of whether rule change, a 'high level' abstract form of novelty, would engage the hippocampus.

4.2 Materials and Methods:

4.21 Subjects

Informed consent was obtained from 10 right-handed subjects (6 male, 4 female; age range 22 - 37 yrs; mean age 27.4; recruited by advertisement). Data from two subjects (1 male, 1 female) were excluded from the analysis due to poor performance.

4.22 Psychological task

During scanning subjects were presented with strings of 4 letters in upper case, at a rate of one every 4 seconds. 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 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". 20 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 2x2 factorial design, with rule change as one factor and letter set (exemplar) change as another factor (see figure 4.1). 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 order of

conditions was randomised and each condition was repeated 3 times. Each activation epoch was followed by a control epoch during which the strings LLLL or RRRR were presented (5 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 4 cells in the 2x2 factorial design. Note that this 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.

4.23 Data acquisition

A total of 480 volumes were acquired per subject plus 5 'dummy' volumes, subsequently discarded, to allow for T1 equilibration effects. Volumes were acquired continuously every 3000 ms. Each volume comprised thirty 3mm axial slices, with an in-plane resolution of 3x3mm, positioned to cover the whole cerebrum. The imaging time series was realigned to correct for interscan movement and normalised 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*., 1995a).

4.24 Data analysis

Data were analysed using Statistical Parametric Mapping (SPM99) employing an event-related model (Josephs *et al*., 1997). Event-related fMRI is used to detect and characterise transient haemodynamic responses to brief stimuli or tasks (Josephs and Henson, 1999; see chapter 2, part II). To identify changes in the BOLD response evoked by single stimuli, responses can be modelled with basis functions of

peri-stimulus time. The basis function used here was a synthetic, canonical haemodynamic response function. This function comprises the weighted sum of two gamma functions to approximate an empirically-derived haemodynamic impulse response (Friston *et al*., 1998a).

The data were first filtered to remove low frequency drifts in signal (cut-off 174 secs). In the analysis testing for the effects of rule change, 4 distinct effects of interest were specified: 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, USA). 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, 4 effects were modelled for each of the 4 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 sec activation epoch (epoch responses) embody mean changes in brain activity, following change in either rule or exemplar. The regressors modelling the

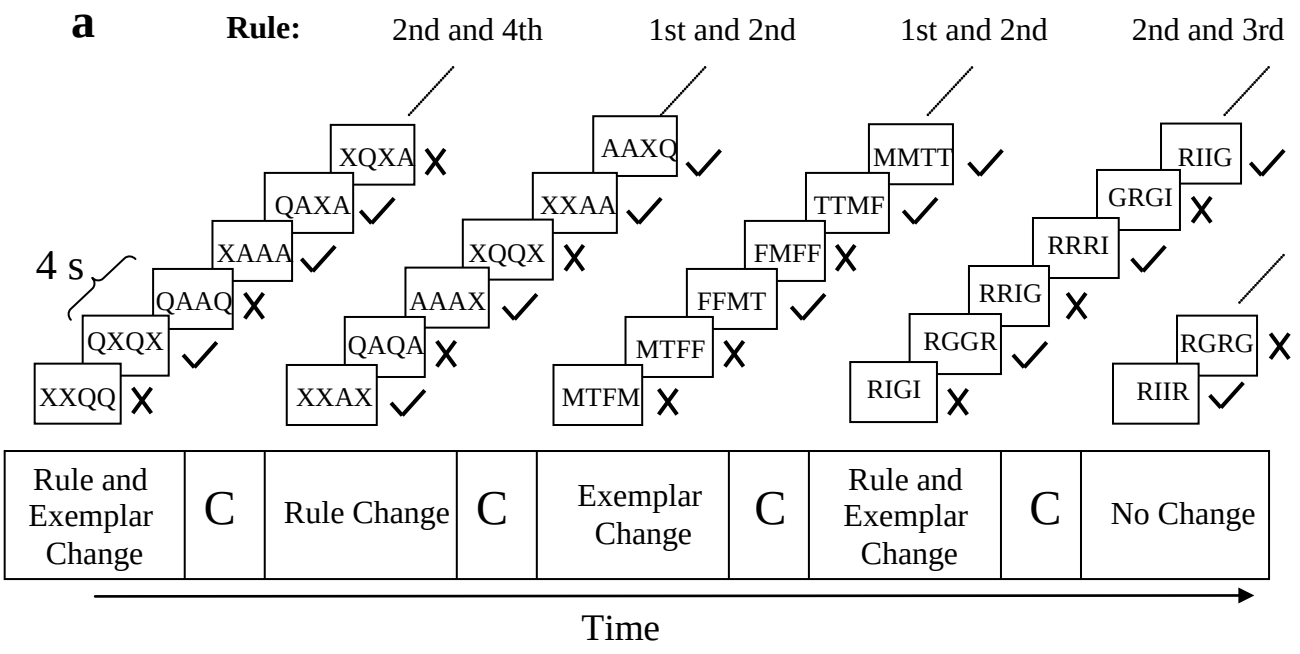
exponential decay embody subject-specific learning-dependent responses within an epoch. Only contrasts involving correct responses were entered into the statistical analyses (there were too few incorrect responses for these regressors to be tested). Movement parameters, determined during realignment, were entered as covariates of no interest, to remove possible movement-related residual effects.

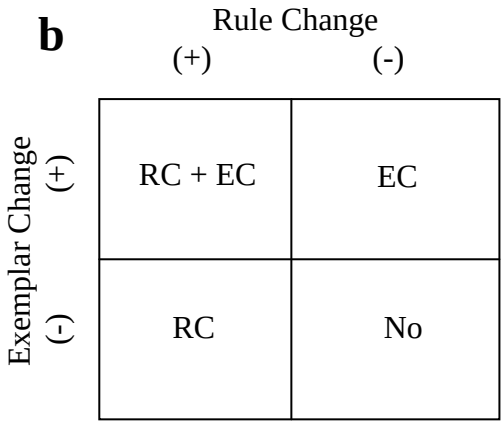
A random effects analysis was the default analysis for experiments in this thesis. This analysis, however, failed to give any corrected p values, most likely due to the small number of subjects included in the analysis. Hence the statistical criteria were relaxed and a fixed effects analysis performed, allowing inferences to be made about the particular group of subjects tested but not the population at large. 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. All rule learning-related effects are reported that survive a height threshold of $p<0.0001$ (uncorrected) at a spatial extent threshold of 5 voxels.

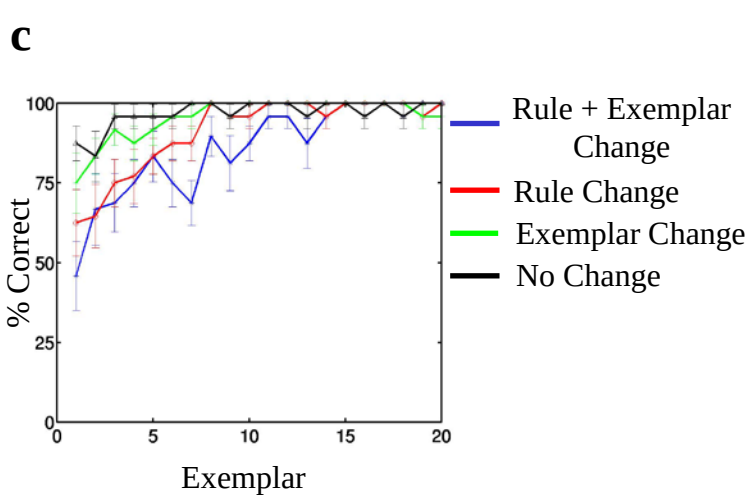
A similar analysis was conducted to test for response adaptation following exemplar change. The purpose of this second analysis was to focus on a region of interest, the anterior hippocampus (used here to refer to the dentate gyrus, CA subfields and subiculum), which was previously implicated in novelty detection (chapter 3). In this analysis, instead of modelling a subject-specific performance-related exponential decay, an arbitrary exponential function was chosen to model adaptation to exemplar novelty, since there was no reason to predict that automatic

responses to novelty would be correlated with subject-specific behaviour. The same function modelled novelty-dependent responses in all subjects.

Figure 4.1. 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 1st to 4th position of each string. (b) The 2x2 factorial design. (c) The average performance of the 8 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 is shown in blue; rule change in red; exemplar change in green; and no change in black. Abbreviations here and in all subsequent figures: RC+EC: epoch following change of both rule and exemplar. RC: epoch following change in just rule. EC: epoch following change in just exemplar. No: epoch following no change.


a
Rule:
2nd and 4th
1st and 2nd
1st and 2nd
2nd and 3rd
4 s
XXQQ
QXQX
QAAQ
XAAA
QAXA
XQXA
XXAX
QAQA
AAAX
XQQX
XXAA
AAXQ
MTFM
MTFF
FFMT
FMFF
TTMF
MMTT
RIGI
RGGR
RRIG
RRRI
GRGI
RIIG
RIIR
RGRG
Rule and Exemplar Change
C
Rule Change
C
Exemplar Change
C
Rule and Exemplar Change
C
No Change
Time


b
Rule Change
(+)
(-)
Exemplar Change
(+)
(-)
RC + EC
EC
RC
No


c
% Correct
100
75
50
25
0
0
5
10
15
20
Exemplar
Rule + Exemplar Change
Rule Change
Exemplar Change
No Change

4.3 Results:

4.31 Behaviour

Figure 4.1c demonstrates that performance fell following a rule change but then improved over trials, reaching 100% by the end of each rule change epoch. As predicted, a repeated measures 2 x 2 x 20 ANOVA demonstrated a significant Rule change [(RC+EC + RC) – (EC + No)] x 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)] x 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 was there a three-way interaction between Rule change, Exemplar change and Time ($F_{4.1,\ 28.6}$=0.570, $p>0.3$). Nonetheless, performance 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 behavioural profile most 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 + 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 categorise on the basis of an abstract rule, rather than a similarity-based process.

4.32 Functional imaging

To determine rule learning-related functional neuroanatomy, I 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 cortex (FPPC; figure 4.2; Table 4.1). Right FPPC (figure 4.2a) was significant at a corrected level ($p<0.05$ corrected for multiple comparisons), with the left hemisphere homologous area significant at $p<0.0001$ uncorrected (figure 4.2b). A further left FPPC region ($p<0.0001$ uncorrected), lying in left superior frontal sulcus, also showed a main effect of rule learning (figure 4.2c). The parameter estimates and time course of the BOLD response clearly reveal that the exponentially-decaying response in right (figure 4.2a) and left (figure 4.2b and c) FPPC was maximal during epochs following a rule change relative to those epochs in which the rule remained the same.

A hippocampal response to exemplar change was also tested for in the same left anterior hippocampal region that was found previously to respond to perceptual and exemplar novelty (chapter 3). Figure 4.3 demonstrates the SPM of the main effect of exemplar change-evoked exponential adaptation (thresholded at $p<0.05$, uncorrected). 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.

Table 4.1: Main effect of rule (p<0.0001 uncorrected)

Brain Region	Talairach Co-ordinates (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 4.2. Main effect of rule change. The SPM (threshold p<0.001) has been rendered onto a canonical T1 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 8 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 3 replications of each condition and averaged across the 8 subjects.

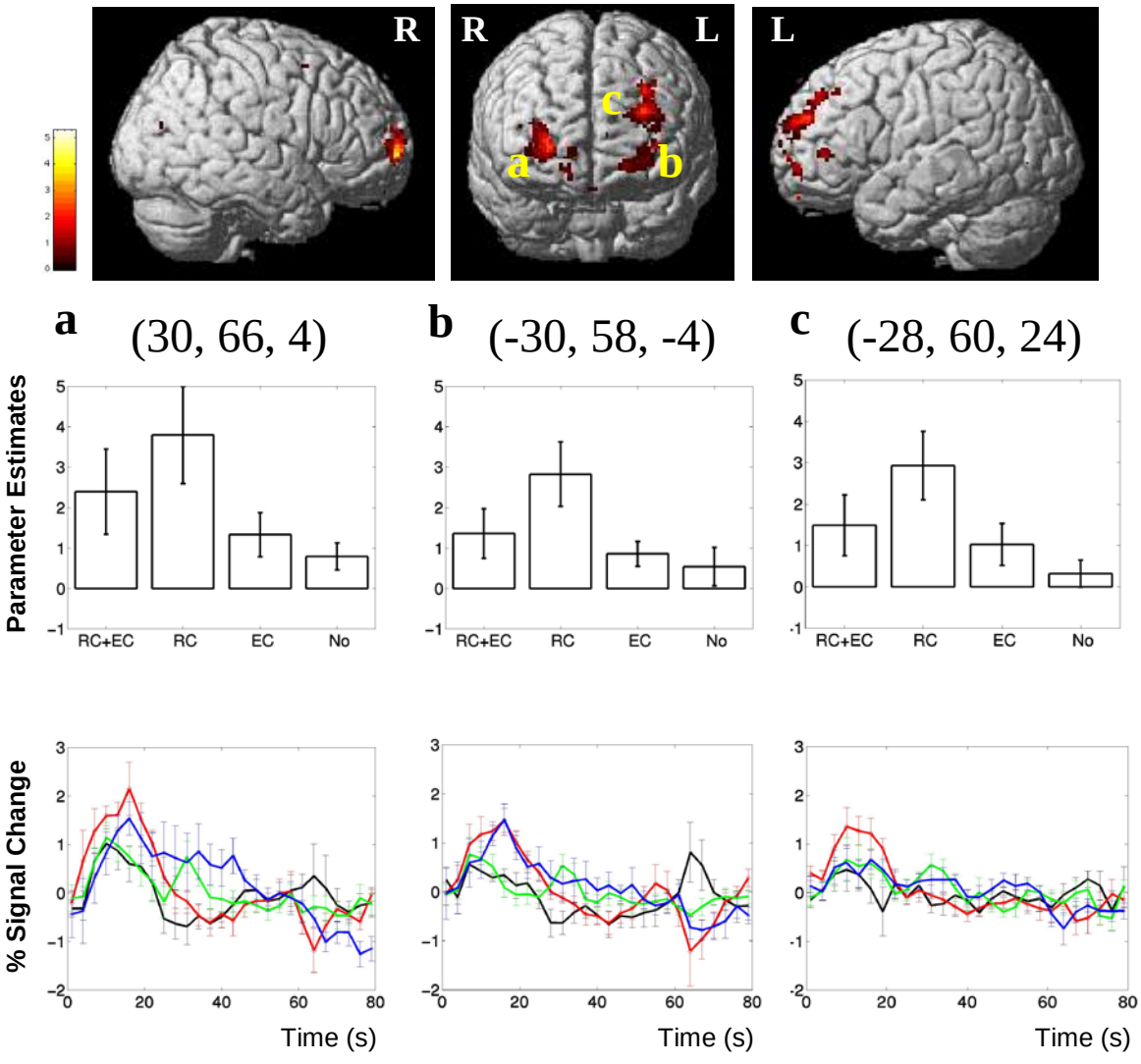
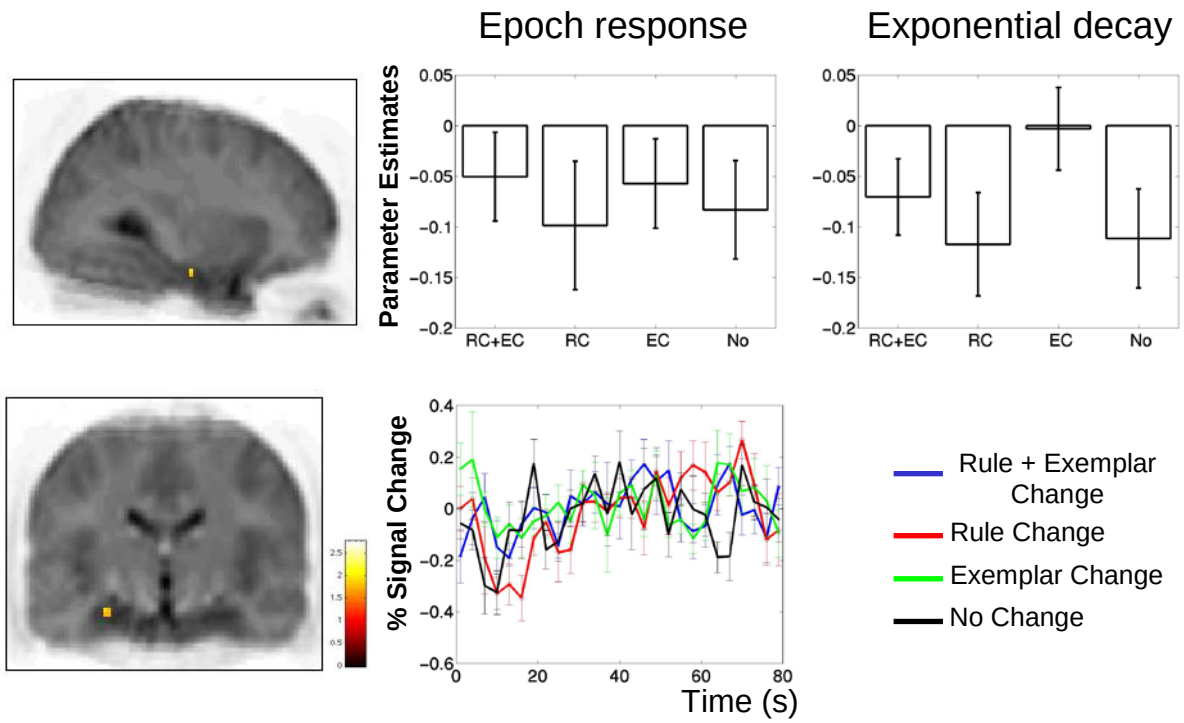


Figure 4.3. Left anterior hippocampus responds to exemplar change. The SPM (threshold p<0.05) of the main effect of exemplar change-evoked exponential adaptation has been superimposed on a coronal section (y=-14) and sagittal section (x=-30) of a functional image to demonstrate left anterior hippocampal activation (-30, -14, -20). This image is the mean functional image (produced for each subject during realignment) averaged for the 8 subjects with grey-scale inversion for ease of illustration. Superimposing the SPM on a functional image avoids the issue of distortion in T1 to T2* coregistration, 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.



4.4 Discussion

4.41 Fronto-polar prefrontal cortex

Different psychological mechanisms have been proposed to account for the human ability to categorise stimuli. The brain regions responsible for these categorisation processes have not been fully characterised. The current behavioural data provide evidence of transfer of categorisation performance to perceptually novel exemplars, confirming that subjects learned to categorise letter strings on the basis of abstract rules and not merely on the basis of similarities between exemplars. The 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 characterised. 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*., 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, 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) and the Raven's progressive matrices test (Prabhakaran *et al*., 1997), show consistent activation in FPPC. Of these reasoning tasks, the 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; 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 enabled more specific inferences to be made through use of an event-related design that modelled correct and incorrect trials separately. Furthermore, the design allowed modelling of neuronal adaptation to reflect each subject's learning rate.

Neuropsychological studies that attempt to dissociate consequences of lesions to different human prefrontal loci, despite their limitations, lend support to the importance of anterior frontal regions in rule learning. Damage to superior medial frontal areas (including rostral BA9 and 10) produces impairment in the WCST that is equivalent to that produced by dorsolateral prefrontal (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. The DLPFC group did, however, show more set losses (failures to consistently apply a categorisation rule once it is determined) than the superior medial frontal group (Stuss *et al*., 2000). This possible DLPFC role in rule application speaks to the previous finding by Fletcher *et al*. (1999) 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 the current finding, it could be suggested that the FPPC is engaged during intentional or explicit rule induction but once a rule is learnt, more posterior prefrontal areas mediate rule application. DLPFC was not found to be differentially activated (increasing or decreasing) following change in rule, which might reflect these regions being 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*., (1999) attributed activation of FPPC exclusively to "branching", a process required when tasks involve setting up and maintaining an overall goal while concurrently setting and achieving sub-goals (see Fletcher and Henson, 2001, for a review). The present 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 review see 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*., 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.* (1998) for their observation of FPPC activation during processing of categorical vs noncategorical patterns. 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.

4.42 Hippocampus

Chapter 3 described a left anterior hippocampal response to both exemplar and perceptual novelty in the context of a developing rule system. The results from the current experiment replicate this finding by demonstrating that change of exemplar vocabulary activates left anterior hippocampus, in close proximity to the region previously activated. Furthermore, because the statistical analysis was restricted to correct responses, the anterior hippocampal novelty response is not confounded by performance, a potential criticism of the experiment presented in chapter 3.

Changing the letters making up the stimuli does not affect category membership. The behavioural results indicate that subjects were responding on the basis of abstract rules hence exemplar changes, like font changes in chapter 3, are not

of behavioural relevance. It could, therefore, be concluded that the left anterior hippocampal response observed in the current experiment indexes perceptual novelty effected by changing the letters subtending the presented stimuli. There are, however, other interpretations. It has been argued that subjects can apply multiple categorisation strategies simultaneously (Smith *et al*., 1998). Hence, although the emphasis of the 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 categorisation. Furthermore, in the current and previous (chapter 3) study, subjects were required to process the relative positions of letters. Relational processing is a hypothesised function of the hippocampus (Cohen *et al*., 1999). Specifically, it is thought that the hippocampus binds together multiple inputs to permit representations of the relations among the constituent elements of a scene (Cohen *et al*., 1999). In the current and previous study, the constituent elements are letters that must be processed relative to one another.

The general response profile for each condition was a transient depression of anterior hippocampal activation. Exemplar change-evoked activation in anterior hippocampus appears, therefore, to be superimposed on this transient task-related decrease in activation. This response profile is similar to that observed in chapter 3, where the enhanced anterior hippocampal response to perceptual novelty was in the context of relative hippocampal deactivation (figure 3.2). The WCST has also been shown to produce a relative decrease in hippocampal activation (Berman *et al*., 1995). These observations suggest that high level cognitive tasks, such as rule

learning, that activate frontal regions may also cause relative hippocampal deactivation.

It could be suggested that reduction in hippocampal activation reflects an active inhibition, which may be effected by processing in prefrontal cortex. There exist extensive, direct connections between hippocampus and prefrontal cortex (Goldman-Rakic *et al*., 1984; Thierry *et al*., 2000; see chapter 7). If this proposed prefrontal modulation of hippocampal activation is a necessary component of the prefrontal processing that enables rule induction, reduced hippocampal activation should only be evident in the epochs following rule change. This was not the case as the left anterior hippocampal response decreased at the start of all activation epochs. However, the transient decrease in performance levels that followed the start of new exemplar and no change epochs was interpreted as subjects pre-empting a rule change. The fact that activation in left anterior hippocampus is reduced at the start of no change epochs may, therefore, reflect an automatic prefrontal-hippocampal modulation recruited when subjects anticipate a rule change. This interpretation requires that the decrease in hippocampal activation following the start of the no change epoch return to baseline as soon as subjects realise that the rule has not changed. Figure 4.3 demonstrates that this is indeed the case, with the decrease in hippocampal activation being more transient following no change than following rule change. This observation therefore supports the possibility that hippocampal responses are actively inhibited by prefrontal cortex during explicit rule induction.

A further explanation for decreased hippocampal activation following rule change is the absence of a context in which to ‘place’ or encode novel items. Once

subjects realise the rule has changed, there is no abstract structure with which to judge the category membership of novel stimuli. One view of hippocampal function (Wallenstein *et al*., 1998) is that it establishes a context on the basis of a constellation of stimuli and uses this context for integrating incoming stimuli. Rule change abolishes the current context causing hippocampal activity to decrease as exemplars cannot be encoded into a meaningful framework. It could, however, be argued that the loss of context following rule change should cause increased hippocampal activation to re-establish a context in which to encode exemplars.

The fact that medial temporal regions were not activated by rule changes agrees with previous observations that medial temporal lobe lesions do not prevent the acquisition of abstract knowledge in categorisation tasks, despite impairing memory for individual items (Knowlton and Squire, 1993). That the hippocampus is not engaged by rule changes is supported by normal performance on the WCST in patients with bilateral temporal lobectomy (Milner, 1963; see chapter 1).

4.43 Summary

The findings suggest that fronto-polar prefrontal cortex selectively mediates rule learning in a categorisation task emphasising 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. Rule change-evoked activation in FPPC most likely reflects problem solving and not simple detection of novel rules, although neuropsychological data indicate that this region may be particularly important for detecting abstract rule changes and switching behaviour accordingly (Stuss *et al*., 2000). Change in surface

features during categorisation engages left anterior hippocampus, supporting the previous proposal of novelty-evoked activation in this region. High-level novelty, that engages prefrontal problem solving operations, produces a transient decrease in hippocampal activation. For further discussion of these results, see Strange *et al*. (2001).