Neuron, Vol. 47, 295–306, July 21, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.neuron.2005.06.008

Formal Learning Theory Dissociates Brain Regions with Different Temporal Integration

Jan Gläscher* and Christian Büchel
Neuroimage Nord
Department of Systems Neuroscience
University Medical Center Hamburg-Eppendorf
Martinistrasse 52
D-20246 Hamburg
Germany

Summary

Learning can be characterized as the extraction of reliable predictions about stimulus occurrences from past experience. In two experiments, we investigated the interval of temporal integration of previous learning trials in different brain regions using implicit and explicit Pavlovian fear conditioning with a dynamically changing reinforcement regime in an experimental setting. With formal learning theory (the Rescorla-Wagner model), temporal integration is characterized by the learning rate. Using fMRI and this theoretical framework, we are able to distinguish between learning-related brain regions that show long temporal integration (e.g., amygdala) and higher perceptual regions that integrate only over a short period of time (e.g., fusiform face area, parahippocampal place area). This approach allows for the investigation of learning-related changes in brain activation, as it can dissociate brain areas that differ with respect to their integration of past learning experiences by either computing long-term outcome predictions or instantaneous reinforcement expectancies.

Introduction

Learning can be characterized by the acquisition of knowledge of reliable relationships between occurrences of contingent events in the environment (Schultz and Dickinson, 2000). For example, a mailman might associate a particular part of his daily delivering route with the occurrence of an aggressive, freely roaming dog. This stimulus-stimulus association manifests itself in probabilistic predictions of an event given some other event. Two necessary prerequisites exist in order to acquire valid and reliable stimulus predictions: (1) a correlation between two events, such that the occurrence of one event predicts the other with a specific probability (contingency), e.g., the co-occurrence of the dog at a particular part of the delivering route on several days; and (2) the organism (e.g., the mailman) needs to process and represent this correlation by means of integrating past experiences of these two events (temporal integration). However, in a dynamically changing environment, the contingency between two events might change (e.g., the owner and dog might go on vacation and return a few weeks later). Consequently, these changes in contingencies would also have to be represented.

How would an organism respond to such variation in learned contingencies? One possible solution would be to immediately adapt the stimulus prediction to reflect only the recent learning trials. While this is advantageous if the contingency change remains intact (e.g., another attack is imminent), it is maladaptive if the contingency violation is only a random event. Another possible solution would be to only slowly adapt the stimulus prediction and rely more strongly on the established contingency of the past learning history. The latter strategy is targeted toward the optimization of long-term behavioral adaptation, while under the former, additional resources for an increased alertness level are made available on a short-term basis. Balancing these two strategies based on the evaluation of the current situation provides the basis for flexible and adaptive behavior. It is therefore likely that the central nervous system of a behaving organism has the computational capabilities of representing both long- and short-term strategies by means of a different degree of temporal integration, possibly in distinct brain regions.

Pavlovian conditioning is a very basic type of learning ideally suited to investigate the formation of stimulus predictions and the interval of temporal integration in an experimentally controlled setting (Pavlov, 1927). In this procedure, a neutral stimulus acquires a predictive value, thus becoming a conditioned stimulus (CS), for the occurrence of a biologically salient, unconditioned stimulus (US) when the two are consistently paired. The rationale behind Pavlovian conditioning rests on the notion of reinforcement contingency (probability of US occurrence given the CS) as the association-forming force of learning, while at the same time decreasing the importance of the spatiotemporal relationship (contiguity) of the two stimuli (Rescorla and Wagner, 1972).

Previous studies investigating the neural correlates of Pavlovian conditioning have identified medial temporal lobe (MTL) structures (e.g., the amygdala and the hippocampus). While the amygdala is thought to be involved in the formation of the CS-US association (LeDoux, 2000; Maren, 2001), the hippocampus is thought to be important for the maintenance of the CS representation in trace conditioning paradigms in which CS and US are separated in time (Eichenbaum and Cohen, 2001). Functional neuroimaging studies have confirmed the involvement of the abovementioned structures in aversive fear-conditioning studies (Buchel et al., 1999; Buchel et al., 1998; Cheng et al., 2003; Jensen et al., 2003; Knight et al., 2004; LaBar et al., 1998; Morris et al., 2001).

Almost all of these studies have acquired imaging data only during acquisition, with the exception of LaBar and colleagues (LaBar et al., 1998), who also scanned the extinction phase. Thus, despite occasional pretraining habituation, none of these studies have investigated the dynamics of changes in contigency. Differential conditioning (reliable US predictions) is commonly demonstrated by (1) categorical comparisons between rein-

*Correspondence: glaescher@uke.uni-hamburg.de

forced (CS+) and nonreinforced (CS–) conditioned stimuli (i.e., Jensen et al., 2003) or (2) increases or decreases in brain activation during CS processing in the form of a time × condition (CS+/CS–) interaction (Buchel et al., 1999; Buchel et al., 1998; Gottfried et al., 2002; Morris et al., 2001). However, changes over time can be confounded by effects not related to learning per se, like habituation. One way to investigate how and where the brain represents the stimulus predictions is to systematically manipulate the CS-US contingency, i.e., dynamically change the probability of reinforcement (US occurrence), thus simulating changes of the correlative structure of events as they can occur in real-world learning settings.

Another prerequisite for Pavlovian conditioning is the temporal integration of past experiences for the detection of the correlative structure between contingent events. This question can be ideally investigated by employing formal learning theories, since they embody the temporal integration as a model parameter. Computational models of Pavlovian conditioning like the Rescorla-Wagner (RW) model (Rescorla and Wagner, 1972) rest on the prediction error as the reinforcement signal (Vogel et al., 2004) and have been very successful in describing behavior (Miller et al., 1995) and neuronal activity (Schultz et al., 1997; Schultz and Dickinson, 2000) in the course of learning. The prediction error refers to the difference between the actual (R_t) and the predicted outcome (V_t) of a particular learning trial *t*. This error is subsequently used to adapt future predictions. The RW model is an important trial-based model of prediction error learning, which conceptualizes the updating of stimulus predictions V_t as the sum of the previous predictions and the weighted prediction error $[V_{t-1} + \epsilon(R_t - V_t)]$ (equations adopted from Dayan and Abbott, 2001). Thus, the updating of stimulus predictions relies on two sources: (1) the past (long-term) learning history that accumulates in V_{t-1} and (2) the instantaneous (short-term) prediction error $(R_t - V_t)$ that relates current outcome to the previous learning history. Crucially, the learning parameter ϵ regulates the influence of the current prediction error in the prediction update. If the parameter ϵ is high, the current prediction error exerts a strong influence on the prediction update, thus ignoring past experience as the primary influence. Conversely, if ϵ is low, the prediction update is primarily driven by past learning experience, mostly disregarding current prediction error (see Figure 1). Thus, the learning parameter ϵ can be seen as a time constant that captures the interval of temporal integration of past learning experiences in different brain regions. Functional neuroimaging is ideally suited to detect this integration, as it is capable of assessing activity changes in multiple brain areas at the same time during learning. Long temporal integration suggests a sustained representation of the acquired value in form of the stimulus prediction, while short temporal integration can be interpreted as an anticipatory signal encoding the instantaneous reinforcement expectancy.

Hence, we developed an fMRI experiment for Pavlovian fear conditioning in which reinforcement contingencies were systematically varied during the course of the experiment. This resulted in experimental phases in which the CS reliably predicted the presence or absence of the US and intermediate phases in which the contingencies were unclear (see Experimental Procedures and Figure 1). We chose two visual stimuli (face/house) as CSs and an aversive pain stimulus as the US (pricking laser pain in experiment 1 and heat pain delivered by a thermode in experiment 2).

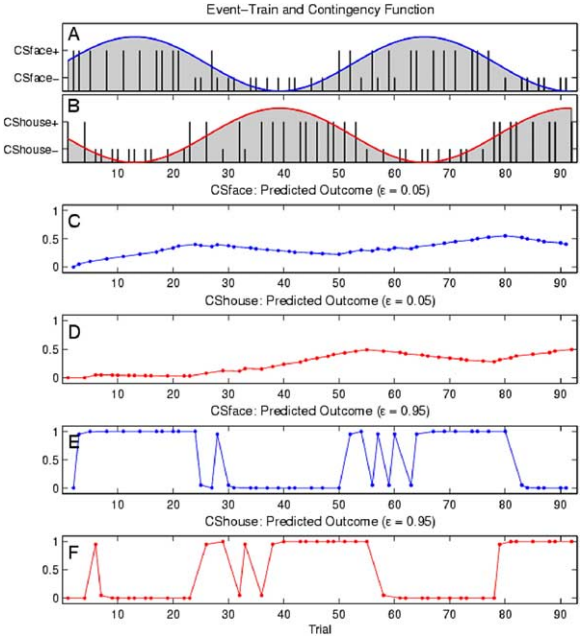

Figure 1. Predictions from the Rescorla-Wagner Model and Experimental Design for Experiment 1

(A and B) Event train (stick function) and contingency threshold function (sine waves) for face CS (blue) and house CS (red). Long sticks indicate reinforced CSs (CS+); short sticks indicate nonreinforced CSs (CS–). CSs were presented in the background; subjects performed a one-back working memory task in the foreground.

(C–F) Predictions derived from the Rescorla-Wagner model for the face CS (blue) and the house CS (red) at two different learning rates (ϵ = 0.05 and 0.95).

In order to prevent subjects from developing elaborate cognitive strategies for predicting US occurrence, we chose an implicit conditioning paradigm in experiment 1 and distracted them with a simple working memory (WM) task (Carter et al., 2003). As a physiological index of learning, we collected skin conductance responses (SCRs) throughout the first experiment. Furthermore, we sought to replicate and generalize the findings in experiment 1 and adopted a similar, but explicit conditioning paradigm in experiment 2, in which subjects were asked to provide online ratings of US expectancy.

Imaging data were analyzed within the RW framework employing either a low or a high learning parameter ϵ to determine the interval of temporal integration. We expected brain regions known for their involvement in learning and memory (such as MTL structures) (Buchel et al., 1998, 1999; Jensen et al., 2003; Morris et al., 2001; Ploghaus et al., 2001) to be best characterized by long temporal integration (low ϵ).

Conversely, by means of short temporal integration, the brain is able to generate an anticipatory signal that

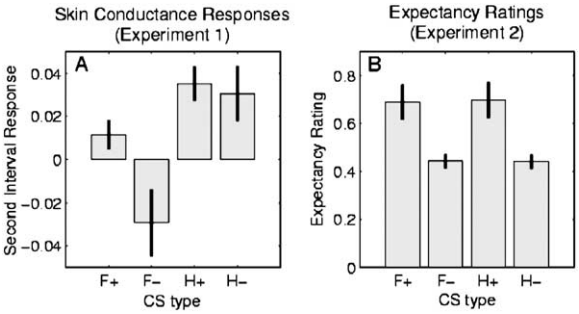

Figure 2. Behavioral Data from Both Experiments

(A) Mean skin conductance response (second interval reponses, see Experimental Procedures) to all four CS event types in experiment 1 across subjects (error bar, SEM). F+/H+, reinforced face/house CS; F–/H–, nonreinforced face/house CS.

(B) Mean expectancy ratings for all four CS types in experiment 2 across subjects (error bar, SEM). Labels as in panel (A).

could selectively modulate the processing in perceptual areas (e.g., processing enhancement in expectation of upcoming aversive reinforcement). Thus, we expected to find short temporal integration (high ϵ) in areas related to perceptual processing in the ventral visual system (Epstein et al., 1999; Kanwisher et al., 1997).

Results and Discussion

Experiment 1

Prior to data analysis, we verified that subjects maintained a high level of target detection accuracy and showed typical activations for painful laser stimulation (see the Supplemental Results available with this article online).

Skin Conductance Data

The demonstration of differential conditioning (CS+ versus CS–) after the acquisition phase is a widely accepted index for learning in a Pavlovian conditioning experiment (Knight et al., 2003; Ohman and Soares, 1994, 1998). However, in a design with variable contingencies, it is difficult to select those trials in which a CS reliably predicts US presence (CS+) or absence (CS–). We compared the mean skin conductance response (SCRs) of "high" versus "low" contingency CSs (for threshold, see Experimental Procedures) in a paired t test across all subjects and found a significant differential conditioning effect ($T_{13} = 2.29$, $p < 0.05$). Detailed inspection of the data revealed that this significant differential conditioning effect was limited to the face CS ($T_{13} = 2.30$, $p < 0.05$) and did not occur for the house CS ($T_{13} = 0.46$, $p > 0.6$) (Figure 2A).

Imaging Data

In order to investigate the regional specificity of the interval of temporal integration, we included regressors encoding the RW predictions with a low and a high learning rate ($\epsilon = 0.05$ and 0.95, respectively, see Figures 1C–1F) in the first level analysis (for identification of ϵ values see Experimental Procedures). The prediction regressor allows for an amplitude modulation of the modelled BOLD response. These prediction regressors were then compared with differential contrasts in a repeated-measures ANOVA at the second level, treating subjects as the random effect.

First, we calculated the contrast low > high ϵ regressors. This effect is demonstrated in the left amygdala (trend level in the right reported here for the similarity of the parameter estimates) (Figure 3A). Other brain regions showing a similar response pattern in the same contrast are bilateral nucleus accumbens, bilateral ventral putamen, and bilateral hippocampus (see Figure S1). Further regions (Table 1, part I) were midcingulate, perigenual cingulate cortex, the hand area of primary somatosensory cortex (S1) contralateral to the stimulated hand, and the red nucleus. Interestingly, the differential effect of low > high ϵ prediction regressor was more pronounced for the face CS, paralleling our results in the autonomic data (see parameter estimates in Figure 3A). Table 1 (part I) lists the Z values for all of the aforementioned regions that correspond to the particular volumes of interest.

The interpretation of the response pattern of parameter estimates (β weights) displayed adjacently to the statistical maps is complicated because the sign and the relative size of both low and high ϵ estimates influence the modulation of the fitted response in the particular voxel. A negative estimate reverses the direction of the modulation, while the relative size of the estimates determines which regressor dominates the modulation of the fitted response. In order to visualize the combined effects of both high and low ϵ modulation, we plotted the fitted predictions in the left amygdala (Figures S3C and S3D). This figure clearly shows a distinct modulatory pattern depending on the CS type (face/house): while the influence of the low ϵ prediction dominates the modulation for the face CS, the modulation for the house CS is primarily driven by the negative high ϵ prediction yielding an increasing modulatory influence on the BOLD response in the left amygdala during phases of nonreinforcement.

In contradistinction, the reverse contrast (high > low ϵ prediction regressors) yielded a significant effect in the left fusiform face area (FFA, Figure 3B). Further regions that showed a similar effect were found in bilateral anterior insula and in the left lateral orbitofrontal cortex (Figure S2). Table 1 (part II) lists the Z values for these regions that correspond to the particular volumes of interest. The pattern of high and low ϵ estimates in these regions suggests a different modulatory influence of the RW prediction than in the amygdala (Figures S1E and S1F): the modulations of BOLD responses to the face CS are almost entirely influenced by the high ϵ prediction, while the negative low ϵ prediction exerts an influence only for the house CS, yielding a response pattern in which the modulation decreases as the house becomes more and more predictable of the US occurrence.

Experiment 2

We conducted a second experiment in order to replicate and generalize our findings with an explicit version of the task and to overcome potential explanatory confounds of the previous experiment. In experiment 1, all subjects were presented with the same event train.

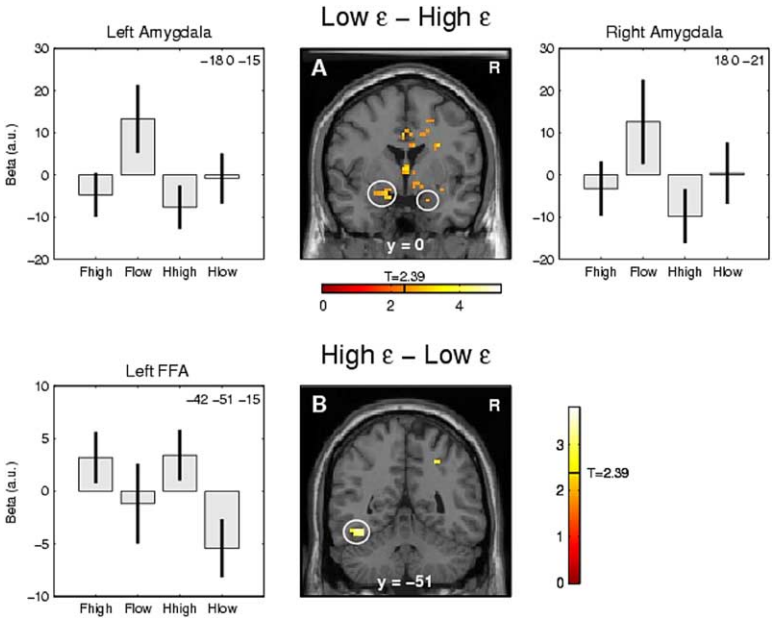

Figure 3. Neuroimaging Results from Experiment 1

(A) Brain activations to low > high ϵ prediction regressors in experiment 1. The SPM (threshold: $p < 0.01$) shows significant clusters of activation in the left amygdala (trend level on the right). The bar graphs to the left and right display the mean parameter estimates of both low and high ϵ prediction regressors (across subjects) for both CSs in the peak activation voxel (in MNI space) of the circled clusters shown in the SPM (error bar, 90% c.i.). Flow/Fhigh, prediction regressor for the face CS with low/high learning rate; Hlow/Hhigh, prediction regressor for the house CS with low/high learning rate.

(B) Brain activations to high > low ϵ prediction regressors in experiment 1. The SPM (threshold: $p < 0.01$) shows significant clusters of activation in the left FFA. Bar graphs display the mean parameter estimates for the peak voxel in the FFA as in panel (A).

Thus, it might be possible that subjects only learned the contingency of the face CS because it was always the first CS to be systematically reinforced. Another potential confound was that the two contingency curves were exact complements of each other: when the face CS was reinforced, the house CS was not, and vice versa. Thus, subjects could have simply learned the face contingency and implicitly inferred the house contingency. Third, it remained unclear whether the observed modulations in the amygdala and the FFA crucially rest on the implicit nature of the conditioning paradigm in experiment 1. Thus, in a different version of the task, we intended to generalize the findings of experiment 1 and rule out confounds and alternative explanations. In experiment 2, we sought to obtain an explicit and direct behavioral measure of the US prediction. Consequently, subjects were asked to give binary ratings of US expectancy within the first 2 s of CS occurrence. These altered instructions change the paradigm from an implicit to an explicit conditioning design. In order to control for inverse contingencies, we shifted the phase of the two contingency curves by only 90° (Figure 4). Finally, we also inserted a short period of random 50% partial reinforcement in the event train

Table 1. Statistical Results of Experiment 1 Corrected for Spherical Search Volumes at the Peak Voxel in MNI Space

I. Low > High ϵ Prediction Regressor								
	Left Hemisphere				Right Hemisphere			
Region	x	y	z	Z	x	y	z	Z
Amygdala	−18	0	−15	2.96 *	30	0	−12	2.51 (n.s, p = 0.099)
Hippocampus	−27	−12	−18	4.15 ***	30	−15	−15	3.59 **
					24	−18	−15	3.58 **
Nucl. Accumbens	−9	6	−9	3.23 *	12	−15	−12	3.44 **
	−12	9	−9	2.96 *	9	9	−3	2.69 *
Ventral Putamen	−30	6	−12	2.76 *	33	6	−12	3.73 **
Red nucleus	−9	−15	−15	2.79 *	12	−15	−12	3.33 **
Mid-cing. Cortex	−3	3	36	3.15 *				
Perigenual ACC					0	36	6	3.84 **
Hand area S1					39	−30	63	3.41 **
II. High > Low ϵ Prediction Regressor								
	Left Hemisphere				Right Hemisphere			
Region	x	y	z	Z	x	y	z	Z
Fusiform Gyrus (FFA)	−42	−51	−15	3.35 *				
Ant. Insula	−33	12	6	3.91 **	39	15	6	3.18 *
Orbitofrontal Cortex	−36	36	−9	3.10 *				

The center of the search volume was determined from the coordinates of other conditioning studies or by anatomical definition. The volumetric extent of the search volume was determined by the approximated volume of the target structure (see Experimental Procedures). *p < 0.05, **p < 0.01, ***p < 0.001 (small volume correction); n.s., nonsignificant.

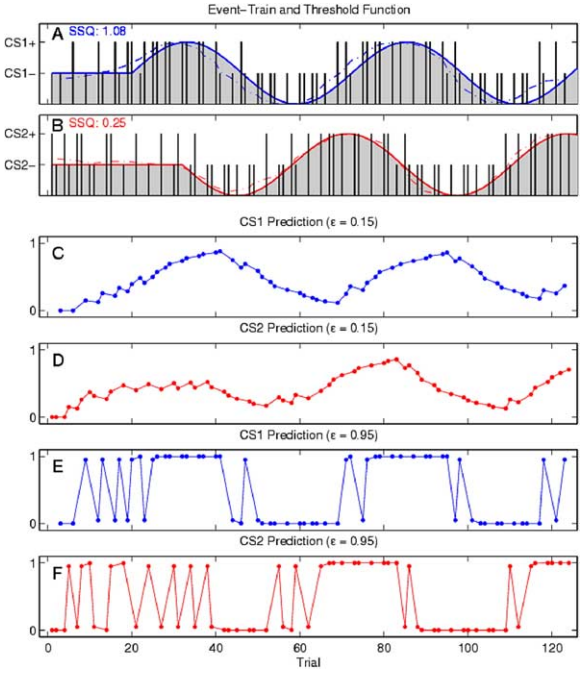

Figure 4. Predictions from the Rescorla-Wagner Model and Experimental Design for Experiment 2

(A and B) Event train (stick function) and contingency threshold function (sine waves) for face CS (blue) and house CS (red). Long sticks indicate reinforced CSs (CS+); short sticks indicate nonreinforced CSs (CS–). The dashed line visualizes the fit of the event train with the idealized contingency curve (see Experimental Procedures).

(C–F) Predictions derived from the Rescorla-Wagner model for the face CS (blue) and the house CS (red) at two different learning rates (ϵ = 0.15 and 0.95).

prior to any systematic contingency changes in order to distract subjects from cognitive search processes for the correct contingency pattern. Importantly, we created individual event trains for each subject. In addition, we counterbalanced the CS assignments across subjects, i.e., for some subjects the face was the first CS to be systematically reinforced, while for others it was the house. Figure 4 shows an example event train and derived RW prediction curves (see Experimental Procedures for more detail). As in experiment 1, we verified that the US (thermode heat pain) elicited similar activations in pain-related regions (see Supplemental Results)

Behavioral Data

We tested our behavioral data in a statistical approach analogous to the SCRs in experiment 1 by comparing the expectancy ratings of those events in which the CS-US contingency was "high" against those in which the contingency was "low" (see Experimental Procedures for analogous thresholds as in experiment 1). These comparisons yielded highly significant learning effects for both CSs at the group level (face: $T_{16} = 4.284$, $p < 0.001$; house: $T_{16} = 4.355$, $p < 0.001$; Figure 2B). At the single-subject level, we identified successful learning when the behavioral responses were positively correlated with the reinforcement regime. This criterion revealed that seven subjects learned the contingencies of both CSs, five additional subjects learned only the contingency of one CS, while five subjects did not learn the contingency of either CS.

Based on these observations in the total sample of 17 subjects, we confined the analysis of the imaging data to those 12 subjects who learned the association of at least one CS and the US, but included only the effect size images for the high and low ϵ modulation for which we found behavioral evidence of learning (see Experimental Procedures).

Imaging Data

In order to directly compare these results with those of experiment 1, we performed the same analysis of the imaging data. As indicated by the exploratory sampling of ϵ parameter space (see Experimental Procedures), we set the low ϵ to 0.15 and the high ϵ to 0.95. The derived predictions at these learning rates were included as parametric modulations in the first level analysis and were later compared with differential contrasts in a repeated-measures ANOVA at the second level.

For the comparison of low > high ϵ regressors, we found a significant effect in the right amygdala (Figure 5A). The corresponding Z and p values for a reduced search volume are listed in Table 2 (part I). In contrast to the findings from experiment 1, we did not observe any effects in the nucleus accumbens, the ventral putamen, or the hippocampus for this comparison.

In the reverse contrast (high > low ϵ regressor), we found significant effects in bilateral FFA and the left parahippocampal place area (PPA), an area known to activate during the processing of spatial configuration, in particular houses (Epstein et al., 1999) (Figures 5B and 5C). Further regions that showed a similar effect were found in the anterior insula and orbitofrontal cortex (Table 2, part II).

Strikingly, the pattern of parameter estimates for the high and low ϵ regressors in amygdala and FFA very closely corresponded to that of experiment 1 (see Figures 3 and 5 for visual comparison). We also plotted the fitted prediction curve for better visualization of the modulatory influences of the high and low ϵ regressors for this experiment (Figure S4). These curves also strongly underline the close correspondence to the fitted prediction curves from experiment 1. Thus, in experiment 2, we demonstrate very precise replications of our findings from experiment 1, despite significant changes to the experimental procedures, especially the cognitive task.

In both experiments, we investigated the representation of CS-US contingencies using Pavlovian fear conditioning. In experiment 1, we chose an implicit variant of the paradigm (Carter et al., 2003), while in experiment 2, we adopted an explicit version requiring the subjects to rate their US expectancy. Behaviorally, we found evidence for a selective learning effect for face CS in experiment 1 and for both CSs in experiment 2. Using a paradigm involving dynamically changing reinforcement contingencies, we sought to determine the interval of temporal integration of past experiences necessary for generating reliable stimulus predictions (Schultz and Dickinson, 2000). These varying contin-

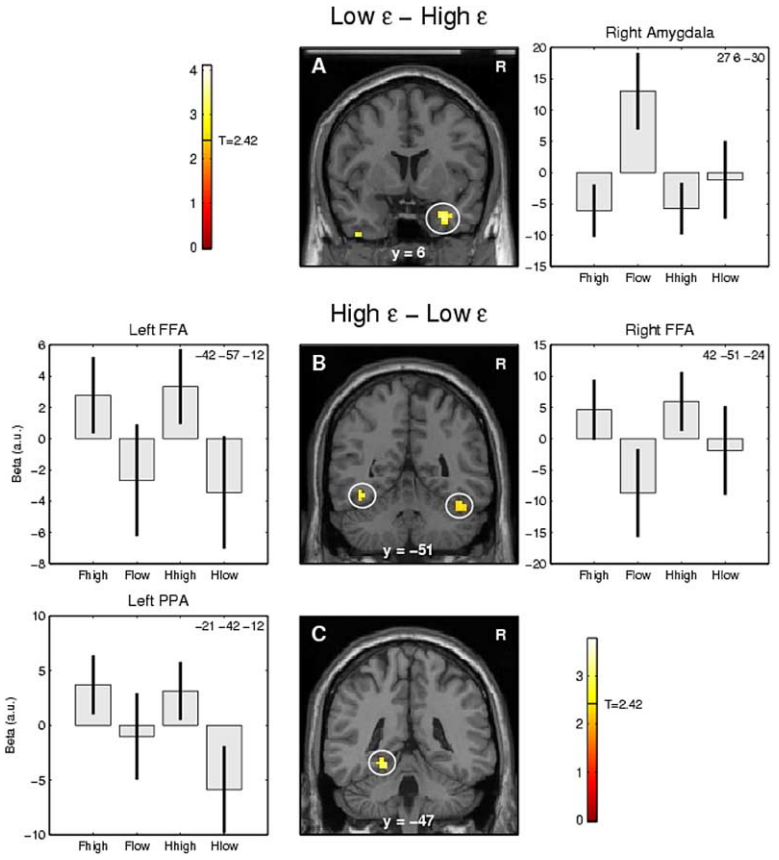

Figure 5. Neuroimaging Results from Experiment 2

(A) Brain activations to low > high ϵ regressors in experiment 2. The SPM (threshold: $p < 0.01$) shows a significant cluster in the right amygdala. Bar graphs display the mean parameter estimates for low and high ϵ regressors for both CSs. Note the close resemblance of these findings with those shown in Figure 3 (labels as in Figure 3).

(B and C) Brain activations to high > low ϵ regressors. The SPMs (threshold: $p < 0.01$) show significant clusters in bilateral FFA and left PPA. Bar graphs display the mean parameter estimates for low and high ϵ regressors for both CSs. These parameter estimates also closely match those of experiment 1. Flow/Fhigh, prediction regressor for the face CS with low/high learning rate; Hlow/Hhigh, prediction regressor for the house CS with low/high learning rate.

gencies simulate changes in stimulus-stimulus associations as they can occur in real-world settings. In line with our hypotheses, we found significant effects of temporal integration in both experiments in two sets of regions in which the BOLD signal covaried with the prediction as derived from the Rescorla-Wagner model, albeit on different time scales. The amygdala displayed a long temporal integration (low learning parameter ϵ). On the contrary, we found effects of short temporal integration (high learning parameter ϵ) in ventral visual areas known for their specific involvement in processing the face and house CS (FFA and PPA). Although we also found long temporal integration in other areas (e.g., hippocampus and ventral striatum) and short temporal integration in anterior insula and orbitofrontal cortex in experiment 1, we decided to adopt a conservative approach and discuss only those areas that show a similar and robust effect in both experiments.

Long Temporal Integration in the Amygdala

There is a widely accepted consensus that the amygdala plays a pivotal role in the formation of CS-US as-

Table 2. Statistical Results of Experiment 2 Corrected for Spherical Search Volumes at the Peak Voxel in MNI Space

I. Small > Large ϵ Prediction Regressor								
	Left Hemisphere				Right Hemisphere			
Region	x	y	z	Z	x	y	z	Z
Amygdala					27	6	–30	3.29 *
II. Large > Small ϵ Prediction Regressor								
	Left Hemisphere				Right Hemisphere			
Region	x	y	z	Z	x	y	z	Z
Fusiform Gyrus (FFA)	–42	–57	–12	2.95 *	42	–51	–24	2.63 *
Parahippocampus (PPA)	–21	–42	–12	2.99 *				
Ant. Insula	–33	15	3	2.59 *				
Orbitofrontal Cortex	–42	39	–9	2.82 (n.s., $p = 0.058$)				

The center of the search volume was determined from the coordinates of other conditioning studies or by anatomical definition. The volumetric extent of the search volume was determined by the approximated volume of the target structure (see Experimental Procedures). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (small volume correction); n.s., nonsignificant.

sociations in Pavlovian conditioning (Buchel et al., 1998, 1999; LeDoux, 2000; Maren, 2001) The association between CS and US is thought to be formed in the basolateral amygdala comprising the lateral, basal, and accessory basal subnuclei (Cardinal et al., 2002; LeDoux, 2000). Because reinforcement contingencies lie at the heart of the association between CS and US, one would expect amygdala activation to covary with these contingencies. Although we found a nearly identical pattern of prediction modulation in both experiments, the pattern itself suggests a differential modulation for the two CSs. The modulation for the face CS follows the pattern of long temporal integration (low ϵ) and is corroborated by a significant differential conditioning effect in the SCR data of experiment 1. However, the modulatory pattern for the house CS reflects short-term integration in the reverse direction: the activation immediately decreases following a US presentation and increases upon nonreinforced trials. This modulation can be characterized as (inhibitory) active unlearning (Myers and Davis, 2002). The animal literature has produced considerable evidence for involvement of the amygdala in inhibitory unlearning (for review see Falls and Davis, 1995). It is beyond the scope of these experiments to disentangle the different modulatory influences on the amygdala in an experimentally controlled fashion. However, we suggest that the observed pattern is not a mere order effect or an effect of a specific sequence of reinforcement because these confounds were carefully controlled for through complete randomization in experiment 2. Rather, we speculate that the reason for the selective prediction modulation to the face CS might lie in its biological relevance. Facial expressions are an important social signal that allow an evaluation of the benevolence or hostility of the social environment (Glascher et al., 2004). Furthermore, is it a well-known fact that the amygdala plays a crucial role in the processing of facial expressions (for review see Adolphs, 2002), suggesting that faces are biologically more salient than houses. Early conditioning studies demonstrated that biologically salient stimuli can be better conditioned than nonsalient stimuli, even in the absence of awareness (Esteves et al., 1994; Ohman and Soares, 1994). The observed difference in the prediction modulation for both CSs suggests that the amygdala closely reflects (or encodes) the long-term reinforcement contingencies for biologically salient stimuli, while simultaneously suppressing the prediction encoding for a competing, but less salient CS (short-term inhibitory unlearning). It would require further investigation to characterize this difference in prediction modulation in more detail, e.g., with two face CSs with potentially different facial expressions.

These differential findings seem to be supported by the behavioral data of experiment 1 (implicit task), in which subjects developed differential SCRs only for the face CS. However, under an explicit cognitive task (experiment 2), differential expectancy ratings were found for both CSs. Given these findings, we suggest that the acquisition of a Pavlovian conditioning under subtle changes in reinforcement contingencies and under implicit processing is facilitated when biologically salient stimuli are employed as CSs (Ohman and Soares, 1994, 1998), whereas directing attention to US expectancies (explicit task) attenuates this bias. Interestingly, the amygdala seems to do the same computations regardless of the task (i.e., the parameter estimates of both experiments closely correspond). Previous studies have suggested that the amygdala may be critically involved in the differential SCRs in experiment 1 (Cheng et al., 2003; Phelps et al., 2001). While our findings are in line with this interpretation, we also suggest that other brain areas may be involved in the differential expectancy effects for both CSs in experiment 2.

The present study departs from earlier neuroimaging studies of Pavlovian conditioning in two respects. With the notable exception of the study by LaBar and colleagues (LaBar et al., 1998), previous imaging studies have only acquired brain activation data from one phase of the conditioning procedure (i.e., acquisition), although most of them included a habituation and an extinction phase in their experimental protocol. In addition, prior conditioning studies have always used stable, nonchanging reinforcement contingencies of either 100% (Cheng et al., 2003; Knight et al., 2004; LaBar et al., 1998) or 50% partial reinforcement (Buchel et al., 1998, 1999) during acquisition. Our study, however, acquires neuroimaging data from different conditioning phases multiple times in combination with dynamically changing reinforcement contingencies. Although some of these studies did not observe amygdala activation (Knight et al., 1999, 2004), most studies report amygdala activation related to altered CS processing during acquisition (Buchel et al., 1998, 1999; LaBar et al., 1998; Morris et al., 2001) and extinction (LaBar et al., 1998) or to the expression of a conditioned skin conductance response (Cheng et al., 2003), even when the US is never presented (Phelps et al., 2001). Consistently, these studies have reported a decrease of amygdala activation in the course of the experiment, a pattern that has been interpreted as novelty or uncertainty coding (Davis and Whalen, 2001; Zald, 2003). Stated differently, in the context of formal learning theories, the more predictable a CS becomes (given stable contingencies), the less activation is observed in the amygdala, suggesting an involvement of this structure in encoding and processing contingency changes. The findings of our two experiments support this notion: the amygdala closely monitors contingency changes for biologically salient stimuli (i.e., face CSs). Because we wanted to systematically manipulate reinforcement contingencies, our phases of stable contingencies were necessarily short. Thus, the reason we did not observe decreases in activation during phases of stable contingencies might be that for the amygdala, which may encode contingency changes on a long-term basis, these phases of "stable" contingencies were not long enough to detect the stability.

It is interesting to note that the low learning rate was slightly higher in experiment 2 ($\epsilon = 0.15$) than in experiment 1 ($\epsilon = 0.05$). This could be an effect of the different cognitive tasks in the two experiments. The learning rate ϵ is also influenced by the associability of CS and US (Pearce and Bouton, 2001). In experiment 2, subjects were asked to submit online expectancy ratings of the US. The explicit nature of the task probably leads to an increased allocation of processing resources, which in turn could increase the associability of the CSs.

Short-Term Anticipation of Reinforcement in Perceptual Areas

In juxtaposition to the amygdala findings, we found circumscribed activations in stimulus-specific regions of the ventral visual stream (FFA/PPA) in both experiments. Strikingly, the patterns of parameter estimates of both experiments also closely correspond to each other.

Here, the pattern of the prediction modulation can be best described as short temporal integration (high ϵ) with a minimal attenuation in phases of stable contingencies. The modulation in these ventral stream areas reflects the outcome history of only the last one or two trials. This is accomplished by amplifying the influence of the prediction error on the update of the current prediction via a high learning rate. This short-term integration of recent reinforcement trials could provide the basis for an anticipatory signal that allocates more perceptual resources to ventral stream areas in order to process learned, biologically salient stimuli in more detail.

Neurons in the FFA have been reported to be modulated by attentional factors (Pessoa et al., 2002; Vuilleumier et al., 2001) and Pavlovian conditioning (Morris et al., 2001). The latter study reported learning-related signal increases in this area over time. Thus, during sustained aversive reinforcement of a face stimulus, the perceptual processing in the FFA is increased. Our pattern of prediction modulation in the FFA and PPA supports and extends these findings because it demonstrates the processing dynamics in situations of changing contingencies: only during phases of reinforcement is the activation elevated in these areas. Furthermore, our results suggest that the modulation of ventral stream areas takes place on a short time scale.

Formal Learning Theories in Neuroimaging

The application of formal learning theories that rest on prediction error as the reinforcement signal (Vogel et al., 2004) has been very successful in modeling behavioral (Miller et al., 1995) and neuronal responses in appetitive learning experiments (Schultz et al., 1997; Schultz and Dickinson, 2000). Within this framework, the neuronal coding of the (short-term) prediction error in appetitive learning has received much research attention (Schultz and Dickinson, 2000), and corresponding neuronal ensembles have been repeatedly located in the ventral tegmental area with projections to the ventral striatum in single-cell recordings in monkeys (Fiorillo et al., 2003; Schultz et al., 1997). Work akin to this using functional neuroimaging in humans reports activations localized in the ventral putamen that correlate with the prediction error (McClure et al., 2003; O'Doherty et al., 2003, 2004; Seymour et al., 2004).

While these studies greatly contribute to the identification of the neural correlates of prediction error, the focus of our investigation lies on the structures encoding the actual predictions. To our knowledge, only one other imaging study also tried to localize the anatomical structures encoding the stimulus predictions. In an elegant second-order aversive conditioning study employing a temporal differences learning model, Seymour and colleagues (Seymour et al., 2004) reported activation for the predictions in the ACC, anterior insula, and midcingulate cortex. These activations are in the vicinity of the foci we reported in experiment 1 (Figure S2, Table 1). Furthermore, our findings help to characterize the observed activation with different amounts of temporal integration.

The distinction between prediction error and predictions in formal learning theories is essential, as the former can be seen as the pacemaker of learning, whereas the latter actually computes and encodes the values of expected outcomes (Montague et al., 2004). However, although both are distinct computational concepts, they are also closely linked. Single-cell recordings and neuroimaging studies have identified the ventral striatum as the "seat" of these computational entities (McClure et al., 2003; O'Doherty et al., 2003; Schultz et al., 1997; Seymour et al., 2004). However, the delineation of the region with respect to predictions and their errors is far from complete. The nucleus accumbens is not only intimately connected to reward processing (Montague et al., 2004), but is also differentially activated during aversive Pavlovian conditioning (Jensen et al., 2003). Our finding of covariation of activity in the nucleus accumbens with slowly changing predictions (low ϵ) in experiment 1 can be seen as in line with these findings. Interestingly, however, we also observed the same covariation in the ventral putamen, which has been previously associated with prediction error (McClure et al., 2003; O'Doherty et al., 2003; Seymour et al., 2004). We do not have a strong opinion on whether the ventral putamen computes both the predictions and the error. It is, however, possible that these computations are accomplished by distinct neuronal subpopulations located in the same brain area which are not resolvable with fMRI.

All of the aforementioned studies employed the Temporal Difference (TD) model, a real-time expansion of the trial-based Rescorla-Wagner model used in the present study. Trial-based learning models like the RW model compute a single prediction and prediction error for each trial, while real-time models like the TD algorithm resolve predictions for each time point within a trial. We chose to model our data with the trial-based RW model because the main focus of the study lies in the different intervals of temporal integration in distinct brain regions rather than in resolving the time course of the activation during CS processing. Additionally, the rate of image acquisition in BOLD fMRI only allows for modeling a few time points in a TD model effectively (McClure et al., 2003; O'Doherty et al., 2003, 2004; Seymour et al., 2004). In fact, the predictions of both models converge with fewer time points modeled in a TD framework.

In summary, we found evidence for long temporal integration of the past learning experiences in the amygdala and short temporal integration in stimulus-specific ventral stream areas (FFA/PPA) that reflects the instantaneous reinforcement expectancy and selectively enhances perceptual processing. Crucially, in order to detect the modulatory influence of reinforcement contingencies on the BOLD signal, we employed a design with varying contingencies.

The parallelism and yet regional distinction of both slow and fast learning-related changes is the potential

neural correlate of the computational solution to two tasks that a behaving organism has to solve in order to be successful in evolutionary survival. Areas that encode the short-term reinforcement history enhance their perceptual processing and propagate an immediate anticipatory signal of upcoming outcomes (e.g., threats), which helps to initiate potentially defensive behavioral reactions. On the other hand, brain regions that integrate over a long period of the past learning history approximate the expected long-term average outcome (i.e., values) necessary for strategic behavior. Balancing the behavioral consequences of these two necessary computational goals could enhance the chances of evolutionary survival and procreation. Our findings further implicate that long-term changes in the acquired value are also encoded, thus allowing for the dynamic adaptation of goals.

Experimental Procedures

Conditioning Procedure

Experiment 1

We chose an implicit Pavlovian conditioning procedure (Carter et al., 2003) to investigate contingency related changes in brain activation. Subjects performed a simple working memory one-back task in which they had to monitor a stream of letters (presented for 1 s with a stimulus onset asynchrony [SOA] of 2 s) and respond with a button press if the same letter was presented twice in a row. One-back targets always appeared in the pause between the US and the following CS approximately every 20–30 s. In the meantime, one of the CSs was presented in the background for 6 s. We chose a jittered SOA of 12 ± 2 s for the CSs. A neutral facial expression drawn from the Ekman Series of Facial Affect (Ekman and Friesen, 1976) and a picture of a house were used as our two CSs. The laser pain stimulus (US), which had a duration of only 1 ms, coincided with the offset of the CS, thus rendering our design as an implicit delay conditioning paradigm.

Experiment 2

In order to obtain stronger behavioral data, we chose an explicit conditioning procedure that renders a direct measure of stimulus prediction. Subjects were asked to give a binary rating of US expectancy via a button press within the first 2 s of the presentation of each CS. For CSs, we used the same visual stimuli as in experiment 1 and presented them also for 6 s (jittered SOA of 12 ± 2 s). However, we chose a heat pain stimulus delivered via a thermode attached to the inside of the subjects' left forearm as a more effective US.

Contingency Variation

In order to demonstrate covariation between brain activation in specific regions and constantly updated stimulus prediction, we manipulated CS-US contingencies by increasing and decreasing the probability of US occurrence in a systematic way. Thus, in a first step we created a pseudorandom CS event train with the restriction that each CS type (face or house) could appear only on two successive trials.

The dynamically increasing and decreasing contingencies were reflected in a low-frequency sine wave (1.75 cycles) for each CS type that spanned across all trials of the experiment (see blue and red curves in Figures 1A and 1B [experiment 1] and Figures 4A and 4B [experiment 2]). In experiment 1, the phase of the two sine waves was shifted against each other by 180°, while in experiment 2, the phase shift was 90°. In addition, we added a 50% partial reinforcement period prior to the systematic contingency variation in experiment 2 in order to alleviate cognitive search processes for the reinforcement schedule. The aforementioned sines then served as threshold functions describing the probability of reinforcement at each trial. To apply this continuous contingency sine wave to the discrete event train, we drew a random number in the amplitude range of the sine wave for each trial. If that random number fell below the threshold function (i.e., within the gray area in Figures 1A and 1B), then the trial was assigned to be reinforced CS (CS+); if the random number fell above the threshold function it was assigned to function as a nonreinforced CS (CS–). This resulted in experimental phases in which the face CS was more consistently reinforced, while the house CS remained nonreinforced and vice versa. We carefully controlled that each CS type (face or house) was followed by an equal number of USs (laser or thermode pain) during the entire experiment. The actual event train for experiment 1 is indicated in Figure 1. Note that there was no abrupt (binary) change between CS+ and CS– phases as occurs when switching from acquisition to extinction; rather, a transitional phase in which both CS+ and CS– trials conforming to the abovementioned selection scheme bridged the two phases of repeated CS+ or CS– trials.

While we used the same event train and CS assignment (blue, face; red, house) in experiment 1, this was changed in experiment 2: computer-generated event trains for each subject were similar but not identical, and the CS assignment was counterbalanced across subjects. Figures 4A and 4B represent an example event train from one of the subjects. In order to evaluate how close each individual event train matched the idealized contingency sine waves, we smoothed the event train with a five-trial full-width-at-half-maximum (FWHM) Gaussian filter (Figures 4A and 4B, dashed lines) and calculated the sum of squares (SSQ) between this smoothed event train and the contingency function. We chose only those event trains that rendered a SSQ < 1.5 for each of the CSs.

Rescorla-Wagner Model

We chose to model our data with the trial-based Rescorla-Wagner model because we were primarily interested in the effects of different learning rates expressed in the model parameter ϵ. Based on the actual event train in Figures 1 and 4, the RW model allows the estimation of constantly updated outcome predictions that can be subsequently used to model the data. According to the RW model, the predicted outcome V_t is calculated with the following equations:

$$V_t = w_{t-1} \times u_t$$

and

$$w_t = w_{t-1} + \epsilon(R_t - V_t) \times u_t$$

where V_t indicates the predicted outcome of trial t, u_t indicates the CS type at trial t (face or house) and can be either 1 or 0 depending if the particular CS type is presented on trial t, R_t indicates the actual outcome of trial t, w_t indicates the change in prediction at trial t due to the prediction error at trial t ($R_t - V_t$), and ϵ indicates the learning parameter that controls the influence of the prediction error in the update of the prediction (Dayan and Abbott, 2001).

We sought to determine the optimal learning parameter of a particular brain region. This can be accomplished by nonlinear optimization or by densely sampling the parameter space of ϵ. We chose the latter approach because it is more feasible within the GLM framework as employed by SPM2. Thus, we conducted analyses for a range of ϵ (0 to 1 in steps of 0.05) and identified the ϵ for which the mean effect size (across subjects, one-sample t test) was maximal. This revealed that the maximal effect sizes for activated voxels were reached at either very low (ϵ = 0.05) or very high (ϵ = 0.95) learning rates in experiment 1. Consequently, we used these two learning rates for further testing (see below). For experiment 2, these maximal effect sizes were reached at ϵ = 0.15 and 0.95. The predicted outcomes for each CS type (face, blue; house, red) for low and high learning parameters are shown in Figures 1 and 4 (C-F) for both experiments.

Experimental Procedure

Experiment 1

Subjects were instructed that they would participate in a working memory task in which they were distracted either by visual stimuli appearing in the background or sometimes by painful stimuli applied to the top of their left hand. It was explained to them that the goal of the study was to monitor their brain activations during the task and that they should therefore pay close attention to the working memory task. Subjects completed the fMRI session in 19 min.

After scanning, subjects were confronted with a postexperimental questionnaire testing for their conscious knowledge of the stimulus contingencies, first under free and then under cued recall conditions (see Supplemental Experimental Procedures for more details).

Experiment 2

Subjects were instructed that they would participate in an experiment that investigates the effects of expectancies on subsequent pain perception and the related brain activation. While it was stressed that they should provide expectancy ratings for every trial, it was not mentioned to them that they were part of a learning/conditioning study. Subjects completed the fMRI session in 24 min.

Subjects

In experiment 1, we scanned 15 right-handed male subjects (mean age: 25.3 ± 2.8 SD). One subject was excluded from the analysis due to a misunderstanding of task instructions. In experiment 2, we scanned 17 subjects (8 males, mean age: 24.9 ± 2.5 SD). All subjects were free of neurological or psychiatric diseases, had normal or corrected-to-normal vision, and signed a consent statement that was approved by the local ethics committee.

Data Acquisition

For skin conductance recordings in experiment 1, we used Ag/AgCl electrodes attached at the hypothenar site of the subject's left palm. Both electrodes were placed on the same dermatome (C8) in order to control for potential recording differences between dermatomes. The signal was amplified using a CED 2502 skin conductance amplifier and then digitized at 100 Hz on a CED Micro1401 mkII (Cambridge, UK) physiological recording unit.

Functional imaging in both experiments was performed on a 3.0T Siemens Trio scanner (Erlangen, Germany). 42 transversal slices of echo-planar (EPI) T_2* weighted images were acquired in each volume with a slice thickness of 2 mm and 1 mm gap (TR = 2410 ms, TE = 25 ms, flip angle 80°, FoV 210 mm^2, matrix 64 × 64). Experiment 1 was completed in a single session with 472 volumes acquired in 19 min. Durations of experiment 2 varied between 24.7 and 25.5 min (615 to 636 scans) due to the individual event trains and jittered SOAs.

Data Processing

SCR data of experiment 1 were resampled to 10 Hz, low-pass filtered (3 s cut-off), and mean corrected. We then calculated the second interval response (SIR) by subtracting the mean of 3 s prestimulus baseline from the maximum SC deflection of the second half of the CS (3 s window). Previous studies have reported that the SIR is affected by contingency variations in the course of learning experiments (Knight et al., 2003; Wolter and Lachnit, 1993).

Image processing and statistical analyses of both experiments were carried out using SPM2. Prior to image processing, we discarded the first four images to alleviate the scan equilibration effect. All volumes were realigned to the first volume, spatially normalized to a standard EPI template (Friston et al., 1995) using 3rd degree B-spline interpolation, and finally smoothed with an isotropic 10 mm full-width-at-half-maximum Gaussian filter to account for anatomical differences between subjects and to allow for valid statistical inference at the group level.

Statistical Analysis of Skin Conductance Data—Experiment 1

The demonstration of differential conditioning (CS+ versus CS–) after the acquisition phase is a widely accepted index for learning in a Pavlovian conditioning experiment (Knight et al., 2003; Ohman and Soares, 1994, 1998). However, in a design with varying contingencies, choosing the appropriate stimuli for the comparison is more difficult. Thus, we compared the SIRs of those SCRs in which CS-US contingency was high against those SIRs in which the CS-US contingency was low. We defined the thresholds for "high" and "low" contingencies based on the range of outcome predictions (ϵ = 0.05) derived by the RW model (Figures 1A and 1B) and selected those responses as high contingency CSs whose prediction value was above the 75th percentile of the range. Similarly, those responses whose prediction value fell below the 25th percentile were assigned to the low contingency class.

Statistical Analysis of Behavioral Data—Experiment 2

In order to closely match the SCR analysis of experiment 1, we chose a similar group approach for the behavioral data of experiment 2. Thus, we compared the expectancy ratings of trials with high CS-US contingency with those of low contingency using the same cut-off thresholds as in experiment 1. Thresholds were based on the outcome prediction at ϵ = 0.15 to match the learning rate used in the analysis of the imaging data. However, in the attempt to demonstrate the behavioral relevance of the activations found in the imaging analysis, we selected only those subjects who showed clear evidence of learning (i.e., a positive correlation between expectancy ratings and the reinforcement regime).

Statistical Analysis of Imaging Data

Data analysis in both experiments was performed using a general linear model as implemented in SPM2. We targeted our analysis to the detection of brain regions that covary with the outcome prediction of different learning rates ϵ. Hence, we included the optimal high and low ϵ RW predictions (as determined from our exploratory analysis above) as parametric modulations in the design matrix.

Design matrices at the single-subject level were created by convolving the stimulus function (i.e., onset of each CS type [face/house], each US [laser or thermode pain], and only for experiment 1 each target in the one-back WM task) with a synthetic hemodynamic response function (HRF) that models the BOLD effect (Friston et al., 1998). We then entered the predictions of the RW model for a low and a high ϵ (the blue and red curves in Figures 1 and 4, C-F) as parametric modulations for each CS type separately. These parametric modulations were also convolved with the HRF. These regressors model BOLD signal changes that covary with the outcome predictions derived from the RW model at a particular ϵ. In addition, we included the rigid-body scan to scan movement parameters from the realignment stage that model the subject's head movement during the experiment.

Subsequently, we raised six first level effect size images of each subject to a repeated-measures ANOVA at the second level for experiment 1. These images were main effects for each CS type (presentation of each CS) and parametric modulators for the low and the high learning rate for each CS type. In experiment 2, we raised the same three effect size images of those CSs to the second level, for which subjects demonstrated a behavioral learning effect (see above). Seven subjects learned both CSs (face and house), two additional subjects learned only the face CS, and three additional subjects learned only the house CS. Thus, the analysis comprised data from 12 subjects. The analyses in both experiments were appropriately corrected for potential nonspherical distribution of the error term.

In order to show regions that exhibit long temporal integration, we then created the differential t contrasts that compared low > high ϵ regressors. Similarly, we also created the differential t contrast of high > low ϵ regressors to detect those regions of short temporal integration. Our statistical threshold was set at $p < 0.05$, corrected for multiple comparisons using false discovery rate (Genovese et al., 2002). For displaying purposes in Figure 3 and 5 and Figures S1 and S2, the statistical maps were thresholded at $p < 0.01$.

For regions with apriori hypotheses, we applied reduced spherical search volumes that approximated the volumetric extent of the target area centered on coordinates obtained from other conditioning studies or by anatomical definition. The activation in the amygdala was corrected with a standardized anatomical mask (Tzourio-Mazoyer et al., 2002). The activations in the FFA and PPA were corrected for a spherical volume of 10 mm. Sphere centers were determined by the face > house contrast and the house > face contrast, respectively (see Supplemental Experimental Procedures and Table S1). The hippocampus, ventral striatum (nucleus accumbens), anterior insula, and orbitofrontal cortex were corrected for an 8 mm spherical search volume centered on coordinates obtained from Ploghaus et al. (2000), Jensen et al. (2003), Porro et al. (2002), and by anatomical definition, respectively. Activation within the red nucleus was corrected with a 6 mm spherical volume centered on coordinates obtain from Buchel et al. (1998). The activation in cingulate and somatosensory cortices was corrected for a

spherical volume of 10 mm radius. Centers were obtained by locating the tip of the genu of the corpus callosum for the perigenual ACC and the somatosensory cortex opposite the hand knob in the motor cortex (Pizella et al., 1999; Yousry et al., 1997).

Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/47/2/295/DC1/.

Acknowledgments

We thank Tobias Sommer, Michael Rose, Adam McNamara, and Eszter Schoell for helpful suggestions on earlier drafts of this manuscript; Ulrike Bingel for help with the laser stimulation; and the Physics and Methods group at Neuroimage Nord. J.G. is supported by the Studienstiftung des Deutschen Volkes. C.B. is supported by the BMBF and the Volkswagen Stiftung. The authors declare that they do not have any competing financial interest.

Received: September 20, 2004
Revised: March 22, 2005
Accepted: June 2, 2005
Published: July 20, 2005

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behav. Cogn. Neurosci. Rev. *1*, 21–62.

Buchel, C., Morris, J., Dolan, R.J., and Friston, K.J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron *20*, 947–957.

Buchel, C., Dolan, R.J., Armony, J.L., and Friston, K.J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. J. Neurosci. *19*, 10869–10876.

Cardinal, R.N., Parkinson, J.A., Hall, J., and Everitt, B.J. (2002). Emotion and motivation: the role of the amygala, ventral striatum, and prefrontal cortex. Neurosci. Biobehav. Rev. *26*, 321–352.

Carter, R.M., Hofstotter, C., Tsuchiya, N., and Koch, C. (2003). Working memory and fear conditioning. Proc. Natl. Acad. Sci. USA *100*, 1399–1404.

Cheng, D.T., Knight, D.C., Smith, C.N., Stein, E.A., and Helmstetter, F.J. (2003). Functional MRI of human amygdala activity during Pavlovian fear conditioning: stimulus processing versus response expression. Behav. Neurosci. *117*, 3–10.

Davis, M., and Whalen, P.J. (2001). The amygdala: vigilance and emotion. Mol. Psychiatry *6*, 13–34.

Dayan, P., and Abbott, L.F. (2001). Theoretical Neuroscience (Cambridge, MA: MIT Press).

Eichenbaum, H., and Cohen, N.J. (2001). From Conditioning to Conscious Recollection (New York: Oxford University Press).

Ekman, P., and Friesen, W. (1976). Pictures of Facial Affect (Palo Alto, CA: Cunsulting Psychologists Press).

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? Neuron *23*, 115–125.

Esteves, F., Parra, C., Dimberg, U., and Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. Psychophysiology *31*, 375–385.

Falls, W.A., and Davis, M. (1995). Behavioral and physiological analysis of fear inhibition. In Neurobiological and Clinical Consequences of Stress: From Normal Adaptation to Post-Traumatic Stress Disorder, M.J. Friedman, D.S. Charney, and A.Y. Deutch, eds. (New York: Lippincott-Raven), pp. 177–202.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science *299*, 1898–1902.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., and Frackowiak, R.S. (1995). Spatial registration and normalization of images. Hum. Brain Mapp. *2*, 1–25.

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

Genovese, C.R., Lazar, N.A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using false discovery rate. Neuroimage *15*, 870–878.

Glascher, J., Tuscher, O., Weiller, C., and Buchel, C. (2004). Elevated responses to constant facial emotions in different faces in the human amygdala: an fMRI study of facial identity and expression. BMC Neurosci. *5*, 45.

Gottfried, J.A., O'Doherty, J., and Dolan, R.J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. J. Neurosci. *22*, 10829–10837.

Jensen, J., McIntosh, A.R., Crawley, A.P., Mikulis, D.J., Remington, G., and Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. Neuron *40*, 1251–1257.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302–4311.

Knight, D.C., Smith, C.N., Stein, E.A., and Helmstetter, F.J. (1999). Functional MRI of human Pavlovian fear conditioning: patterns of activation as a function of learning. Neuroreport *10*, 3665–3670.

Knight, D.C., Nguyen, H.T., and Bandettini, P.A. (2003). Expression of conditional fear with and without awareness. Proc. Natl. Acad. Sci. USA *100*, 15280–15283.

Knight, D.C., Cheng, D.T., Smith, C.N., Stein, E.A., and Helmstetter, F.J. (2004). Neural substrates mediating human delay and trace fear conditioning. J. Neurosci. *24*, 218–228.

LaBar, K.S., Gatenby, J.C., Gore, J.C., LeDoux, J.E., and Phelps, E.A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. Neuron *20*, 937–945.

LeDoux, J.E. (2000). Emotion circuits in the brain. Annu. Rev. Neurosci. *23*, 155–184.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. Annu. Rev. Neurosci. *24*, 897–931.

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

Miller, R.R., Barnet, R.C., and Grahame, N.J. (1995). Assessment of the Rescorla-Wagner model. Psychol. Bull. *117*, 363–386.

Montague, P.R., Hyman, S.E., and Cohen, J.D. (2004). Computational roles for dopamine in behavioural control. Nature *431*, 760–767.

Morris, J.S., Buchel, C., and Dolan, R.J. (2001). Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. Neuroimage *13*, 1044–1052.

Myers, K.M., and Davis, M. (2002). Behavioral and neural analysis of extinction. Neuron *36*, 567–584.

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

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452–454.

Ohman, A., and Soares, J.J. (1994). "Unconscious anxiety": phobic responses to masked stimuli. J. Abnorm. Psychol. *103*, 231–240.

Ohman, A., and Soares, J.J. (1998). Emotional conditioning to masked stimuli: expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. J. Exp. Psychol. Gen. *127*, 69–82.

Pavlov, I.P. (1927). Conditioned Reflexes (London: Oxford University Press).

Pearce, J.M., and Bouton, M.E. (2001). Theories of associative learning in animals. Annu. Rev. Psychol. *52*, 111–139.

Pessoa, L., McKenna, M., Gutierrez, E., and Ungerleider, L.G. (2002). Neural processing of emotional faces requires attention. Proc. Natl. Acad. Sci. USA *99*, 11458–11463.

Phelps, E.A., O'Connor, K.J., Gatenby, J.C., Gore, J.C., Grillon, C., and Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. Nat. Neurosci. *4*, 437–441.

Pizella, V., Tecchio, F., Romani, G.L., and Rossini, P.M. (1999). Functional localization of the sensory hand area with respect to the motor central gyrus knob. Neuroreport *10*, 3809–3814.

Ploghaus, A., Tracey, I., Clare, S., Gati, J.S., Rawlins, J.N., and Matthews, P.M. (2000). Learning about pain: the neural substrate of the prediction error for aversive events. Proc. Natl. Acad. Sci. USA *97*, 9281–9286.

Ploghaus, A., Narain, C., Beckmann, C.F., Clare, S., Bantick, S., Wise, R., Matthews, P.M., Rawlins, J.N., and Tracey, I. (2001). Exacerbation of pain by anxiety is associated with activity in a hippocampal network. J. Neurosci. *21*, 9896–9903.

Porro, C.A., Baraldi, P., Pagnoni, G., Serafini, M., Facchin, P., Maieron, M., and Nichelli, P. (2002). Does anticipation of pain affect cortical nociceptive systems? J. Neurosci. *22*, 3206–3214.

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

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. Annu. Rev. Neurosci. *23*, 473–500.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., and Frackowiak, R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429*, 664–667.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and 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.

Vogel, E.H., Castro, M.E., and Saavedra, M.A. (2004). Quantitative models of Pavlovian conditioning. Brain Res. Bull. *63*, 173–202.

Vuilleumier, P., Armony, J.L., Driver, J., and Dolan, R.J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron *30*, 829–841.

Wolter, J., and Lachnit, H. (1993). Are anticipatory first and second interval skin conductance responses indicators of predicted aversiveness? Integr. Physiol. Behav. Sci. *28*, 163–166.

Yousry, T.A., Schmid, U.D., Alkadhi, H., Pernaud, A., Buettner, A., and Winkler, P. (1997). Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. Brain *120*, 141–157.

Zald, D.H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. Brain Res. Brain Res. Rev. *41*, 88–123.