ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 31 (2006) 1802 – 1816

Differential hippocampal and prefrontal-striatal contributions to instance-based and rule-based learning

Christian F. Doeller,[a,b,c,*] Bertram Opitz,[c] Christoph M. Krick,[d] Axel Mecklinger,[c] and Wolfgang Reith[d]

[a]*Institute of Cognitive Neuroscience, University College London, London, UK*
[b]*Department of Anatomy and Developmental Biology, University College London, London, UK*
[c]*Experimental Neuropsychology Unit, Department of Psychology, Saarland University, Saarbrücken, Germany*
[d]*Department of Neuroradiology, Saarland University Hospital, Homburg, Germany*

Received 2 June 2005; revised 27 January 2006; accepted 3 February 2006
Available online 24 March 2006

It is a topic of current interest whether learning in humans relies on the acquisition of abstract rule knowledge (rule-based learning) or whether it depends on superficial item-specific information (instance-based learning). Here, we identified brain regions that mediate either of the two learning mechanisms by combining fMRI with an experimental protocol shown to be able to dissociate both learning mechanisms. Subjects had to learn object-position conjunctions in several trials and blocks. In a learning condition, either objects (Experiment 1) or positions (Experiment 2) were held constant within-blocks. In contrast to a control condition in which object–position conjunctions were trial-unique, a performance increase within and across-blocks was observed in the learning condition of both experiments. We hypothesized that within-block learning mainly relies on instance-based processes, whereas across-block learning might depend on rule-based mechanisms. A within-block parametric fMRI analysis revealed a learning-related increase of lateral prefrontal and striatal activity and a learning-related decrease of hippocampal activity in both experiments. By contrast, across-block learning was associated with an activation modulation in distinct prefrontal-striatal brain regions, but not in the hippocampus. These data indicate that hippocampal and prefrontal-striatal brain regions differentially contribute to instance-based and rule-based learning.
© 2006 Elsevier Inc. All rights reserved.

Keywords: fMRI; Instance-based learning; Prefrontal cortex; Rule-based learning; Striatum

Introduction

Learning regularities across multiple episodes is a core cognitive ability. Controversy currently surrounds whether humans learn the surface structure of regular input pattern based on the superficial similarity between learning instances or whether humans acquire abstract rule knowledge (cf. Pothos, 2005; Shanks and St. John, 1994; Shanks, 1995). Three main learning tasks have been used extensively in experimental psychology: artificial grammar learning, category learning, and sequence learning tasks. Results of several studies provide evidence for the notion that learning in these tasks is partly based on the knowledge of (fragments of) learning instances (Nosofsky, 1986; Perruchet, 1994). In contrast, rule-based accounts assume that subjects acquire a set of abstract rules, defining an input pattern as grammatical, as a category member, or as a regular sequence, respectively (Ashby and Perrin, 1988; Reber, 1989). Alternative views posit that learning is subserved by both instance-based and rule-based processes (Dominey et al., 1998; Erickson and Kruschke, 1998; Knowlton and Squire, 1996; Meulemans and Van der Linden, 1997; Shanks and St. John, 1994; Shanks, 1995).

Tightly coupled with the debate about instance-based vs. rule-based learning is the question which brain structures might subserve either mechanism. However, only a few studies examined the neural correlates of instance-based vs. rule-based learning. In an artificial grammar learning study, Fletcher et al. (1999) demonstrated that learning within experimental blocks is mediated by the right lateral prefrontal cortex (PFC), whereas the left lateral PFC subserves learning across the entire experiment. The authors argue that within-block learning effects mainly rely on explicit retrieval of individual items based on the surface structure of items, i.e., instance-based learning. Further, the authors propose that, in contrast, across-block learning is based on the acquisition of abstract rule knowledge. In another study (Strange et al., 2001), subjects were required to learn rules which define the category

* Corresponding author. Institute of Cognitive Neuroscience and Department of Anatomy and Developmental Biology, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK.
E-mail address: c.doeller@ucl.ac.uk (C.F. Doeller).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.02.006

membership of four-letter strings. Changes in abstract rules were associated with an increase of anterior PFC activity, whereas hippocampal activation was modulated by the introduction of new instances. Recent fMRI and patient studies revealed that the hippocampus is associated with instance-based learning (Lieberman et al., 2004; Opitz and Friederici, 2004), whereas the basal ganglia (Lieberman et al., 2004; Teichmann et al., 2005) and the lateral PFC (Opitz and Friederici, 2004) subserve rule-based learning. However, a computational model has implicated the basal ganglia in instance-based learning and the lateral PFC in rule-based learning, respectively (Dominey et al., 1998). Apart from the Dominey et al. (1998) and the Teichmann et al. (2005) study (which used letter sequences and arithmetic operations, respectively), most of the studies examined artificial grammar learning. Taken together, these studies converge to suggest that the lateral PFC subserves rule-based learning, However, the exact location of lateral PFC activation varies between studies, depending on stimulus properties and task requirements. Furthermore, most of the studies point to an involvement of the basal ganglia in rule-based learning and a specific role of the hippocampus in instance-based learning.

In the present fMRI study, we investigated this issue by adopting the experimental logic of the Fletcher et al. (1999) study. Using a slightly modified version of a recently developed learning paradigm (Doeller et al., 2005), we aimed at investigating instance-based vs. rule-based learning in a non-linguistic domain. In two experiments, subjects were required to memorize six object–position conjunctions in each trial of several experimental blocks. Both experiments included two conditions, a learning condition and a control condition. In the learning condition, either objects (Experiment 1) or positions (Experiment 2) were held constant in each trial of the experimental blocks (systematic repetition of specific objects or positions), by this introducing regularities across episodes, whereas in the control condition, object–position conjunctions were trial-unique (*no* systematic repetition of specific objects or positions). Hence, we define learning as the successful adoption of these invariant objects and positions in object–position conjunctions across trials, i.e., object or spatial regularities. We expected that the introduction of invariant objects and positions entails increased task performance across trials within-blocks of the learning condition. These *within*-block learning effects were supposed to reflect mainly instance-based learning, since subjects' judgments could rely solely on the similarity between study and test items. In contrast to our recent fMRI study (Doeller et al., 2005), trials were blocked by condition to minimize the probability that subjects based their judgement on a common strategy for both conditions. Critically, the set of invariant objects and positions changed from block to block in the learning condition. By this, subjects were able to transfer their knowledge about regularities to new instances when a new block starts. This transfer has been implicated as a possible experimental test to dissociate instance-based and rule-based learning (e.g., Gomez and Schvaneveldt, 1994; Mathews et al., 1989). If subjects acquire abstract rule knowledge, they should be able to transfer this knowledge to new instances. Thus, a performance modulation *across* learning blocks was assumed to be a main index of rule-based learning. Based on the above reviewed literature and our previous results (Doeller et al., 2005), we predicted a learning-related decrease of hippocampal and an increase of prefrontal-striatal activation as a function of learning within-blocks. In contrast, we expected prefrontal-striatal – but not hippocampal – involvement during across-block learning (cf. Fletcher et al., 1999; Lieberman et al., 2004; Opitz and Friederici, 2004).

Materials and methods

Subjects

Twenty-four subjects participated in the study, 12 subjects in Experiment 1 (aged 22–33, mean age 24.6 years, 6 females) and 12 subjects in Experiment 2 (aged 22–29, mean age 24.3 years, 5 females). All subjects were right-handed with normal or corrected-to-normal vision and were paid for participating. Informed consent was obtained before scanning. All participants reported to be in good health with no history of neurological disease.

Stimuli, task, and design

Sixteen pictures denoting real-life objects were used as stimulus material in both experiments. The stimulus set contained eight living and eight non-living objects. Stimuli were presented within a 4×4 grid. In each experimental trial (Fig. 1), six different objects were presented sequentially at six different positions (sample phase). Each object was presented for 600 ms, followed by a 100 ms interstimulus interval. After a 2000 ms delay (fixation cross), subjects were shown a probe stimulus for 1000 ms. Participants were required to indicate whether or not the current object–position conjunction (the probe stimulus) was identical to one of the six object–position conjunctions presented during the sample phase. Responses were delivered by a button press with the right or left index finger (1000 ms response window). Response-to-hand mappings were counterbalanced across subjects. Probes in each block (see below) comprised 50% old (old object at old position) and 50% new object–position conjunctions (3 equally distributed categories: old object/new position, new object/old position, and new object/new position). Visual feedback was provided for 500 ms immediately after probe offset. An exponentially distributed intertrial interval (ITI) of 2.5–7.5 s (mean: 3.5 s) was used (cf. Doeller et al., 2005, for a detailed description of the experimental procedures).

The design of both experiments included two conditions, a control condition and a learning condition. To minimize the probability that subjects based their judgment on a common strategy for both conditions, trials were blocked by condition. In contrast to our previous study (Doeller et al., 2005), in which blocks of both conditions were presented in randomized order, subjects performed four blocks in sequence in each condition in the present study. A randomized sequential presentation of blocks of the learning and the control condition (as in our previous study) seems to prevent across-block learning. The presentation of control blocks might hamper the proposed transfer mechanism from one block to the next block in the learning condition (Doeller and Opitz, 2004). In both conditions, participants were informed about the beginning of a new block. Each experimental block comprised 36 trials. The order of conditions was balanced across subjects. The 3D structural MR sequence (see below) was measured in the middle of the experiments to separate the scanning sessions of both conditions. Participants were unaware of the experimental manipulation, i.e., the existence of two different conditions. In the control condition, object–position conjunctions were unique in each trial. Here, each object–

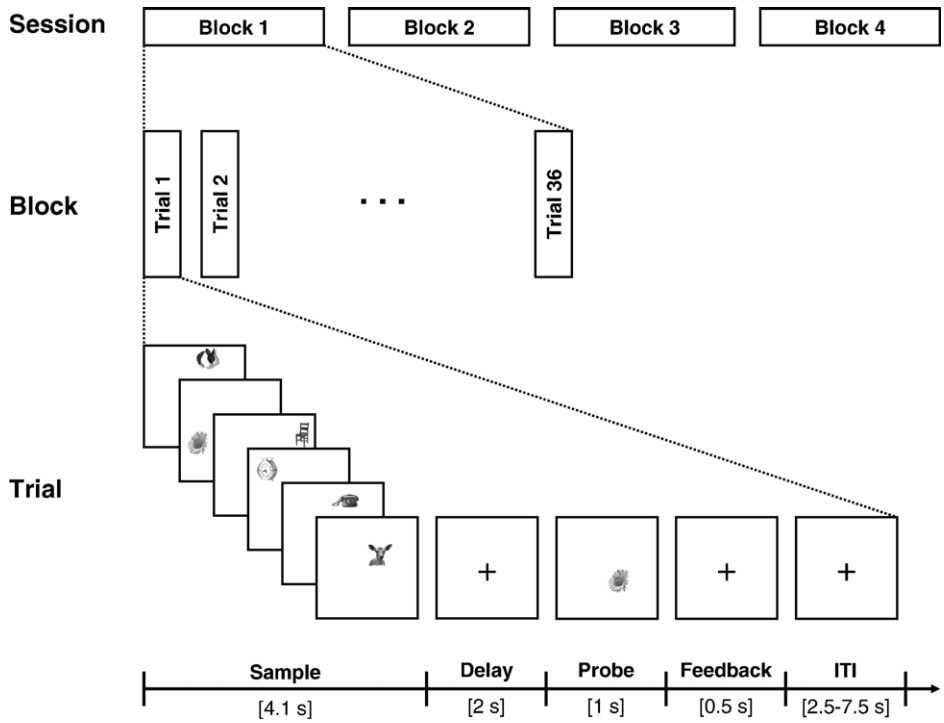

Fig. 1. Session, block, and trial structure. Subjects performed four blocks in each experimental condition. Each block comprised 36 trials. During the sample phase of the trial, six different objects were presented sequentially at six different positions within a 4 × 4 grid (grid not shown in the Figure). Upon presentation of a probe stimulus, participants were required to indicate whether the probe is identical to one of the six object position conjunctions presented during the sample phase. In the present example, the probe comprised an old object at a new position, therefore requiring a "new" response. Duration (in seconds) of the respective trial phases are depicted at the bottom of the Figure. ITI, intertrial interval.

position conjunction in each trial and each block was selected randomly from the set of objects and positions. In contrast, in the learning condition either objects (Experiment 1) or positions (Experiment 2) were held constant within each trial of the experimental blocks. In the learning condition of Experiment 1, invariant objects were presented at variable positions (object regularities). Accordingly, in the learning condition of Experiment 2, variable objects were presented at invariant positions (spatial regularities). The number of invariant objects and positions, respectively, resembled the number of object–position conjunctions presented during the sample phase, i.e., the same six objects or positions were repeated within one block: in each trial, in the learning condition of Experiment 1, the same six objects were presented in random sequence at variable positions, in each trial of one block in the learning condition of Experiment 2, variable objects were presented at the same six positions (in random sequence). The fixed configuration of invariant objects and positions changed from block to block, i.e., when a new block starts, a new set of objects or positions in object–position conjunctions was repeated. All other task parameters were held constant across conditions and experiments. In particular, the control condition was identical in Experiments 1 and 2. Subjects were given a training session immediately before the experiment, including 36 training trials.

Behavioral analysis

For each subject, Pr values (proportion hits – proportion false alarms; Feenan and Snodgrass, 1990) were used as main performance measure in both experiments. Instead of the false alarm rate to all new probes only false alarms to probes comprising new objects (Experiment 1) and new positions (Experiment 2), were included. We have previously demonstrated that learning of such object–position conjunctions is – beside an increase in hit rate – most clearly revealed by a reduction of false alarms to these specific probe types, since subjects could reject these probes solely on the basis of the knowledge of invariant objects and positions, respectively (Doeller and Opitz, 2004; Doeller et al., 2005). In the following, this performance measure will be referred to as 'adjusted Pr'. In the control condition of Experiments 1 and 2, adjusted Pr values were computed in the same manner as in the corresponding learning condition. To analyze within-block learning effects, mean performance measures in nine consecutive trials (i.e., trials 1–9, 10–18, 19–27, and 28–36) were averaged across all blocks separately for both conditions. To evaluate across-block learning effects, mean adjusted Pr values for each entire block of both conditions were averaged separately for blocks 1–2 and 3–4, respectively.

Imaging parameters

BOLD-sensitive, T_2^*-weighted functional images and T_1-weighted structural images (3D MP-RAGE) were acquired at 1.5 T (Siemens Sonata). Functional data were acquired using a gradient-echo EPI pulse sequence, with the following parameters: $T_R = 1.8$ s, $T_E = 50$ ms, flip angle = 85°, slice thickness = 4 mm, interslice gap = 1 mm, in-plane resolution = 3.5 × 3.5 mm, FoV = 224 mm^2, 20 axial slices parallel to AC–PC plane. The first four volumes were discarded to allow for T_1 equilibration. In both experiments, the two conditions were conducted in different functional sessions, separat-

ed by the structural 3D MP-RAGE sequence. An additional 2D T_1-weighted structural sequence (T_R = 600 ms, T_E = 13 ms, flip angle = 80°, slice thickness = 4 mm, interslice gap = 1 mm, in-plane resolution = 0.9 × 0.9 mm, FoV = 224 mm^2, 20 axial slices parallel to AC–PC plane) was measured in-plane with respect to the functional sequence and applied during the coregistration procedure (see below).

Imaging preprocessing

The imaging analysis was performed with SPM2 (http://www.fil.ion.ucl.ac.uk/spm/spm2.html). FMRI time series were sinc–interpolated in time to correct for differences in slice acquisition time and motion-corrected, using a 6-parameter rigid body spatial transformation and a B-spline interpolation. An additional unwarping procedure was applied to account for movement-by-magnetization inhomogeneity interactions and by this to minimize the movement-related residual variance (Andersson et al., 2001). Coregistration included two steps. First, images of the 2D T_1-weighted sequence, which was measured in-plane with respect to the EPI sequence were coregistered to the mean functional image. Second, images of the 3D MPRAGE sequence were coregistered to the images of the resliced 2D T_1-weighted sequence. Subsequently, resliced 3D anatomical images were normalized to the standard T_1 template (MNI reference brain). Based on the determined parameters, the normalization algorithm was then applied to the functional volumes. Finally, the normalized functional images were resampled into 2 mm isotropic voxels and spatially smoothed with an isotropic 7-mm FWHM Gaussian kernel.

Statistical analysis of fMRI time series

Time series were analyzed using a two-stage analysis. At a first level, for each trial of both conditions trial-specific effects were modeled using the canonical HRF (Friston et al., 1998) separately for the sample phase (epoch-related; duration = 4.1 s) and the probe phase (event-related). Additional parametric analyses were conducted (see below). Data were high-pass filtered to 1/128 Hz and scaled for global activity. Parameters for each covariate were estimated by a least-mean-squares fit of the model to the time series using a subject-specific fixed-effects model within the general linear model. During the estimation procedure, serial correlations were estimated with a restricted maximum likelihood (ReML) algorithm using an AR(1) plus white noise model. In SPM2, the ReML estimates (hyperparameters) are then used to correct for non-sphericity (Friston et al., 2002). Linear contrasts of the parameter estimates for each regressor were calculated for each subject and brought to the second level random effects analysis. All analyses were restricted to the sample phase. MNI-coordinates of all reported activations have been transformed to the canonical Talairach space (http://www.mrc-cbu.cam.ac.uk/imaging/mnispace.html).

To investigate brain regions involved during regularity learning, we conducted parametric fMRI analyses in the learning condition of both experiments to explicitly model learning-related activation patterns (cf. Doeller et al., 2005). For this purpose, individual fMRI time series (associated with each sample phase) were weighted with parametric modulation functions, separately modeling (1) within-block learning (Fig. 5), (2) across-block learning (Fig. 7), and (3) the across-block modulation of within-block learning effects (Fig. 8).

Thus, the regressors of these models reflected either learning-related activation increase or decrease within and across-blocks, respectively, and the across-block modulation of within-block learning effects (see below). The first set of model functions was derived by averaging the individual within-block learning functions (mean adjusted Pr values for nine consecutive trials, collapsed across all blocks and subjects), separately for both experiments. A second set of model functions was used to analyze across-block learning effects. Here, performance measures (adjusted Pr values) were collapsed across all trials, separately for each block and averaged across subjects (separately for each experiment). In a third analysis, we used the slope of the mean fitted learning functions separately for each entire block as a model function. For this purpose, we (i) fitted the mean learning functions (adjusted Pr values) for *each* block with a logarithmic function $y = a\ln(t_i) + b$ and (ii) calculated the slope of these functions. Mean learning functions were derived by averaging individual learning functions for each block collapsed across all 24 subjects of both experiments. We used averaged performance functions rather than individual learning functions as parametric modulation functions in all three analysis to increase the model fit by means of reducing error variance.

Areas exhibiting unspecific time effects were excluded from the analyses. To control for these effects, the respective contrasts were exclusively masked with regions showing an activation modulation as a function of time in the control conditions, i.e., activation increases or decreases within or across-blocks. These mask images in the control conditions were derived by using identical parametric modulation functions and statistical thresholds as in the analyses of the respective learning conditions.

It is important to note that this parametric imaging analysis does not allow for a direct quantitative comparison of within-block and across-block learning effects since both underlying parametric modulation functions have a completely different rate constant. However, the mass-univariate statistical approach definitely permits the conclusion that both learning effects rely on qualitatively different brain networks.

Given our a priori hypotheses, all fMRI analyses were restricted to the lateral prefrontal cortex, the striatum, and the hippocampus. In the lateral PFC we focussed on areas which have been reported in previous imaging studies on incremental learning (cf. Doeller et al., 2005; Fletcher et al., 2001, 2005; Opitz and Friederici, 2003; Schendan et al., 2003; Strange et al., 2001; Turner et al., 2004). Accordingly, prefrontal regions of interest were defined about peak locations at 30,66,4 (superior frontal gyrus), 40,11,31/−48,19,30/−40,30,24 (posterior and mid-portion middle frontal gyrus), −40,8,18/40,28,19 (inferior frontal gyrus, pars opercularis), −57,20,17/−36,40,4 (inferior frontal gyrus, pars triangularis), and 36,21,25 (inferior frontal sulcus), together with the corresponding contralateral coordinates (all coordinates in Talairach space). Based on our a priori hypothesis and in agreement with the literature, significant activated regions were identified using a statistical threshold of $P < 0.0005$ (uncorrected; cluster size: 5 contiguous voxels). An additional region-of-interest (ROI) analysis was performed for the striatum and the hippocampus to further examine our previous results of within-block learning ($P < 0.01$, small volume corrected, SVC; Worsley et al., 1996). Striatal and hippocampal ROIs were determined according to a detailed neuroanatomy atlas (Warner, 2001). Appropriate mask image was generated using MRIcro (http://www.psychology.

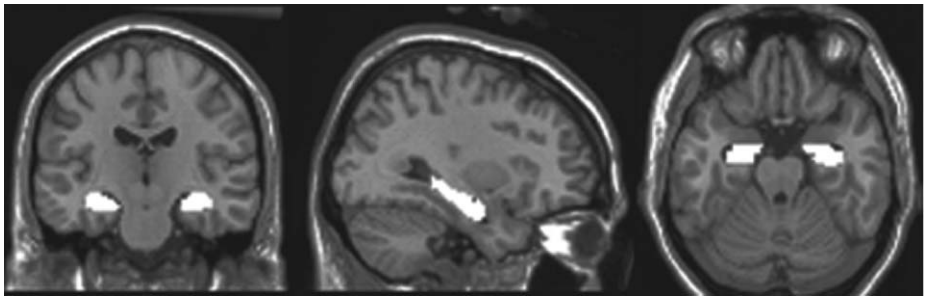

Fig. 2. The mask image defining the hippocampal region of interest is superimposed on coronal, sagittal, and axial sections (from left to right) of the MNI T_1-weighted MRI template (see Materials and methods).

nottingham.ac.uk/sta_/cr1/mricro.html) (see Fig. 2, which shows hippocampal ROIs).

Results

Behavioral results: within-block learning

In a first step, within-block learning effects were analyzed (Fig. 3). In both experiments, performance (adjusted Pr values) increased across trials in the learning condition, but not in the control condition. To examine changes of performance during the time course of the blocks, comparisons between the first (trials 1–9) and last (trials 28–36) quarter of trials within-blocks were conducted. In Experiment 1 (Fig. 3A, left panel), mean adjusted Pr values increased from the first to the last quarter of trials within-blocks in the learning condition, $F(1, 11) = 4.86$, $P < 0.05$, but not in the control condition, $F(1, 11) < 1$. Accordingly, performance was better in the learning condition as compared to the control condition in the last quarter of trials, $F(1, 11) = 16.12$, $P < 0.005$, but not in the first quarter of trials, $F(1, 11) = 2.95$, $P > 0.1$. A marginally significant linear trend across trials was restricted to the learning condition, $F(1, 11) = 4.32$, $P < 0.07$ (control condition: $F(1, 11) = 1.50$, $P > 0.2$). In Experiment 2 (Fig. 3A, right panel), an increase of adjusted Pr values between the first and last quarter of trials was observed in the learning condition, $F(1, 11) = 5.40$, $P < 0.05$, but not in the control condition: $F(1, 11) = 1.29$, $P > 0.2$. In contrast to the first quarter of trials, $F(1, 11) = 1.28$, $P > 0.2$, performance between conditions differed at a marginally significant level in the last quarter of trial, $F(1, 11) = 4.05$, $P < 0.07$. Moreover, mean adjusted Pr values increased linearly in the learning condition, $F(1, 11) = 5.70$, $P < 0.05$, but not in the control condition, $F(1, 11) = 1.38$, $P > 0.2$. Thus, these data indicate that an increase of performance across trials within-blocks was restricted to the learning condition in both experiments. However, similar comparisons for the hit rate revealed no differences between the first and last quarter of trials in both conditions (Fig. 3B), neither in Experiment 1 (learning condition: $F(1, 11) = 2.94$, $P > 0.1$; control condition: $F(1, 11) = 1.72$, $P > 0.2$) nor in Experiment 2 (learning and control condition: $F(1, 11) < 1$).

For both experiments, additional statistical analyses were conducted separately for false alarms to probes including new objects and new positions, respectively. A dissociation between both experiments was expected with respect to the two types of

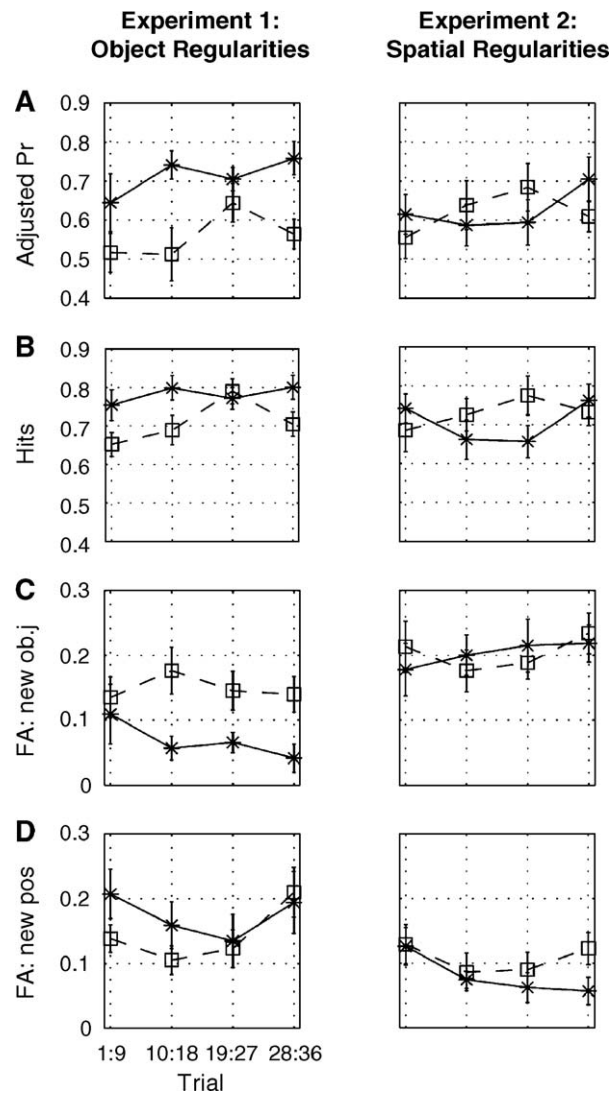

Fig. 3. Behavioral Results: Within-block Learning. (A) Mean adjusted Pr values, (B) mean hit rate, (C) mean false alarm rate for probes including new objects, and (D) mean false alarm rate for probes including new positions. The performance measures are averaged across trials 1–9, 10–18, 19–27, and 28–36, respectively, and collapsed across all blocks, separately for the learning condition (solid) and the control condition (dashed) in Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels).

false alarms. In Experiment 1 (object regularities), a reduction of false alarms to probes including new objects was predicted, whereas in Experiment 2 (spatial regularities), we hypothesized a selective reduction of false alarms to probes including new positions, since subjects could reject these specific probes solely on the basis of their knowledge of invariant objects and positions, respectively. In Experiment 1, the false alarm rate to probes including new objects (Fig. 3C, left) was significantly reduced in the learning condition relative to the control condition, $F(1, 11) = 22.71$, $P < 0.001$. Additional comparisons between both conditions of Experiment 1 revealed no differences with respect to this false alarm type in the first quarter of trials, $F(1, 11) < 1$. However, both conditions differed significantly in the last quarter of trials, $F(1, 11) = 9.12$, $P < 0.05$. By contrast, no differences between conditions were observed for false alarms to probes including new positions (Fig. 3D, left), neither for all trials, $F(1, 11) = 1.25$, $P > 0.2$, nor for the first, $F(1, 11) = 3.53$, $P > 0.08$, and the last quarter of trials, $F(1, 11) < 1$. As predicted, analyses of false alarms to probes including new objects in Experiment 2 (Fig. 3C, right) revealed no differences between both the two conditions, neither for all trials, $F(1, 11) < 1$, nor for the first, $F(1, 11) = 1.36$, $P > 0.2$, and the last quarter of trials, $F(1, 11) < 1$. In contrast, the false alarm rate to probes including new positions (Fig. 3D, right) tended to be reduced in the last quarter of trials in the learning as compared to the control condition, $F(1, 11) = 4.49$, $P < 0.06$. No differences between conditions were observed for the first quarter of trials, $F(1, 11) < 1$. Accordingly, false alarms of this type significantly decreased between the first and last quarter of trials selectively in the learning condition, $F(1, 11) = 5.71$, $P < 0.05$ (control condition: $F(1, 11) < 1$). Furthermore, a significant linear trend across trials was restricted to the learning condition, $F(1, 11) = 5.76$, $P < 0.05$ (control condition: $F(1, 11) < 1$).

Behavioral results: across-block learning

In a second analysis step, across-block learning effects were examined (Fig. 4). Similar to the within-block analysis, a performance increase across-blocks was restricted to the learning condition in both experiments. This observation was confirmed by a comparison between mean adjusted Pr values for blocks 1–2 vs. 3–4, collapsed across all trials within-blocks. In both experiments, mean adjusted Pr values increased across-blocks in the learning condition (Experiment 1: $F(1, 11) = 13.90$, $P < 0.005$, Fig. 4A, left panel; Experiment 2: $F(1, 11) = 6.50$, $P < 0.05$, Fig. 4A, right panel), but not in the control condition ($F(1, 11) < 1$ in both experiments). A similar pattern was observed for the hit rate: In both experiments the mean hit rate was higher in blocks 3–4 as compared to blocks 1–2 in the learning condition (Experiment 1: $F(1, 11) = 8.10$, $P < s0.05$; Experiment 2: $F(1, 11) = 6.20$, $P < 0.05$), but not in the control condition (Experiment 1: $F(1, 11) < 1$; Experiment 2: $F(1, 11) = 1.59$, $P > 0.2$).

Similar to the within-block learning effects, we conducted additional analyses of false alarms to probes including new objects and new positions, respectively. In Experiment 1, the ANOVA revealed a main effect of condition for probes including new objects, $F(1, 11) = 22.70$, $P < 0.001$ (Fig. 4C, left). Accordingly, false alarms of this type were reduced in blocks 1–2, $F(1, 11) = 10.03$, $P < 0.01$, and blocks 3–4, $F(1, 11) = 14.43$, $P < 0.005$, in the learning relative to the control condition. The false alarm rate to probes including new positions (Fig. 4D, left) did not differ between conditions, $F(1, 11) = 1.25$, $P > 0.2$, neither in blocks 1–2, $F(1, 11) < 1$, nor in blocks 3–4, $F(1, 11) < 1$. In accordance with the results of the within-block learning effects, a dissociation between experiments was observed.

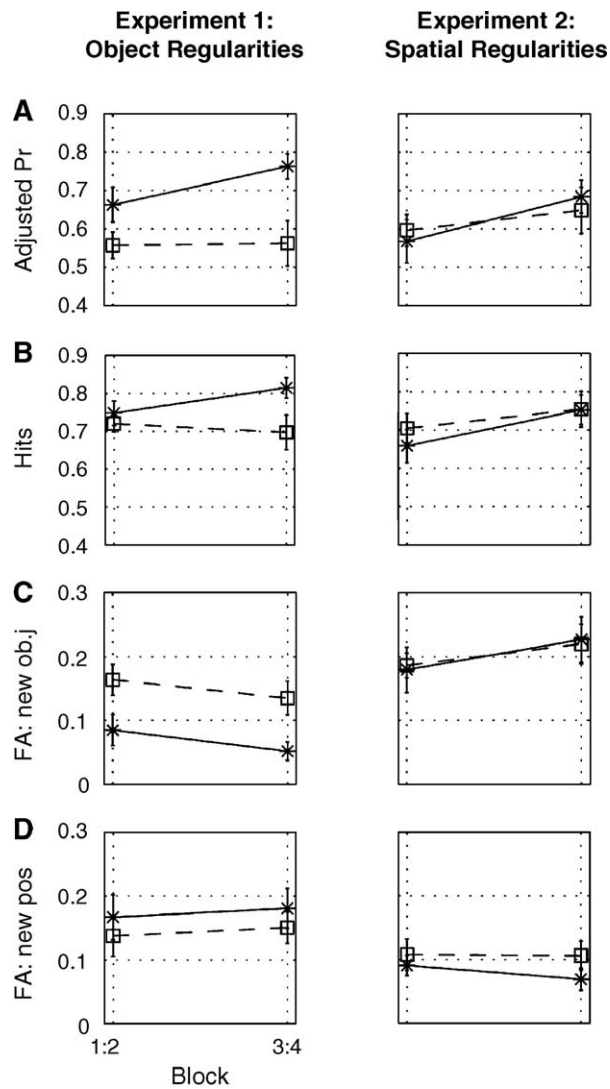

Fig. 4. Behavioral Results: Across-block Learning. (A) Mean adjusted Pr values, (B) mean hit rate, (C) mean false alarm rate for probes including new objects, and (D) mean false alarm rate for probes including new positions. The performance measures are averaged across-blocks 1–2 and 3–4, respectively and collapsed across all trials, separately for the learning condition (solid) and the control condition (dashed) in Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels).

In Experiment 2, the false alarm rate to probes including new objects (Fig. 4C, right) was similar in the learning and control condition ($F(1, 11) < 1$, for the analysis of all blocks and separate analyses for blocks 1–2 and 3–4, respectively). By contrast, subjects showed a reduced false alarm rate to probes including new positions (Fig. 4D, right) in the learning relative to the control condition (all blocks: $F(1, 11) = 7.85$, $P < 0.05$, blocks 1–2: $F(1, 11) = 1.91$, $P > 0.1$; blocks 3–4: $F(1, 11) = 5.87$, $P < 0.05$).

To evaluate the across-block modulation of the within-block learning effects, we calculated performance differences (adjusted Pr values) between the last quarter of blocks and the first quarter of blocks, separately for blocks 1–2 and blocks 3–4. These

difference values were considered as a measure of learning within-blocks. Difference values entered into an ANOVA with the factors condition (control vs. learning) and block (difference values for blocks 1–2 and blocks 3–4, respectively). In this analysis we were mainly interested in rule-based mechanism which were assumed to be similar in both the object and the spatial domain. In addition, given the lower number of datapoints in this analysis compared to the above analyses, we increased the power of this analysis by including data from Experiments 1 and 2. This analysis revealed a marginally significant condition × block interaction, $F(1, 23) = 4.19$, $P < 0.06$ (learning condition: 0.14 ± 0.05 [block 1–2]; 0.07 ± 0.06 [block 3–4]; control condition: −0.03 ± 0.05 [block 1–2]; 0.13 ± 0.06 [block 3–4]). Moreover, difference values were higher in the learning relative to the control condition in blocks 1–2, $F(1, 23) = 4.28$, $P < 0.05$, but not in blocks 3–4, $F(1, 23) < 1$, indicating that the performance increase within-blocks in the learning condition was attenuated across-blocks relative to the control condition.

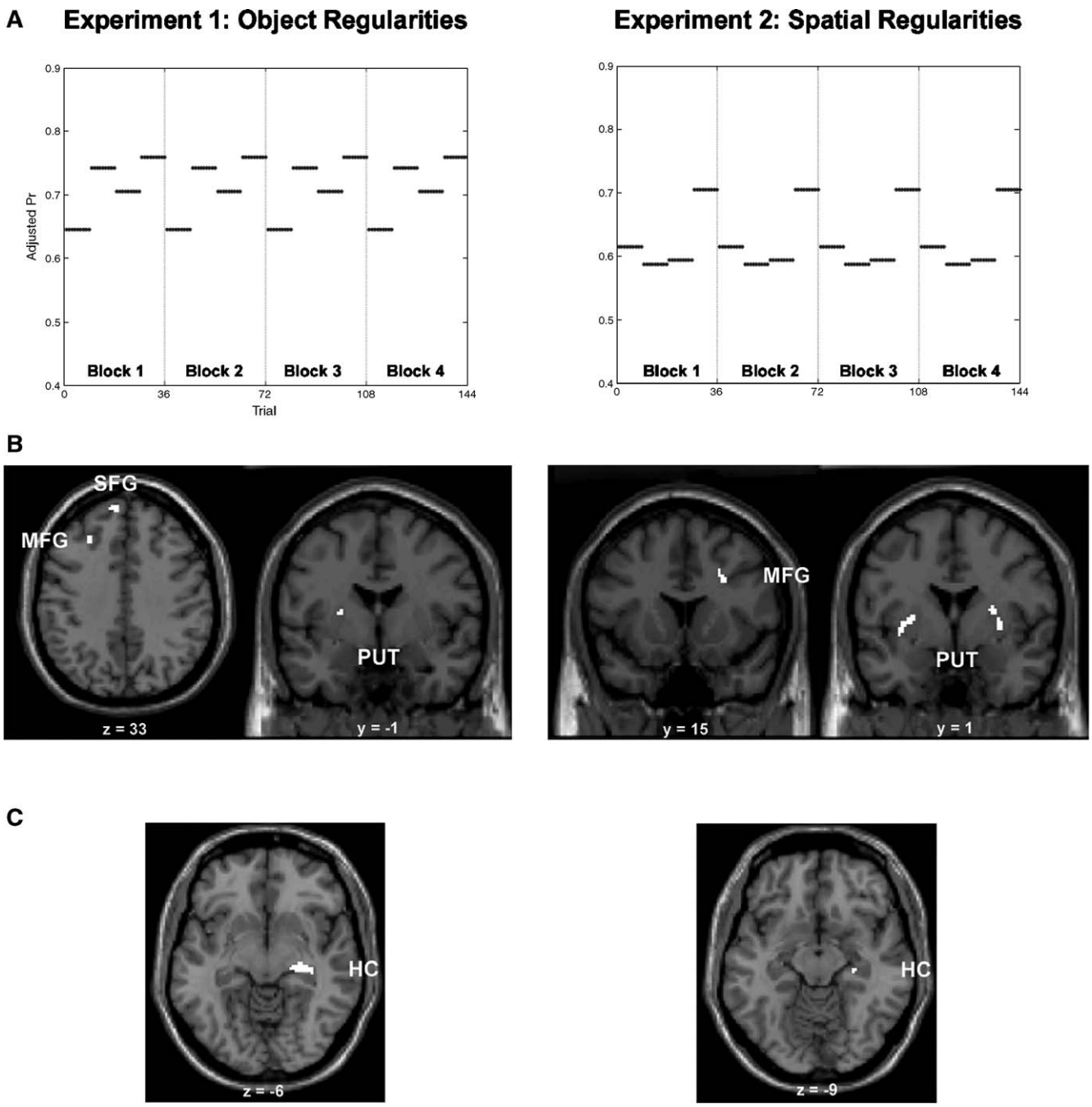

Fig. 5. Imaging results: within-block learning. Parametric modulation functions and imaging results are depicted separately for Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels). (A) Parametric modulation functions for block 1 (trials 1–36), block 2 (trials 37–72), block 3 (trials 73–108), and block 4 (trials 109–144). (B) Learning-related increase of activity within-blocks in the superior frontal gyrus (SFG), middle frontal gyrus (MFG), and putamen (PUT). (C) Learning-related decrease of activity within-blocks in the hippocampus (HC). SPMs are superimposed on coronal or axial sections of the MNI T_1-weighted MRI template. Talairach y coordinate or z coordinate is given below each image.

Table 1
Imaging results: within-block learning

Region	BA	Hemisphere	Talairach			Z score
			x	y	z	
Experiment 1: object regularities						
(1) Learning-related increase						
Superior frontal gyrus (SFG)	9	L	−8	54	32	3.89
Middle frontal gyrus (MFG)	8	L	−24	31	33	4.36
Putamen (PUT)	–	L	−24	−1	9	*2.66
(2) Learning-related decrease						
Hippocampus (HC)	–	R	26	−22	−6	*3.87
Experiment 2: spatial regularities						
(1) Learning-related increase						
Middle frontal gyrus (MFG)	8	R	24	15	32	3.69
Putamen (PUT)	–	L	−26	4	3	*3.23
	–	R	26	1	11	*2.70
(2) Learning-related decrease						
Hippocampus (HC)	–	R	28	−24	−9	*2.71

Brain regions showing a learning-related activation pattern within-blocks, separately for Experiments 1 and 2 (from anterior to posterior), described in terms of Brodmann areas (BA), hemisphere (L, left; R, right), Talairach coordinates (mm; transformed from the MNI-space), and peak Z score. SPMs were thresholded at $P < 0.0005$ (uncorrected), 5 voxel extent, except for $*P < 0.01$ (small volume corrected). Contrasts were exclusively masked with regions showing time effects (activation increase or decrease within-blocks) in the control condition of both experiments.

Imaging results: within-block learning

In a first step, within-block learning was evaluated in a parametric fMRI analysis (Fig. 5; Table 1). For this purpose, fMRI time series in the learning condition were weighted with mean within-block learning functions, separately for both experiments. Contrasts in this and the subsequent across-block fMRI analysis were exclusively masked with regions showing time effects (activation increase or decrease within- and across-blocks, respectively) in the control condition of both experiments (see Materials and methods). In Experiment 1, a learning-related increase of activity within-blocks was observed in the left superior frontal gyrus (SFG; Brodmann Area [BA] 9; peak Talairach coordinates x, y, z: −8, 54, 32) and the left middle frontal gyrus (MFG; BA 8, −24, 31, 33). Furthermore, the left putamen (−24, −1, 9) showed a learning-related increase within-blocks. A similar prefrontal-striatal pattern was observed in Experiment 2. Right MFG (BA 8: 24, 15, 32) and bilateral putamen (left: −26, 4, 3; right: 26, 1, 11) activity increased as a function of learning within-blocks. By contrast, the right hippocampus showed a learning-related decrease of activation within-blocks in both experiments (Experiment 1: 26, −22, −6; Experiment 2: 28, −24, −9). Examples of prefrontal, striatal, and hippocampal BOLD responses are provided in Fig. 6.

Imaging results: across-block learning

In a second analysis step, the neural correlates of across-block learning were investigated (Fig. 7; Table 2). Mean performance functions of each entire block were used as parametric modulation functions in the fMRI analysis (see Materials and methods). This analysis revealed a learning-related increase of activity across-blocks in the left anterior MFG (BA 46; −22, 43, 2) in Experiment 1 and in the left posterior MFG (BA 8; −30, 9, 35) in Experiment 2. In contrast, the left putamen showed a learning-related decrease of activity in Experiment 1 (−18, 13, −9) and Experiment 2 (−20, 12, −1). No learning-related activation pattern across-blocks could be observed in the hippocampus (at a lowered threshold of $P < 0.01$, SVC), neither in Experiment 1 nor in Experiment 2. Examples of prefrontal and striatal BOLD responses are reported in Fig. 9.

Finally, modulations of the within-block learning effects across-blocks were analyzed in a separate model, using the slope of the

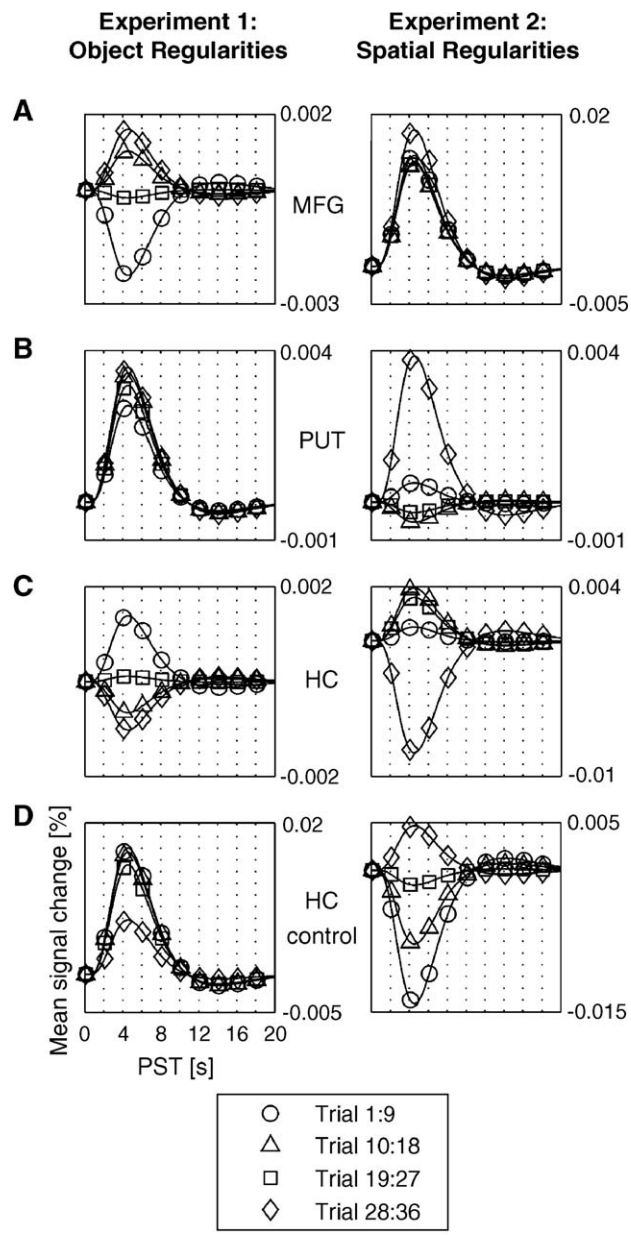

Fig. 6. Imaging Results: Within-block Learning. Time courses of the best-fitting parametric BOLD response (from the within-block analysis, see Fig. 5) for two representative subjects are depicted separately for Experiment 1 (object regularities; left panels; subject #HK) and Experiment 2 (spatial regularities; right panels; subject #LD) for (A) the middle frontal gyrus (MFG), (B) the putamen (PUT), and the right hippocampus (HC) in the (C) learning and the (D) control condition. Responses are plotted separately for trials 1–9, trials 10–18, trials 19–27, and trials 28–36, relative to grand mean over voxels, and time-locked to sample phase onset against post-stimulus time (PST).

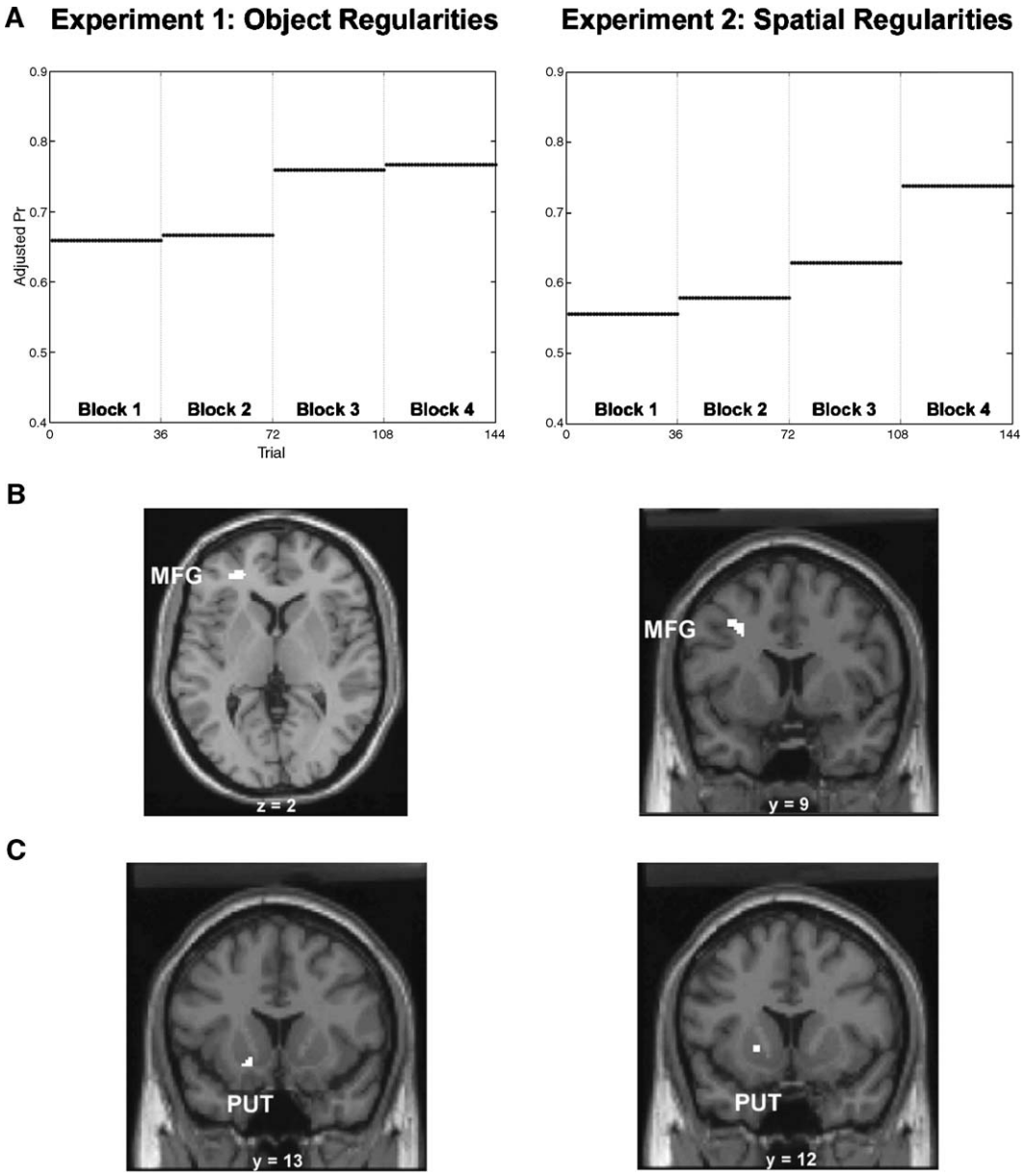

Fig. 7. Imaging results: across-block learning. Parametric modulation functions and imaging results are depicted separately for Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels). (A) Parametric modulation functions separately for blocks 1–4. (B) Learning-related increase of activity across-blocks in the middle frontal gyrus (MFG). (C) Learning-related decrease of activity across-blocks in the putamen (PUT). See Fig. 5 for details.

mean fitted learning function separately for each of the four blocks as model function (collapsed across both experiments; see Materials and methods; Figs. 8 and 9; Table 2). The mean slope decreased across-blocks and we suppose that this gradual decrease of the slope reflects the across-block modulation of the within-block learning effect. This analysis revealed activations in the left SFG (BA 10; −28, 53, 14), the bilateral putamen (left: −22, 11, −4; right: 24, 13, −6), and the left caudate nucleus (−14, 8, 5).

Discussion

The present study aimed at specifying the neural correlates of instance-based and rule-based learning. The first learning-process should be mainly reflected in a performance increase within experimental blocks (within-block learning), whereas the latter learning process should predominantly entail a performance increase across learning blocks (across-block learning). In Experiment 1 and Experiment 2, subjects had to learn object regularities and spatial regularities, respectively. Our behavioral analyses revealed within-block and across-block learning effects in both experiments. In both experiments, within-block learning was associated with a decrease of hippocampal and an increase of middle/superior frontal gyrus and putamen activity, whereas across-block learning was reflected in an increase of middle frontal gyrus and an attenuation of putamen activation. Based on these results, we argue that the introduction of object and spatial regularities entails reduced relational binding requirements, since

Table 2
Imaging results: across-block learning

Region	BA	Hemisphere	Talairach			Z score
			x	y	z	
Experiment 1: object regularities						
(1) Learning-related increase						
Middle frontal gyrus (MFG)	46	L	−22	43	2	4.46
(2) Learning-related decrease						
Putamen (PUT)	–	L	−18	13	−9	3.60
Experiment 2: spatial regularities						
(1) Learning-related increase						
Middle frontal gyrus (MFG)	8	L	−30	9	35	4.94
(2) Learning-related decrease						
Putamen (PUT)	–	L	−20	12	−1	3.53
Across-block modulation of within-block learning effects						
Superior frontal gyrus (SFG)	10	L	−28	53	14	3.72
Putamen (PUT)	–	R	24	13	−6	4.87
	–	L	−22	11	−4	3.81
	–	L	−16	9	−11	3.89
Caudate nucleus (CN)	–	L	−14	8	5	4.38

Brain regions showing a learning-related activation pattern across-blocks and an across-block modulation of within-block learning effects (for further details see Table 1 legend). SPMs were thresholded at $P < 0.0005$ (uncorrected), 5 voxel extent. Contrasts were exclusively masked with regions showing time effects in the control condition of both experiments.

invariant objects are bound to variable positions or vice versa. Furthermore, we propose that prefrontal-striatal brain regions mainly contribute to rule-based learning, whereas the hippocampus solely subserves instance-based learning.

In both experiments, performance (Pr values) increased within and across-blocks. Contrary, in the control condition performance remained constant within and across-blocks. Furthermore, the analysis of false alarms revealed a dissociation between both experiments. During learning object regularities (Experiment 1), solely false alarms to probes including new objects decreased within and across-blocks, whereas in Experiment 2 (spatial regularities), this selective decrease of false alarms was restricted to probes including new positions. Here, subjects benefit most of all from regularity learning since they can reject the probes solely on the basis of the knowledge of invariant objects and positions, respectively. A possible explanation for this differential domain-specific reduction of false alarms might be an enhanced selectivity of object and spatial representations maintained in memory, respectively (cf. Yeshurun and Carrasco, 1998). However, these learning effects were more pronounced in Experiment 1 as compared to Experiment 2. This might be due to the high distinctiveness of the objects used in this study which possibly enhances processing of objects relative to positions. By this the extraction (i.e., the detection and representation of the set of invariant [across multiple occurrences] episodic features [objects/positions] in the learning condition) of invariant objects is facilitated relative to the extraction of invariant positions, which might entail more robust learning effects in the object than in the spatial domain (cf. Doeller and Opitz, 2004).

However, in contrast to our previous study (Doeller et al., 2005), we could not find a significant increase of hit rates in both learning conditions within-blocks. Thus, in contrast to reduced relational binding, an alternative explanation for the increased performance in the present experiment might be enhanced item memory (memory for single objects in Experiment 1 and memory for single positions in Experiment 2 rather than memory for object–position conjunctions). If subjects would solely rely on the knowledge of the invariant stimulus feature irrespective of the second, variable stimulus feature, an item memory explanation would predict an increase of false alarms for probes comprising an old object at a new position (Experiment 1) and for probes comprising a new object at an old position (Experiment 2). However, a post hoc analysis revealed no differences of the false alarm rate for these specific probe types between the first and the last quarter of blocks in both experiments (Experiment 1: $F(1, 11) < 1$; Experiment 2: $F(1, 11) = 2.05$, $P > 0.10$). These data support our view that improved performance is mainly due to reduced binding requirements rather than an increase of pure item memory.

Hippocampal contribution to instance-based learning

The learning–related decrease of hippocampal activity within-blocks in Experiments 1 and 2 is consistent with our previous study (Doeller et al., 2005) as well as with studies showing hippocampal decrease during other types of gradual/incremental learning, e.g., sequence learning (Schendan et al., 2003), probabilistic classification learning (Poldrack et al., 2001; Seger and Cincotta, 2005), and artificial grammar learning (Opitz and Friederici, 2003, 2004; Strange et al., 2001). In accordance with models of hippocampal

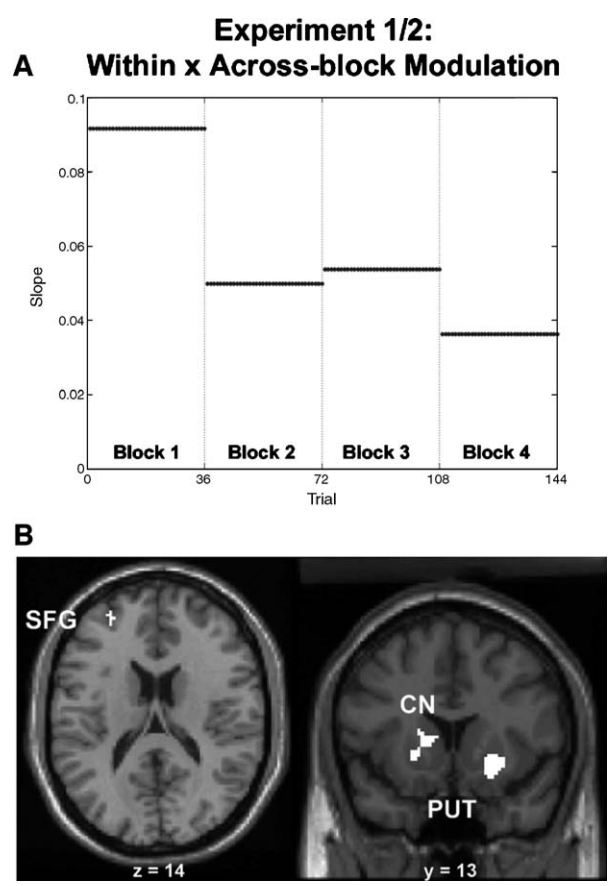

Fig. 8. Imaging Results: Across-block Modulation of Within-block Learning effects. (A) Model functions separately for blocks 1–4. (B) Slope-related activity in the SFG, the putamen, and the caudate nucleus (CN). See Materials and methods and Fig. 5 for details.

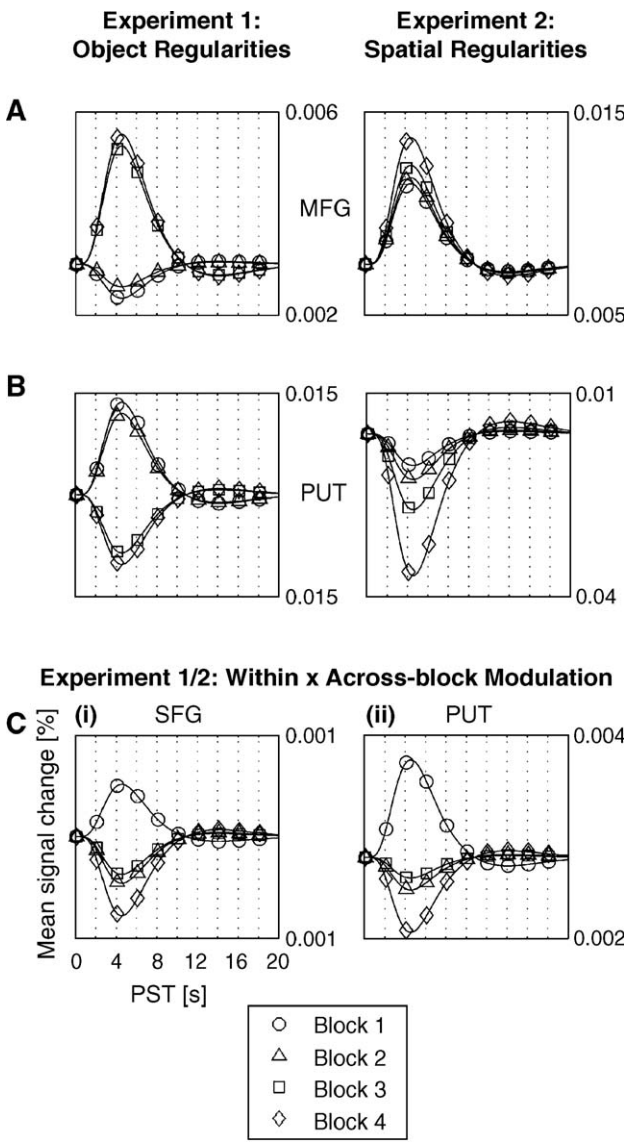

Fig. 9. Imaging Results: Within-block Learning. (A)–(B): Time courses of the best-fitting parametric BOLD response from the across-block analysis (Fig. 7) for two representative subjects are depicted separately for Experiment 1 (object regularities; left panels; subject #HK) and Experiment 2 (spatial regularities; right panels; subject #LD) for (A) the middle frontal gyrus (MFG), and (B) the putamen (PUT). In (C) responses from the slope-related analysis (Fig. 8) are depicted separately for the superior frontal gyrus (SFG; left) and the putamen (PUT; right). Responses are plotted separately for blocks 1–4, relative to grand mean over voxels, and timelocked to sample phase onset against post-stimulus time (PST).

function (Eichenbaum, 2000; Norman and O'Reilly, 2003) we argue that the activation *decrease* is due to the gradual reduction of relational binding requirements during learning, since invariant objects have to be bound to variable positions (Experiment 1) or variable objects have to be bound to invariant positions (Experiment 2), respectively. By contrast, in the control condition, variable objects have to be bound to variable positions in each trial leading to constant, high binding requirements.

As a consequence, hippocampal involvement diminishes as a function of learning. The behavioral data support this binding hypothesis rather than a pure item memory explanation. Moreover, our view is in agreement with studies on the medial temporal lobe which indicate that the rhinal cortex – and not the hippocampus – is involved in item memory (Brown and Aggleton, 2001; Davachi and Wagner, 2002).

In fact the hippocampus supports associative or relational memory (Cohen et al., 1999), i.e., binding of different features into a coherent representation of an episode. Binding is a general mechanism, which seems to be important in different cognitive domains (Treisman, 1999; Zimmer et al., in press), with the hippocampus playing a critical role during binding in episodic memory (Cohen et al., 1999; Ryan et al., 2000) and language processing (Meyer et al., 2005; Opitz and Friederici, 2004). Furthermore, it is conceivable that hippocampal involvement in working memory, especially for novel, complex information (Ranganath and D'Esposito, 2001; Ranganath et al., 2005) is due to binding of to be remembered features in working-memory which supports successful long-term memory formation (cf. Ranganath and Blumenfeld, 2005). Thus, the learning-related hippocampal activation pattern in our working-memory task might reflect the reduced working memory binding operations, possibly in the service of long-term memory formation.

Interestingly, in the two aforementioned studies (Opitz and Friederici, 2004; Strange et al., 2001) the introduction of new exemplars (Strange et al., 2001) and changes of superficial features of sentences (Opitz and Friederici, 2004) entailed a phasic increase of hippocampal activation which rapidly attenuated as a function of learning (cf. Strange et al., 1999). In contrast, in both studies hippocampal activation was not affected by the introduction of new grammatical rules. In line with these findings, hippocampal activity decreased as a function of learning within-blocks but not across-blocks in the present study. At the beginning of each new block, the set of invariant objects and positions was changed. Consequently, these changes of superficial features, i.e., instances gave rise to hippocampal involvement at the beginning of the blocks with its activity gradually decreasing as a function of increasing performance within-blocks. Critically, the across-block performance increase, which is supposed to be mainly based on strategic rule-based mechanisms, especially the transfer of knowledge about the existence of regularities to new instances in a new block, did not give rise to a modulation of hippocampal activation across-blocks. Based on these data and in line with previous imaging studies (Lieberman et al., 2004; Opitz and Friederici, 2004; Strange et al., 2001) we propose that the hippocampus contributes to learning mainly by means of instance-based processes.

However, one objection against our interpretation of the learning-related activation decrease in the hippocampus could be that the hippocampal pattern might be related to increased predictability of events in the learning relative to the control condition. For instance, using a choice reaction time task where subjects were required to respond to target stimuli consisting of color-shape conjunctions, Strange et al. (2005) demonstrated that hippocampal activation was modulated by the entropy of a stimulus sequence, i.e., the predictability of events. This is a novel and interesting finding, since predictability-related brain signals (in conditioning or causal associative learning experiments) have been consistently observed in the striatum (e.g., McClure et al., 2003; O'Doherty et al., 2003) and to a lesser extent in the prefrontal cortex (Fletcher et al., 2001; Turner et al., 2004). However, in contrast to the Strange et al. (2005) study, where the authors used four different event types (four possible conjunctions of two colors and two shapes), our object–position conjunctions in each trial

were selected from a much larger set of 16 possible objects/16 possible positions (control condition), 6 objects/16 positions (learning condition Experiment 1), and 16 objects/6 positions (learning condition Experiment 2), respectively. Given this huge number of possible object–position conjunctions in both the control condition ($n = 256$) and the learning condition ($n = 96$) the probability of occurrence of a specific object–position conjunction in one block is very small. On average, even in the learning condition a singular object–position conjunction is presented only 2.25 times in one block (0.85 times in the control condition). This makes it rather unlikely that differences in the predictability of events account for the hippocampal activation pattern within-blocks. However, further studies are required to disentangle the specific role of the hippocampus in memory binding and encoding the predictability of events.

Prefrontal–striatal contributions to rule-based learning

Distinct lateral prefrontal brain regions exhibited a learning-related modulation of activity within and across-blocks in both experiments. In showing a prefrontal activation increase as a function of learning within-blocks in both experiments, we could replicate our previous findings (Doeller et al., 2005). However, it should be noted that the prefrontal activation pattern in our previous work was slightly more lateral than the present PFC activations. This might be due to task differences. In the present study, both conditions were presented in separate sessions. By contrast, in our previous study, blocks of both conditions were presented in randomized order. Furthermore, in contrast to our previous study, in the present experiment subjects were not informed about the presence of task relevant regularities. Thus, these task changes might have entailed marginally different within-block PFC activations.

We argue that the lateral PFC is involved in (i) the extraction of task-relevant regularities and (ii) the adjustment of ongoing processing of stimuli on the basis of these extracted regularities, which might be represented in the PFC by means of mnemonic codes (Becker and Lim, 2003). It is supposed that these mnemonic codes facilitate encoding of new object-position conjunctions and give rise to increased performance during the time course of learning by providing a more efficient encoding strategy. The present results indicate that separable lateral prefrontal regions exhibited a learning-related activation pattern within- and across-blocks. In our previous work we pointed out that the extraction mechanism is most dominant during an early stage of the learning process, whereas the adjustment mechanism gradually evolves during learning, since the development of mnemonic codes relies on the extracted regularities. In the present study, it is conceivable that the former mechanism is most prevalent *within* learning blocks, especially within the early learning blocks, whereas the latter mechanism gains more influence during later blocks, i.e., evolves across learning blocks.

Furthermore, we argue that this within-block extraction process is supported by integrative working memory operations, which might be important for the integration of the invariant object and spatial features, respectively. This is in accordance with two fMRI studies showing higher lateral PFC activity associated with the integration of items into higher-level chunks (Bor et al., 2003) and the integration of verbal and spatial information (relative to the separate representation of both types of information; Prabhakaran et al., 2000).

An alternative interpretation for the observed prefrontal results within-blocks remains possible. Relying solely on the knowledge about the set of invariant objects or positions while performing the memory task might lead to a false response in about 17% of the possible probes, namely those composed of old objects at new positions (Experiment 1) and of new objects at old positions (Experiment 2), respectively. In the case of these specific probe types, the lateral PFC might be involved in interference suppression and resolution (cf. Bunge et al., 2001; Miller et al., 1996; Sakai et al., 2002).

The lateral PFC has been implicated in various kinds of gradual learning (Fletcher et al., 2001; Opitz and Friederici, 2003; Schendan et al., 2003; Strange et al., 2001; Turner et al., 2004). Furthermore, the lateral PFC is involved in the acquisition and the implementation of task-relevant rules (Freedman et al., 2001; Rougier et al., 2005; Wallis et al., 2001). It is conceivable that the proposed mnemonic codes (Becker and Lim, 2003) provide rule-like aggregated representations of extracted regularities of invariant stimulus features, most pronounced during later stages of learning. The across-block modulation of middle frontal gyrus activity might indicate that such rule-like representations are flexibly adapted to new task requirements, such as a new set of invariant stimulus features at the beginning of a new block (cf. Duncan, 2001), and hence producing "higher levels of generalization" (Rougier et al., 2005, p. 7343) as learning proceeds. Alternatively, across-block learning (and the prefrontal activation increase across-blocks) might reflect an incremental process of organization of the input pattern (Jamieson and Mewhort, 2005). Fletcher et al. (1999) observed a virtually identical increase of left middle frontal gyrus activity across-blocks during artificial grammar learning. The authors argue that this incremental activation pattern reflects the "implementation of semantic knowledge" (Fletcher et al., 1999, p. 176) of abstract grammar rules. In accordance with this view, we argue that rule-like representations of task-relevant regularities in object–position conjunctions gradually evolve across-blocks, since information is aggregated across multiple episodes (cf. Rougier et al., 2005). This left-lateralized PFC activation increase across-blocks is also in line with the view that the left PFC is involved in semantic encoding operations (Gazzaniga, 2000; Habib et al., 2003). Furthermore, the inferior–superior gradient of MFG activation for learning invariant stimulus features in the object (inferior portion of the MFG) and spatial domain (superior portion of the MFG) is in accordance with previous memory models of the PFC (Goldman-Rakic, 1996).

Similar to our previous study, the putamen exhibited a learning-related activation increase within-blocks. In addition, we observed a learning-related activation decrease across-blocks in this brain region. Furthermore, putamen and caudate nucleus activity correlated with the across-block attenuation of the slope of the within-block learning functions. Within learning blocks, we observed a complementary activation pattern in the hippocampus and the striatum. This differential involvement of the hippocampus and the striatum is in accordance with the view that the hippocampus and the basal ganglia comprise different memory systems. In contrast to the hippocampus which supports declarative memory, the striatum has been associated with nondeclarative learning and memory operations, namely incremental or habit learning (Knowlton et al., 1996; Packard and Knowlton, 2002; Poldrack et al., 2001; Squire and Zola, 1996).

In addition, the striatum has been implicated in stimulus–reward and stimulus–stimulus learning by representing predictions

of future reward delivery (O'Doherty et al., 2003; Schultz, 2002). Based on this evidence and assumptions of formal learning theories (cf. Sutton and Barto, 1990), the learning-related increase of striatal activity within-blocks might indicate that the striatum codes the increased predictability of invariant stimulus features and by this implements a statistical model of the task environment (Dayan et al., 2000). The across-block attenuation of these striatal processes might reflect the facilitated update of this statistical model.

Our behavioral analyses point to an interaction of within- and across-block learning. However, this might be partially due to an increase of general skills in the control condition. On a neural level, this interaction was reflected in an attenuation of superior frontal gyrus and putamen/caudate activity (slope-related analysis). Taken together, these data indicate that both learning processes are – at least to a certain extent – interdependent, i.e., within-block learning effects are attenuated across-blocks. The maximum of performance within-blocks seems to be reached earlier in the last as compared to the first blocks, i.e., within-block learning gets faster in the final phase of the experiment (cf. Fletcher et al., 1999, 2001), where predictability of regularities increases and by this uncertainty diminishes faster (cf. Rescorla and Wagner, 1972; Sutton and Barto, 1990).

Finally, it could be speculated that there is a cooperative functional interaction between lateral prefrontal and striatal structures during learning within and across-blocks in the present task, possibly with different time courses in either structure (cf. Pasupathy and Miller, 2005). Recent computational models (Atallah et al., 2004; Braver and Cohen, 2000; Frank et al., 2001) support this view by showing a crucial role of the striatum in providing a gating mechanism which controls the flexible update of prefrontal representations by means of a reinforcement signal. Thus, striatal reinforcement signals, coding the predictive relationships among invariant stimulus features might be involved in updating information about regularities maintained in the PFC, hence stabilizing prefrontal rule-like representations (cf. Rougier et al., 2005).

In summary, the present results suggest that instance-based and rule-based learning presumably depend on different brain regions. The hippocampus seems to be solely involved during instance-based learning, whereas distinct lateral prefrontal and striatal structures mainly subserve rule-based learning.

Acknowledgments

This work was supported by grants from the German Research Foundation, DFG, Research Group FOR-448. The authors thank Lea Meyer and Henning Loebbecke for their assistance during data acquisition and Ali Jeewajee for helpful comments on a previous version of the manuscript.

References

Andersson, J.L.R., Hutton, C., Ashburner, J., Turner, R., Friston, K., 2001. Modeling geometric deformations in EPI time series. NeuroImage 13, 903–919.

Ashby, F.G., Perrin, N.A., 1988. Toward a unified theory of similarity and recognition. Psychol. Rev. 95, 124–150.

Atallah, H.E., Frank, M.J., O'Reilly, R., 2004. Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. Neurobiol. Learn. Mem. 82, 253–267.

Becker, S., Lim, J., 2003. A computational model of prefrontal control in free recall: strategic memory use in the California verbal memory task. J. Cogn. Neurosci. 15, 821–832.

Bor, D., Duncan, J., Wiseman, R.J., Owen, A.M., 2003. Encoding strategies dissociate prefrontal activity from working memory demand. Neuron 37, 361–367.

Braver, T.S., Cohen, J.D., 2000. On the control of control: the role of dopamine in regulating prefrontal function and working memory. In: Monsell, S., Driver, J. (Eds.), In Control of Cognitive Processes: Attention And Performance XVIII. MIT Press, Cambridge, pp. 713–737.

Brown, M.W., Aggleton, J.P., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat. Rev., Neurosci. 2, 51–61.

Bunge, S.A., Ochsner, K.N., Desmond, J.E., Glover, G.H., Gabrieli, J.D.E., 2001. Prefrontal regions involved in keeping information in and out of mind. Brain 124, 2074–2086.

Cohen, N.J., 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.

Davachi, L., Wagner, A.D., 2002. Hippocampal contributions to episodic encoding: insights from relational and item-based learning. J. Neurophysiol. 88, 982–990.

Dayan, P., Kakade, S., Montague, P.R., 2000. Learning and selective attention. Nat. Neurosci. 3, 1218–1223.

Doeller, C.F., Opitz, B., 2004. Cognitive and neural mechanisms subserving regularity learning. In: Mecklinger, A., Zimmer, H., Lindenberger, U. (Eds.), Bound in Memory: Insights from Behavioral and Neuropsychological Studies. Shaker Verlag, Aachen, pp. 129–157.

Doeller, C.F., Opitz, B., Krick, C.M., Mecklinger, A., Reith, W., 2005. Prefrontal–hippocampal dynamics involved in learning regularities across episodes. Cereb. Cortex 15, 1123–1133.

Dominey, P.F., Lelekov, T., Ventre-Dominey, J., Jaennerod, M., 1998. Dissociable processes for learning the surface structure and abstract structure of sensorimotor sequences. J. Cogn. Neurosci. 10, 734–751.

Duncan, J., 2001. An adaptive coding model of neural function in prefrontal cortex. Nat. Rev., Neurosci. 2, 820–829.

Eichenbaum, H., 2000. A cortical–hippocampal system for declarative memory. Nat. Rev., Neurosci. 1, 41–50.

Erickson, M.A., Kruschke, J.K., 1998. Rules and exemplars in category learning. J. Exp. Psychol. Gen. 127, 107–140.

Feenan, K., Snodgrass, J.G., 1990. The effect of context on discrimination and bias in recognition memory for pictures and words. Mem. Cogn. 18, 517–527.

Fletcher, P., Büchel, C., Josephs, O., Friston, K., Dolan, R.J., 1999. Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. Cereb. Cortex 9, 168–178.

Fletcher, P.C., Anderson, J.M., Shanks, D.R., Honey, R., Carpenter, T.A., Donovan, T., Papadakis, N., Bullmore, E.T., 2001. Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. Nat. Neurosci. 4, 1043–1048.

Fletcher, P., Zafiris, O., Frith, C.D., Honey, R.A.E., Corlett, P.R., Zilles, K., Fink, G.R., 2005. On the benefits of not trying: brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. Cereb. Cortex 15, 1002–1015.

Frank, M.J., Loughry, B., O'Reilly, R., 2001. Interaction between frontal cortex and basal ganglia in working memory: a computational model. Cogn. Affect. Behav. Neurosci. 1, 137–160.

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2001. Categorial representation of visual stimuli in the primate prefrontal cortex. Science 291, 312–316.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A.P., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. NeuroImage 7, 30–40.

Friston, K.J., Penny, W., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16, 465–483.

Gazzaniga, M.S., 2000. Cerebral specialization and interhemispheric

communication: does the corpus callosum enable the human condition? Brain 123, 1293–1326.

Goldman-Rakic, P.S., 1996. Regional and cellular fractionation of working memory. Proc. Natl. Acad. Sci. U. S. A. 93, 13473–13480.

Gomez, R., Schvaneveldt, R.W., 1994. What is learned from artificial grammars? Transfer tests of simple association. J. Exp. Psychol., Learn. Mem. Cogn. 20, 396–410.

Habib, R., Nyberg, L., Tulving, E., 2003. Hemispheric asymmetries of memory: the HERA model revisited. Trends Cogn. Sci. 7, 241–245.

Jamieson, R.K., Mewhort, D.J.K., 2005. The influence of grammatical, local, and organizational redundancy on implicit learning: an analysis using information theory. J. Exp. Psychol., Learn. Mem. Cogn. 31, 9–23.

Knowlton, B.J., Squire, L.R., 1996. artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. J. Exp. Psychol., Learn. Mem. Cogn. 22, 169–181.

Knowlton, B.J., Mangels, J.A., Squire, L.R., 1996. A neostriatal habit learning system in humans. Science 273, 1399–1402.

Lieberman, M.D., Chang, G.Y., Chiao, J., Bookheimer, S.Y., Knowlton, B.J., 2004. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. J. Cogn. Neurosci. 16, 427–438.

Mathews, R.C., Buss, R.R., Stanley, W.B., Blanchard-Fields, F., Cho, J.R., Druhan, B., 1989. The role of implicit and explicit processes in learning from examples: a synergistic effect. J. Exp. Psychol., Learn. Mem. Cogn. 15, 1083–1100.

McClure, S.M., Berns, G.S., Montague, P.R., 2003. Temporal prediction errors in a passive learning task activate human striatum. Neuron 38, 339–346.

Meulemans, T., Van der Linden, M., 1997. Associative chunk strength in artificial grammar learning. J. Exp. Psychol., Learn. Mem. Cogn. 23, 1007–1028.

Meyer, P., Mecklinger, A., Grunwald, T., Fell, J., Elger, C.E., Friederici, A.D., 2005. Language processing within the human medial temporal lobe. Hippocampus 15, 451–459.

Miller, E.K., Erickson, C.A., Desimone, R., 1996. Neural mechanism of visual working memory in prefrontal cortex of the macaque. J. Neurosci. 16, 5154–5167.

Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. Psychol. Rev. 110, 611–646.

Nosofsky, R.M., 1986. Attention, similarity, and the identification categorization relationship. J. Exp. Psychol. Gen. 115, 39–57.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38, 329–337.

Opitz, B., Friederici, A.D., 2003. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. NeuroImage 19, 1730–1737.

Opitz, B., Friederici, A.D., 2004. Brain correlates of language learning: the neural dissociation of rule-based vs. similarity-based learning. J. Neurosci. 24, 8436–8440.

Packard, M.G., Knowlton, B., 2002. Learning and memory functions of the basal ganglia. Annu. Rev. Neurosci. 25, 563–593.

Pasupathy, A., Miller, E.K., 2005. Different time-courses of learning-related activity in the prefrontal cortex and the striatum. Nature 433, 873–876.

Perruchet, P., 1994. Defining the knowledge units of a synthetic language: commentary on Vokey and Brooks (1992). J. Exp. Psychol., Learn. Mem. Cogn. 20, 223–228.

Poldrack, R.A., Clark, J., Paré-Blagoev, E.J., Shohamy, D., Creso Moyano, J., Myers, C., Gluck, M.A., 2001. Interactive memory systems in the human brain. Nature 414, 546–550.

Pothos, E.M., 2005. The rules versus similarity distinction. Behav. Brain Sci. 28, 1–14.

Prabhakaran, V., Narayanan, K., Gabrieli, J.D.E., 2000. Integration of diverse information in working memory within the frontal lobe. Nat. Neurosci. 3, 85–90.

Ranganath, C., Blumenfeld, R.S., 2005. Doubts about double dissociations between short- and long-term memory. Trends Cogn. Sci. 9, 374–380.

Ranganath, C., D'Esposito, M., 2001. Medial temporal lobe activity associated with active maintenance of novel information. Neuron 31, 865–873.

Ranganath, C., Cohen, M.X., Brozinsky, C.J., 2005. Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. J. Cogn. Neurosci. 17, 994–1010.

Reber, A.S., 1989. Implicit learning and tacit knowledge. J. Exp. Psychol. Gen. 118, 219–235.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II. Current Research and Theory. Appleton-Century-Crofts, New York, pp. 64–99.

Rougier, N.P., Noelle, D.C., Braver, T.S., Cohen, J.D., O'Reilly, R.C., 2005. Prefrontal cortex and flexible cognitive control: rules without symbols. Proc. Natl. Acad. Sci. U. S. A. 102, 7338–7343.

Ryan, J.D., Althoff, R.R., Whitlow, S., Cohen, N.J., 2000. Amnesia is a deficit in relational memory. Psychol. Sci. 11, 454–461.

Sakai, K., Rowe, J.B., Passingham, R.E., 2002. Active maintenance in prefrontal area 46 creates distractor-resistant memory. Nat. Neurosci. 5, 479–484.

Schendan, H.E., Searl, M.M., Melrose, R.J., Stern, C.E., 2003. An fMRI study of the medial temporal lobe in implicit and explicit sequence learning. Neuron 37, 1013–1025.

Schultz, W., 2002. Getting formal with dopamine and reward. Neuron 36, 241–263.

Seger, C.A., Cincotta, C.M., 2005. The roles of the caudate nucleus in human classification learning. J. Neurosci. 25, 2941–2951.

Shanks, D.R., 1995. The Psychology of Associative Learning. Cambridge University Press, Cambridge.

Shanks, D.R., St. John, M.F., 1994. Characteristics of dissociable human learning systems. Behav. Brain Sci. 17, 367–447.

Squire, L.R., Zola, S.M., 1996. Structure and function of declarative and nondeclarative memory systems. Proc. Natl. Acad. Sci. U. S. A. 93, 13515–13522.

Strange, B.A., Fletcher, P.C., Henson, R.N.A., Friston, K.J., Dolan, R.J., 1999. Segregating the functions of human hippocampus. Proc. Natl. Acad. Sci. U. S. A. 96, 4034–4039.

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

Strange, B.A., Duggins, A., Penny, W., Dolan, R.J., Friston, K.J., 2005. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Netw. 18, 225–230.

Sutton, R.S., Barto, A.G., 1990. Time-derivative models of Pavlovian reinforcement. In: Gabriel, M., Moore, J. (Eds.), Learning and Computational Neuroscience: Foundations of Adaptive Networks. MIT Press, Cambridge, MA, pp. 497–537.

Teichmann, M., Dupoux, E., Koider, S., Brugières, P., Boissé, M.-F., Baudic, S., Cesaro, P., Peschanski, M., Bachoud-Lévi, A.-C., 2005. The role of the striatum in rule application: the model of Huntington's disease at early stage. Brain 128, 1155–1167.

Treisman, A., 1999. Solutions to the binding problem: progress through controversy and convergence. Neuron 24, 105–110.

Turner, D.C., Aitken, M.R., Shanks, D.R., Sahakian, B.J., Robbins, T.W., Schwarzbauer, C., Fletcher, P.C., 2004. The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. Cereb. Cortex 14, 872–880.

Wallis, J.D., Anderson, K.C., Miller, E.K., 2001. Single neurons in prefrontal cortex encode abstract rules. Nature 411, 953–956.

Warner, J.J., 2001. Atlas of Neuroanatomy. Butterworth-Heinemann, Boston.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.

Yeshurun, Y., Carrasco, M., 1998. Attention improves or impairs visual performance by enhancing spatial resolution. Nature 396, 72–75.

Zimmer, H., Mecklinger, A., Lindenberger, U., in press. Binding in Human Memory: A Neurocognitive Approach. Oxford University Press, Oxford.