◆ Human Brain Mapping 35:3122–3131 (2014) ◆

Hippocampal BOLD Response During Category Learning Predicts Subsequent Performance on Transfer Generalization

Francesco Fera,[1] Luca Passamonti,[2] Mohammad M. Herzallah,[3,4]* Catherine E. Myers,[5,6] Pierangelo Veltri,[1] Giuseppina Morganti,[2] Aldo Quattrone,[1,2] and Mark A. Gluck[3]

[1]Department of Surgical and Medical Sciences, University "Magna Graecia", 88100, Catanzaro, Italy
[2]National Research Council, Neuroimaging Research Unit, 88100, Catanzaro, Italy
[3]Center for Molecular and Behavioral Neurocience, Rutgers University, Newark, New Jersey
[4]Al-Quds Cognitive Neuroscience Lab, Faculty of Medicine, Al-Quds University, Abu Dis, Palestinian Territories
[5]Department of Veterans Affairs, New Jersey Health Care System, East Orange, New Jersey
[6]Department of Psychology, Rutgers University, Newark, New Jersey

Abstract: To test a prediction of our previous computational model of cortico-hippocampal interaction (Gluck and Myers [1993, 2001]) for characterizing individual differences in category learning, we studied young healthy subjects using an fMRI-adapted category-learning task that has two phases, an initial phase in which associations are learned through trial-and-error feedback followed by a generalization phase in which previously learned rules can be applied to novel associations (Myers et al. [2003]). As expected by our model, we found a negative correlation between learning-related hippocampal responses and accuracy during transfer, demonstrating that hippocampal adaptation during learning is associated with better behavioral scores during transfer generalization. In addition, we found an inverse relationship between Blood Oxygenation Level Dependent (BOLD) activity in the striatum and that in the hippocampal formation and the orbitofrontal cortex during the initial learning phase. Conversely, activity in the dorsolateral prefrontal cortex, orbitofrontal cortex and parietal lobes dominated over that of the hippocampal formation during the generalization phase. These findings provide evidence in support of theories of the neural substrates of category learning which argue that the hippocampal region plays a critical role during learning for appropriately encoding and representing newly learned information so that that this learning can be successfully applied and generalized to subsequent novel task demands. *Hum Brain Mapp 35:3122–3131, 2014.* © 2013 Wiley Periodicals, Inc.

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: National Science Foundation; Contract grant number: NSF-0718153; Contract grant sponsor: National Research Council (Consiglio Nazionale delle Ricerche), Italy.

*Correspondence to: Mohammad M. Herzallah, Center for Molecular and Behavioral Neuroscience, 197 University Avenue, Room 209, Newark, New Jersey 07102. E-mail: mohammad.m.herzallah@gmail.com

Received for publication 13 November 2012; Revised 3 July 2013; Accepted 30 July 2013.

DOI 10.1002/hbm.22389

Published online 18 October 2013 in Wiley Online Library (wileyonlinelibrary.com).

Key words: cortico-hippocampal interaction; category learning; transfer generalization; basal ganglia; BOLD activity; computational model

INTRODUCTION

In this study, we developed an fMRI-adapted version of a two-phase learning and transfer generalization task (Fig. 1) that we have previously developed and validated in various clinical populations with basal ganglia and hippocampal formation deficits [Johnson et al., 2008; Myers et al., 2002; Shohamy et al., 2006]. Prior research has shown that learning new rules from positive outcomes depends primarily on the basal ganglia, whereas the hippocampal formation structures are critical for encoding stimulus-stimulus regularities present during learning (including contextual cues), facilitating subsequent generalization of this new learning to future novel task demands and other contexts [Gabrieli, 1998; Poldrack et al., 2001; Myers et al., 2003; Squire and Zola, 1996].

While several brain imaging studies in two-phase learning and generalization tasks have previously been reported in healthy subjects by our lab and others [Johnson et al., 2008], some important questions remain unaddressed, including: What are the brain regions that support the transfer generalization of previously acquired knowledge to novel contexts? In a two-phase task, to what extent do individual differences in behavioral performance during the transfer generalization phase depend on the variability of brain responses associated with initial learning [Eichenbaum, 2000]? Can we use brain imaging responses in the initial learning phase to predict behavioral performance during the subsequent transfer generalization phase?

Functional neuroimaging (fMRI) studies have demonstrated that basal ganglia blood oxygenation level dependent (BOLD) activity increases during initial stimulus-response learning of a cognitive task [Poldrack et al., 2001]. This has been conjectured to facilitate the storage of acquired contingencies so as to maximize obtaining positive outcomes [Daw et al., 2005; Frank and Claus, 2006]. Conversely, Gluck and Myers [1993] proposed a computational model where the hippocampal formation provides both compression of redundancy (of reliably co-occurring inputs) and differentiation of stimuli predicting upcoming events [Gluck and Myers, 1993; Gluck et al., 2005]. Based on the Gluck and Myers model, the hippocampal formation plays a critical role in transfer generalization of already learned contingencies by setting up representations that allow subsequent generalization following contextual changes [Di Paola et al., 2008; Heckers et al., 2004; Preston et al., 2004; Keri et al., 2005]. Various experimental and behavioral studies have confirmed predictions of the Gluck and Myers model of the hippocampal formation; for example, previous fMRI studies suggest that the hippocampal formation BOLD activity moves from a state of high activity very early on during learning to an almost idle state by the end of learning [Poldrack et al., 2001]. Further, nondemented elderly with hippocampal atrophy show a selective deficit on generalization of previously learned rules [Myers et al., 2002, 2003]. In contrast, other studies have shown that patients with frontal and midbrain deficits show intact transfer generalization of learning [Chase et al., 2008].

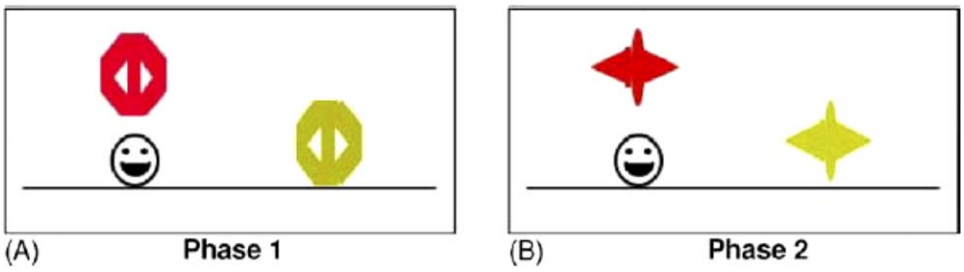

Figure 1.

(A) Screen events on a sample trial of phase 1 (concurrent discrimination) of the two-phase learning and generalization task. On each trial, the object pair is presented in random left-right order and a prompt appears. If the participant responds correctly, the chosen object is raised to reveal a smiley face icon underneath. In this pair, the color of the objects differs but shape is the same and therefore irrelevant; in other pairs (not shown), the shape of the objects differs but color is the same (irrelevant). (B) Screen events on a sample trial of phase 2 (transfer generalization): events are similar to phase 1, but the objects are changed so that the relevant dimension (here, the color) is the same, whereas the irrelevant dimension (here, the shape) is novel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The role of frontal regions has always been highlighted in learning stimulus-response contingencies. The orbitofrontal cortex (OFC) has been consistently implicated in reward learning and expectation as well as in updating stimulus-outcome contingencies [Chase et al., 2008; Murray et al., 2007]. Behavioral studies have implicated the dorso-lateral prefrontal cortex (DLPFC) in regulating attentional shifting between different stimulus contingencies [Loose et al., 2006]. The tight anatomical connections between DLPFC, OFC and parietal/extrastriate cortices indicate that this network mediates the increased need of cognitive processing [Ichihara-Takeda and Funahashi, 2008], similar to what happens during transfer generalization.

In the current study, subjects were presented with pairs of stimuli and asked to find a smiley face that is hidden under one of the stimuli (learning phase). Within each pair, the stimuli had the same color or shape but not both. Thus, one of the two dimensions—color or shape—was irrelevant within that pair. Subsequently, subjects were required to use what they had already learned to predict correct responses when the irrelevant dimension was changed in each pair (transfer generalization phase) [Gluck and Myers, 1993; Eichenbaum, 2000; 2001; Eichenbaum et al., 1989; Jiang et al., 2007].

Based on previous reports highlighting the importance of the bidirectional connections between the hippocampal formation and neocortex in cognitive function [Gluck and Myers, 1993; Eichenbaum, 2000, 2001; Eichenbaum et al., 1989; Jiang et al., 2007], we expected to detect response in the hippocampal formation and prefrontal cortices that would relate to initial learning, with the hypothesis that hippocampal activity tends to adapt with the progressive acquisition of feedback-based stimulus-response associations learned and stored in the basal ganglia [Johnson et al., 2008]. Later in learning, this is expected to be followed by a relative attenuation of the same response and a simultaneous increase in striatal activation as consolidation of contingencies progresses. Further, we predicted that the responses of hippocampal formation and OFC during transfer generalization would differ from that observed during the learning phase in order to maintain optimal level of performance, based on previous results that emphasize the role of OFC in transfer generalization [Chase et al., 2008].

METHODS

Participants

Twenty-eight Caucasian healthy volunteers (10 males, 18 females, age range 21–36 ± 4.2), with normal MRI of the brain and no history of head trauma or neuropsychiatric disorders took part in this study.

Subjects were positioned to lie comfortable in the scanner and foam pads were used to prevent head movements. In-house software was used to design the experimental paradigm, which was back-projected onto an MR compatible screen. Behavioral responses were recorded through a button box response device, connected to the USB port of a PC.

fMRI Task

We employed a version of a two-phase learning and generalization task originally developed by Myers et al. [2002]. The entire task was split into 4 stages, which were organized as follows (Fig. 2).

Stage 1 (early learning) and Stage 2 (late learning): subjects were presented with pairs of abstract objects that differed either in color or shape, but not both (so that there was one relevant and one irrelevant dimension for the discrimination) (Fig. 1). The left-right ordering of the 2 objects was counterbalanced; one member of each pair was arbitrarily designed as rewarding, but the cue-outcome was deterministic, so that the same dimension (shape or color) was always predictive for that pair. On each trial subjects were required to predict, by a key press on a button box

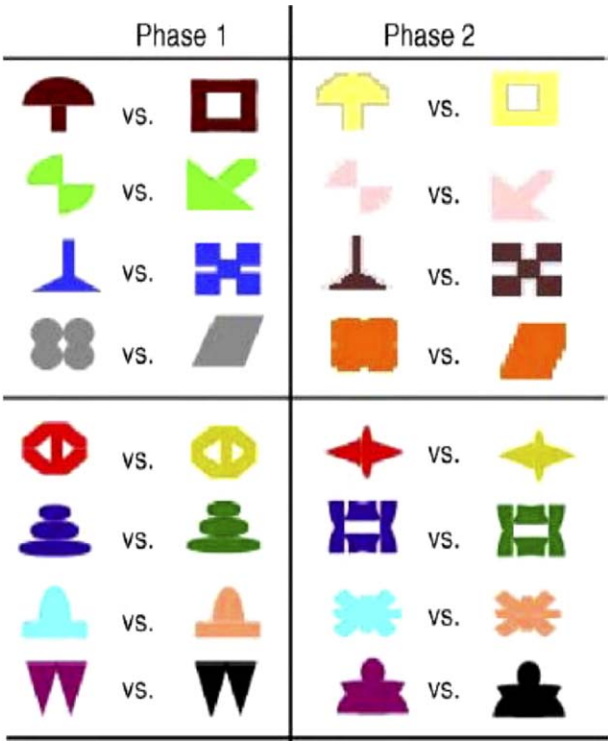

Figure 2.
Stimulus set used for concurrent discrimination (phase 1, left panels) and transfer (phase 2, right panels). Each pair of objects differed either by color (top panels) or by shape (bottom panels). For transfer (phase 2, right panels), the relevant dimension stayed the same, while the irrelevant dimension was changed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

device, which of the two objects was associated with reward (a smiling face). The chosen object was raised after the key press, to reveal the feedback. Incorrect choices were also associated with raising of the chosen object, but no smiley face appeared. Both Stage 1 and Stage 2 included eight pairs, presented in random order for each of three training blocks (24 trials per stage).

- Stage 3 (no-feedback): identical stimuli to those presented in Stages 1 and 2 were presented, but no feedback was provided. This stage again included three blocks (24 trials).
- Stage 4 (transfer): the object pairs were modified so that one relevant dimension (e.g., color) was the same, whereas the irrelevant dimension (e.g. shape) changed. As in Stage 3, no feedback was provided. There were 3 blocks (24 trials) in this stage.

During each of the four stages, blocks with the training stimuli alternated with control blocks during which subjects saw two grey squares side by side; one of the squares was clearly marked in transparency (smiley face) as the correct choice. The duration of each block was 24 s, whereas the stimulus presentation time was 2,500 ms with an inter-stimulus interval of 500 ms, for a total scanning time of 9 min and 36 s. The task total length was not long enough to justify any break in the fMRI scanning, considering that each time a new session was introduced, it had to be accounted for in the analysis model with a consequent reduction of the degrees of freedom.

Image Acquisition and Preprocessing

fMRI scanning was performed on a 3.0 T Unit (General Electric, Discovery, MR-750) with an 8-channels head coil. Whole brain data were acquired with echo-planar T2*-weighted imaging (EPI) sensitive to the BOLD signal contrast (35 axial slices, 3 mm thickness; repetition time, 2,000 ms; echo time, 30 ms; voxel size, 3 x 3 x 3 mm^3). Data were analysed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). EPI were realigned to the first scan to correct for head movements; next they were normalized to the EPI standard template in the MNI (Montreal Neurological Institute) space and smoothed with a Gaussian kernel of full width at half maximum of 8 mm. The mean EPI was computed for each subject and inspected to ensure that none showed excessive signal dropout in the medial temporal cortex and orbitofrontal cortex.

fMRI Analyses

For each participant, we used a general linear model (GLM) to assess brain regional specific effects of task parameters on BOLD indices of activation. First-level models included each type of block as an experimental factor (early learning, late learning no-feedback, transfer, and baseline conditions). Low frequency signal drift was removed by using a high-pass filter (cut-off, 128 s); in addition an autoregressive modeling (AR[1]) of temporal autocorrelations was applied. Contrast images comparing each task condition vs. baseline were generated and entered into a second-level ANOVA to produce an SPM-F map assessing the main effect of task condition (early learning, late learning, no-feedback, transfer). Furthermore, contrast images obtained from the comparison (i) late-early learning and (ii) transfer > no-feedback were entered into a second-level multiple regression GLM to explore correlations between (i) learning-related and (ii) transfer-related individual differences in accuracy (i.e., late—early learning and transfer—no-feedback, respectively) and the corresponding BOLD responses. Multiple regression GLM were also used to explore correlations between variability in brain activations during learning (i.e., late learning >early learning) and individual differences in accuracy during transfer (i.e. transfer accuracy—no-feedback accuracy).

The hippocampus, entorhinal cortex, the OFC, and dorso-lateral prefrontal cortex (DLPFC) were defined as regions of interest (ROIs) given their fundamental role in learning and memory. All ROIs were anatomical regions defined using the "aal.02" atlas for automated anatomical labelling [Tzourio-Mazoyer et al., 2002]. Two approaches were applied for thresholding second-level maps. First, for a priori ROIs, the threshold was set at $P<0.05$, family wise error (FWE) correction for multiple comparisons in small volumes (i.e. small volume correction [svc]) [Worsley et al., 1996; Friston, 1997]. Brain regions that were not predicted a priori, but met a threshold of $P < 0.001$, uncorrected, for 10 or more contiguous voxels, were also reported.

RESULTS

Behavioral Findings

Accuracy (i.e., number of correct responses) and reaction times (RT) recorded during the fMRI paradigm were entered into separate ANOVA models assessing the main effect of each task stage (early-, late-learning, no-feedback, generalization transfer). Subjects responded significantly quicker over learning trials ($F_{(df27)}=3.2$, $P < 0.05$) and with greater accuracy ($F_{(df27)}=14.7$, $P<0.001$, Fig. 3). As expected, participants maintained optimal levels of accuracy during Stage 4 (generalization transfer).

fMRI Findings

First, we used an ANOVA model to assess the main effect of task phases onto whole-brain activations. Statistically significant BOLD responses were identified in several brain areas including *a priori* regions of interest (ROI) (Fig. 4, Supporting Information Table 1). This analysis was followed-up by *post-hoc* t-tests that compared brain

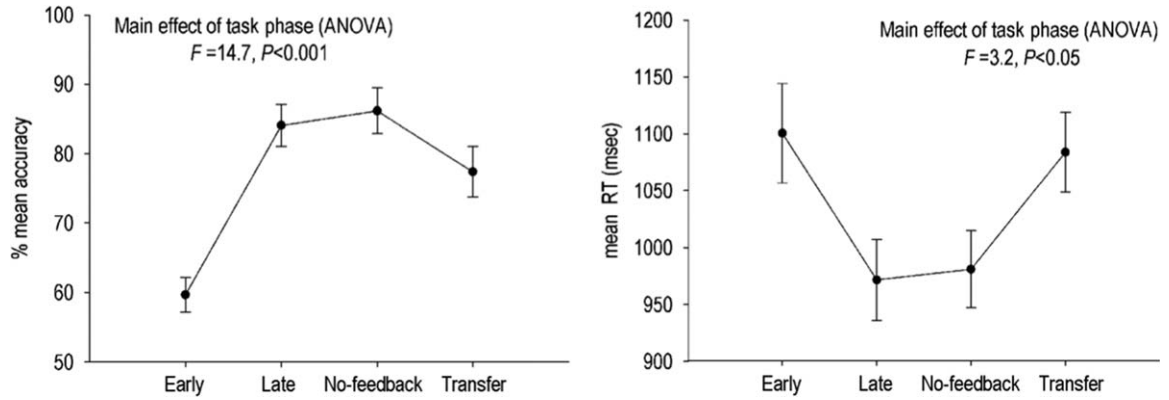

Figure 3.
Behavioral performance during each stage of the task. RT, reaction times.

responses between pairs of task conditions (i.e., early learning>late learning, no-feedback>transfer and *vice versa*). Contrasting early>late learning revealed significant responses in bilateral hippocampus, OFC and other regions outside the ROIs (Supporting Information Table 2); in contrast, the reverse comparison (i.e., late>early learning) showed a significant bilateral response within the caudate (Supporting Information Table 2). While the contrast no-feedback>transfer did not identify any brain region, the opposite comparison (transfer>no-feedback) revealed significant responses in the OFC, DLPFC, and other regions of the attentional network, i.e. the superior parietal lobe, fusiform gyrus, and extrastriate visual cortices (Supporting Information Table 2).

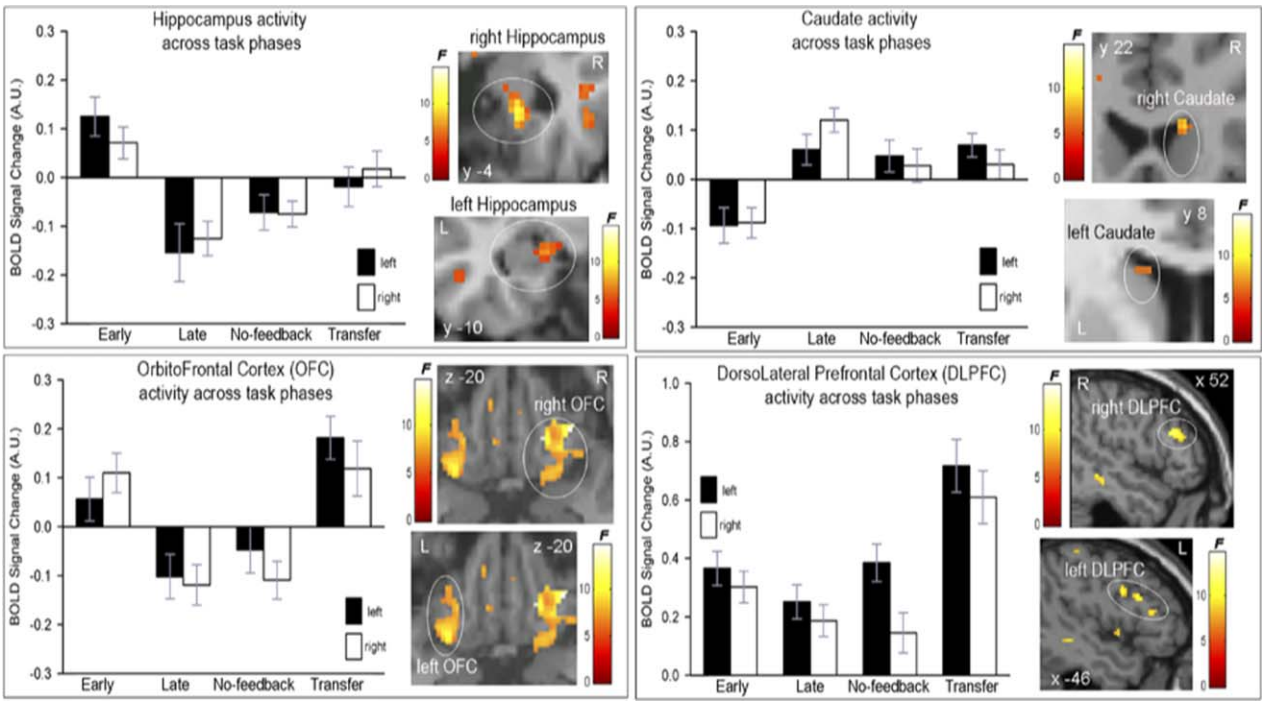

Figure 4.
Blood Oxygenated Level Dependent (BOLD) activity (A.U., arbitrary units) during each task stage (early and late learning, no-feedback and transfer generalization) in different regions of interest. Bar plots represent mean values ± standard errors. Color bars represent F statistics. Coordinates (X, Y, Z) are given in the Montreal Neurological Institute (MNI) space. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

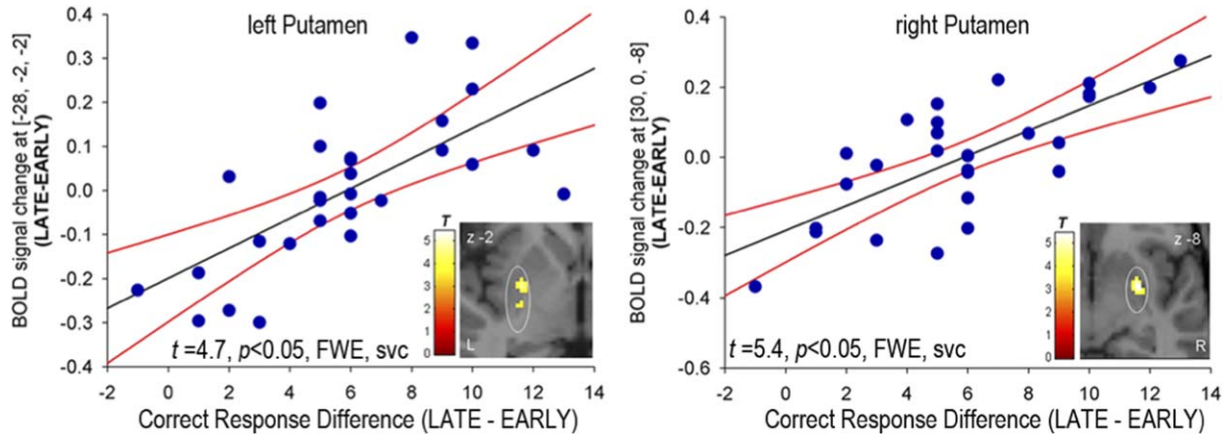

Figure 5.

Subject-specific Blood Oxygenated Level Dependent (BOLD) activity in the bilateral putamen (for the contrast late>early learning) positively correlates with individual differences in behavioral accuracy during learning (defined as the difference between correct responses during the late task stage minus correct responses during the early task stage). Black lines represent regression lines while red lines represent the 95% confidence interval. Color bars represent T statistics. FWE, svc, Family Wise Error, small volume correction. Coordinates (X, Y, Z) are given in the Montreal Neurological Institute (MM) space. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Second, multiple regression analyses tested for correlations between individual variability in task-related brain responses and differences in accuracy. The responses of bilateral putamen (for the contrast late>early learning stage) positively correlated with the accuracy difference between late and early learning (left putamen: $t = 4.7$, $P < 0.05$, FWE, svc; right putamen: $t = 5.4$, $P < 0.05$, FWE, svc (Fig. 5).

In contrast, negative correlations between hippocampal responses and difference in accuracy between transfer and no-feedback were detected when comparing transfer vs. no-feedback (left hippocampus: $t = 3.9$, $P < 0.05$, FWE, svc; right hippocampus: $t = 4.0$, $P < 0.05$, FWE, svc) (Fig. 6).

Third, to check for significant correlations between learning-related variability in brain responses (i.e. late

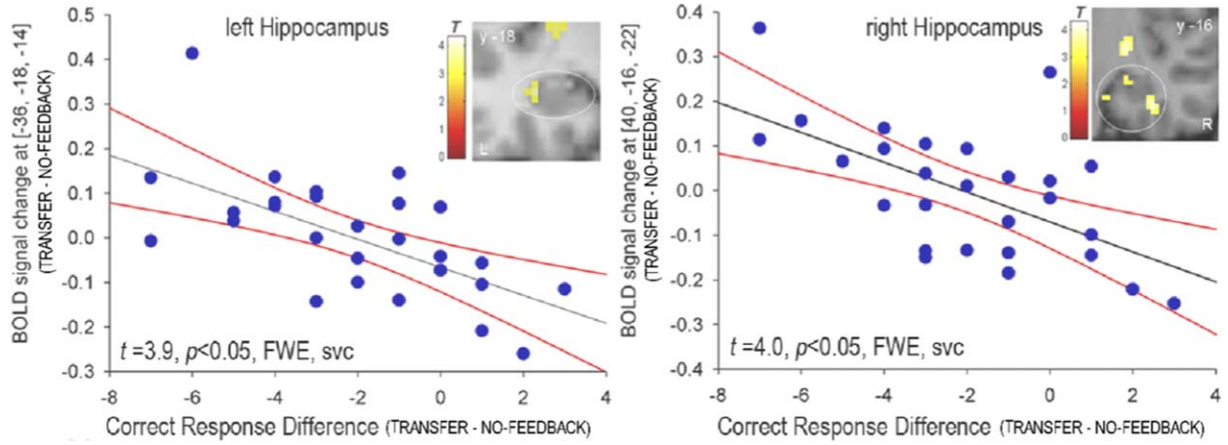

Figure 6.

Subject-specific Blood Oxygenated Level Dependent (BOLD) activity in the bilateral hippocampus (for the contrast transfer>no-feedback) negatively correlates with individual differences in behavioral accuracy during testing (defined as the difference between correct responses during the transfer task stage minus correct responses during the no-feedback task stage). Black lines represent regression lines while red lines represent the 95% confidence intervals. Color bars represent T statistics. FWE, svc, Family Wise Error, small volume correction. Coordinates (X, Y, Z) are given in the Montreal Neurological Institute (MNI) space. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

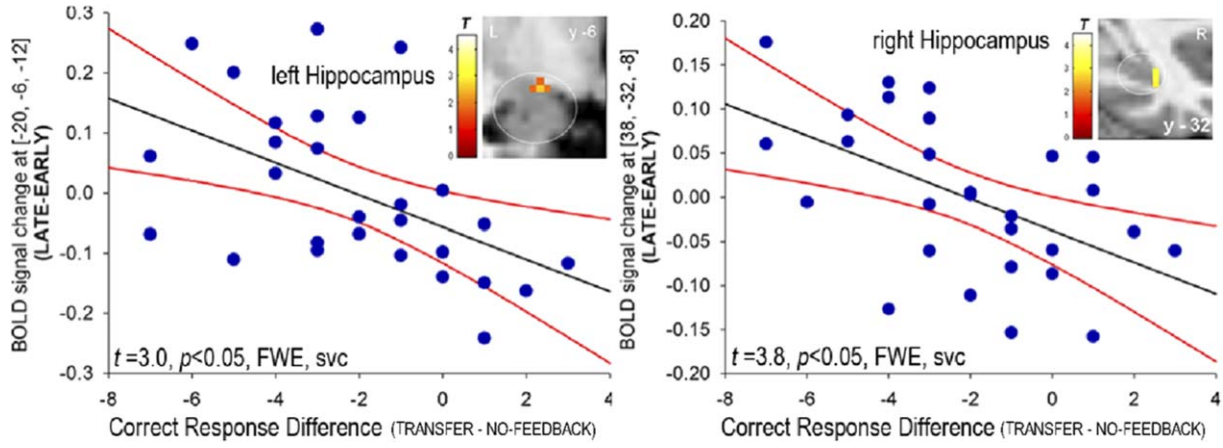

Figure 7.

Subject-specific Blood Oxygenated Level Dependent (BOLD) activity in the bilateral Hippocampus (for the contrast late>early learning) negatively correlates with individual differences in behavioral accuracy during transfer (defined as the difference between correct responses during the transfer task stage minus correct responses during the no-feedback task stage). Black lines represent regression lines while red lines represent the 95% confidence intervals. Color bars represent T statistics. FWE, svc, Family Wise Error, small volume correction. Coordinates (X, Y, Z) are given in the Montreal Neurological Institute (MNI) space. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

learning>early learning) and changes in accuracy during testing (i.e. transfer accuracy − no-feedback accuracy), we entered imaging and behavioural data into multiple regression models that investigated whether discriminative learning in the hippocampus would predict successive facilitation in transfer generalization, as hypothesized by a previous computational model [Gluck and Myers, 1993]. In the multiple regression model, correlational analysis between behavioural scores—calculated as the differences between the number of correct responses during the "transfer" phase of the task minus the "no feedback" phase of the task—and the BOLD response recorded during the late-learning phase of the task vs. the early-learning phase of the task. We found a negative correlation between learning-related hippocampal responses and accuracy during transfer, demonstrating that hippocampal adaptation during the initial learning phase is associated with better scores during subsequent transfer generalization (left hippocampus: $t = 3.0$, $P < 0.05$, FWE, svc; right hippocampus: $t = 3.8$, $P < 0.05$, FWE, svc) (Fig. 7, bottom graphs).

DISCUSSION

Our results demonstrate that the hippocampal formation, basal ganglia and neocortical regions are differentially engaged in learning as a function of task demands. During learning (late>early task stage), a decreased BOLD response of both hippocampal formation and OFC was accompanied by an increased response in the basal ganglia, which is consistent with the hypothesis that hippocampal activity tends to adapt with the progressive acquisition of feedback-based stimulus-response associations learned and stored in the basal ganglia [Johnson et al., 2008]. On the other hand, BOLD responses of DLPFC/OFC and parietal cortices dominated over that of hippocampal formation during the transfer phase of the task, suggesting that the neural circuitry involving these regions plays a major role during this phase [Gluck et al., 2008].

Previous fMRI studies displayed an inverse correlation between hippocampal formation and basal ganglia responses during category learning [Poldrack et al., 2001], where hippocampal formation BOLD activity moves from a state of high activity very early on during learning to an almost idle state by the end of the learning phase. In line with past results, we found a decreased hippocampal formation response that was paralleled by an enhanced caudate response. A possible explanation can be that, as the task contingencies are learned, more automated basal ganglia processing that maximizes individual goal-directed behavior becomes dominant [Poldrack et al., 2001]. Once automaticity in performance is reached, the basal ganglia gradually become more important to guide action selection [Seger and Cincotta, 2005; Cincotta and Seger, 2007]. In our findings, what further supports this conjecture is that individual differences in putamen activity during learning positively correlated with behavioral scores that reflect the learning rate (late learning accuracy—early learning accuracy).

Prior physiological and behavioral studies suggest that the hippocampal formation contributes to the encoding of

new information and shows response adaptation if initial learning is intact [Eichenbaum et al., 1989; Myers et al., 1996; Poldrack et al., 2001]. Our own computational models of cortico-hippocampal function propose that as learning progresses, the hippocampal formation develops internal representations of inputs that discriminate cues predicting rewards from cues that do not predict reward and, at the same time, compresses redundant features of irrelevant features including compressing the non-diagnostic stimulus cues with the tonic (and hence irrelevant) background contextual cues. The initial learning phase is followed by a transfer generalization phase in which irrelevant stimulus dimensions are changed (with no consequence) while relevant ones remain the same. Weak compression of the irrelevant cues with the background context and/or weak discrimination of relevant cues by the hippocampal formation [as proposed by Gluck and Myers, 1993] would lead to impaired transfer generalization of newly acquired rules to a future novel context, as we demonstrated in nondemented patients with hippocampal atrophy [Myers et al., 2002].

Here, we provided further evidence that the degree of neurophysiological adaptation of the hippocampal BOLD response during learning predicts behavioral accuracy during transfer generalization. Given that only one of the two dimensions (either shape or color) is relevant on each trial, the other dimension can be effectively ignored and treated as part of the background context (that is, compressed with the representations of the tonic background contextual cues). Hence, the greater the hippocampal BOLD response adaptation during learning, the better the compression of the irrelevant features needed for efficient and successful transfer generalization. Here, we argue that fast hippocampal processing in healthy subjects (resulting in better generalization) leads to faster attenuation of the hippocampal signal. The negative correlation between hippocampal formation BOLD activity and behavioral responses during transfer corroborates this interpretation and significantly contradicts the current view that familiarity is the only contributor to the hippocampal formation response adaptation [Lekeu et al., 2003; Strange et al., 1999, 2005].

During transfer generalization on this task, an altered set of stimuli is shown and no feedback is provided. In order to successfully generalize previously acquired rules (during learning), subjects have to represent task contingencies in an abstract form and use them to optimize principles that guide behavior. A previous study by Shohamy and Wagner [2008] showed a positive correlation between hippocampal formation response and generalization performance [Shohamy and Wagner, 2008], apparently contradicting our findings. Nonetheless, at least two methodological issues may explain these differences: (i) in the Shohamy and Wagner experiment, a pre-exposure phase (12 min) was administered to subjects before scanning to minimize novelty effects while in our experiment no such pre-exposure was conducted; (ii) compared to our task, Shohamy and Wagner used a more complex acquired-equivalence task with associative overlaps between partial elements of an event similar to the tasks that we (including Shohamy) used in many of our own previous studies [Shohamy and Wagner, 2008]. Given the relatively simpler structure of the task used in our current study, the assimilation of rules and the subsequent switch to automaticity of information processing may have occurred more quickly than in the Shohamy and Wagner [2008] study. Therefore, an initial boost of hippocampal formation activity in our study may have been masked by a rapid translation to adaptation to the same response. It is possible that the absence of a pre-exposure phase in our study significantly impacted the correlation between changes in hippocampal formation response during learning and subsequent generalization accuracy, and can vary between subjects according to subtle structural differences within the hippocampal formation, although this conjecture remains to be empirically tested. Alternatively, the hippocampal formation may mediate generalization via two distinct mechanisms, acting at different hierarchical levels that provide either compression of simple stimuli or direct expression of integrated information. Again, this suggests the importance of looking at the complexity of stimuli, and not just the learning and transfer paradigms, in future behavioural, clinical, and imaging studies. Finally, Shohamy and Wagner [2008] found that individual differences in generalization ranged from 38 to 100%, suggesting that different abilities at baseline (mean was 81%) might be driven by two coexisting functions, and that other higher-order processes are likely to contribute to both. Again, this suggests important avenues for further exploration and the possibility of representational changes occurring at multiple levels of abstraction.

A recent study by Wimmer and Shohamy [2012] confirmed a key prediction of our cortico-hippocampal model [Gluck and Myers, 1993] for the role of the hippocampal formation in the sensory preconditioning task [Wimmer and Shohamy, 2012]. As expected by our model, Wimmer and Shohamy found that hippocampal activation during reward learning (following an association phase) correlated with subsequent generalization of reward to previously nonrewarded stimuli. Further, they reported that hippocampal-striatal connectivity measures during reward learning positively correlate with behavioral accuracy in the reward generalization phase. However, the Wimmer and Shohamy paper did not discuss early hippocampal and striatal activation during initial associative learning. Moreover, Wimmer and Shohamy did not correlate hippocampal vs. striatal activation early on during associative learning with subsequent reward or generalization learning. In contrast, we show in our study that the adaptation of hippocampal responses early on in learning predicts better performance in generalization function. In addition, the Wimmer and Shohamy paper did not report activation of other brain areas (other than the hippocampus and the

striatum) that have almost always been implicated in associative and reward learning, such as the DLPFC and the OFC. In our study, we show how these key areas interact with the hippocampus and striatum during associative and generalization learning. Future studies ought to use sensory preconditioning tasks and more inclusive approaches in data analysis to capture the role of frontal cortices in this learning paradigm.

It is also important to point out that in our current study, the OFC response presents as a "U-shaped" function, with enhanced activation at the beginning (early learning) and the end (transfer generalization) of the task, and decreased activation during intermediate stages of the task (late learning and no-feedback), as shown in Figure 4. The OFC has been consistently implicated in reward learning, reward expectation and correction of behavioral outputs when the association strength between stimulus and outcome decreases [Chase et al., 2008; Murray et al., 2007]. Hence, the robust OFC response evoked during early learning may reflect the acquisition of associations between stimulus-stimulus contingencies and feedback, and/or the correction of wrong choices. Learning of correct rules is indeed demanded during early learning to develop correct stimulus-response representations. In general, the initial learning phase in our task is acquired very quickly by subjects, because of the relatively easy encoding of the stimulus-stimulus contingencies. Within this framework, it is possible that the fast acquisition of rules is mediated by OFC, which can be tested by adding more trials to the initial learning phase and comparing OFC across trial blocks. When feedback-based learning is completed and the adjustment of behavioral responses is no longer necessary, the striatum (caudate and putamen) may manage the automaticity of behavioral responses. Under this scenario, the basal ganglia activity would be likely to compete with that of hippocampus and OFC, which are, as a consequence, disengaged. On the other hand, a newly robust OFC response during transfer could express reward expectation and/or the re-mapping of stimulus-response associations, while suppressing inappropriate behavioral outputs. The contribution of OFC to transfer can be further investigated by adding a generalization phase to a classical reward-learning paradigm.

Past behavioral studies have implicated the DLPFC in regulating attentional shifting between different stimulus contingencies [Loose et al., 2006]. The tight anatomical connections between DLPFC, OFC and parietal/extrastriate cortices indicate that this network mediates the increased need of cognitive processing [Ichihara-Takeda and Funahashi, 2008]. In particular, a robust DLPFC response is expected during the executive aspects of learning, when selecting a choice from two alternatives is required, irrespective of accuracy [Acuna et al., 2002; Gluck et al., 2006]. Our data demonstrate that a persistent activity of DLPFC exists throughout initial phase, while learning within a novel context is associated with an additional boost of DLPFC activity. We speculate that the change in task demands induces a "re-learning" process mediated by OFC, which in turn engages a DLPFC response to identify and manipulate new task rules. Further imaging studies that vary such task demands and utilize connectivity analysis between frontal cortices and the hippocampus would be required to validate this conjecture.

Our findings indicate that the degree of initial-learning-related responses within the hippocampal formation significantly predicts accuracy during transfer generalization, thus providing evidence for an additional mechanism underlying hippocampal adaptation, above and beyond familiarity effects. However, further illustration of and dissociation of the roles of aforementioned brain regions is required to fully understand the contribution of each to learning and memory.

REFERENCES

Acuna BD, Eliassen JC, Donoghue JP, Sanes JN (2002): Frontal and parietal lobe activation during transitive inference in humans. Cereb Cortex 12:1312–1321.

Chase HW, Clark L, Myers CE, Gluck MA, Sahakian BJ, Bullmore ET, Robbins TW (2008): The role of the orbitofrontal cortex in human discrimination learning. Neuropsychologia 46:1326–1337.

Cincotta CM, Seger CA (2007): Dissociation between striatal regions while learning to categorize via feedback and via observation. J Cogn Neurosci 19:249–265.

Daw ND, Niv Y, Dayan P (2005): Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.

Di Paola M, Caltagirone C, Fadda L, Sabatini U, Serra L, Carlesimo GA (2008): Hippocampal atrophy is the critical brain change in patients with hypoxic amnesia. Hippocampus 18:719–728.

Eichenbaum H (2000): A cortical-hippocampal system for declarative memory. Nat Rev Neurosci 1:41–50.

Eichenbaum H (2001): The hippocampus and declarative memory: Cognitive mechanisms and neural codes. Behav Brain Res 127: 199–207.

Eichenbaum, H, Mathews P, Cohen NJ (1989): Further studies of hippocampal representation during odor discrimination learning. Behav Neurosci 103:1207–1216.

Frank MJ, Claus ED (2006): Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, reversal. Psychol Rev 113:300–326.

Friston KJ (1997): Testing for anatomically specified regional effects. Hum Brain Mapp 5:133–136.

Gabrieli JD (1998): Cognitive neuroscience of human memory. Annu Rev Psychol 49:87–115.

Gluck MA, Myers CE (1993): Hippocampal mediation of stimulus representation: A computational theory. Hippocampus 3:491–516.

Gluck MA, Myers C, Meeter M (2005): Cortico-hippocampal interaction and adaptive stimulus representation: a neurocomputational theory of associative learning and memory. Neural Netw 18:1265–1279.

Gluck MA, Myers CE, Nicolle MM, Johnson S (2006): Computational models of the hippocampal region: implications for

prediction of risk for Alzheimer's disease in non-demented elderly. Curr Alzheimer Res 3:247–257.

Gluck MA, Poldrack RA, Keri S (2008): The cognitive neuroscience of category learning. Neurosci Biobehav Rev 32:193–196.

Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D (2004): Hippocampal activation during transitive inference in humans. Hippocampus 14:153–162.

Ichihara-Takeda S, Funahashi S (2008): Activity of primate orbitofrontal and dorsolateral prefrontal neurons: effect of reward schedule on task-related activity. J Cogn Neurosci 20:563–579.

Jiang X, Bradley E, Rini RA, Zeffiro T, Vanmeter J, Riesenhuber M (2007): Categorization training results in shape- and category-selective human neural plasticity. Neuron 53:891–903.

Johnson SC, Schmitz TW, Asthana S, Gluck MA, Myers C (2008): Associative learning over trials activates the hippocampus in healthy elderly but not mild cognitive impairment. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 15:129–145.

Keri S, Nagy O, Kelemen O, Myers CE, Gluck MA (2005): Dissociation between medial temporal lobe and basal ganglia memory systems in schizophrenia. Schizophr Res 77:321–328.

Lekeu F, Van Der Linden M, Degueldre C, Lemaire C, Luxen A, Franck G, Moonen G, Salmon E (2003): Effects of Alzheimer's disease on the recognition of novel versus familiar words: Neuropsychological and clinico-metabolic data. Neuropsychology 17:143–154.

Loose R, Kaufmann C, Tucha O, Auer DP, Lange KW (2006): Neural networks of response shifting: Influence of task speed and stimulus material. Brain Res 1090:146–155.

Murray EA, O'doherty JP, Schoenbaum G (2007): What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. J Neurosci 27:8166–8169.

Myers CE, Ermita BR, Harris K, Hasselmo M, Solomon P, Gluck MA (1996): A computational model of cholinergic disruption of septohippocampal activity in classical eyeblink conditioning. Neurobiol Learn Mem 66:51–66.

Myers CE, Kluger A, Golomb J, Ferris S, De Leon MJ, Schnirman G, Gluck MA (2002): Hippocampal atrophy disrupts transfer generalization in nondemented elderly. J Geriatr Psychiatry Neurol 15:82–90.

Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, Golomb J, Schnirman G, Schwartz R (2003): Dissociating hippocampal versus basal ganglia contributions to learning and transfer. J Cogn Neurosci 15:185–193.

Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA (2001): Interactive memory systems in the human brain. Nature 414:546–550.

Preston AR, Shrager Y, Dudukovic NM, Gabrieli JD (2004): Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus 14:148–152.

Seger CA, Cincotta CM (2005): The roles of the caudate nucleus in human classification learning. J Neurosci 25:2941–2951.

Shohamy D, Wagner AD (2008): Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. Neuron 60:378–389.

Shohamy D, Myers CE, Geghman KD, Sage J, Gluck MA (2006): L-dopa impairs learning, but spares generalization, in Parkinson's disease. Neuropsychologia 44:774–784.

Squire LR, Zola SM (1996): Structure and function of declarative and nondeclarative memory systems. Proc Natl Acad Sci USA 93:13515–13522.

Strange BA, Fletcher PC, Henson RN, Friston KJ, Dolan RJ (1999): Segregating the functions of human hippocampus. Proc Natl Acad Sci USA 96:4034–4039.

Strange BA, Hurlemann R, Duggins A, Heinze HJ, Dolan RJ (2005): Dissociating intentional learning from relative novelty responses in the medial temporal lobe. Neuroimage 25:51–62.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

Wimmer GE, Shohamy D (2012): Preference by association: How memory mechanisms in the hippocampus bias decisions. Science 338:270–273.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996): A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58–73.