

Neurobiological Mechanisms Underlying the Blocking Effect in Aversive Learning

Falk Eippert, Matthias Gamer, and Christian Büchel

Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

Current theories of classical conditioning assume that learning depends on the predictive relationship between events, not just on their temporal contiguity. Here we employ the classic experiment substantiating this reasoning—the blocking paradigm—in combination with functional magnetic resonance imaging (fMRI) to investigate whether human amygdala responses in aversive learning conform to these assumptions. In accordance with blocking, we demonstrate that significantly stronger behavioral and amygdala responses are evoked by conditioned stimuli that are predictive of the unconditioned stimulus than by conditioned stimuli that have received the same pairing with the unconditioned stimulus, yet have no predictive value. When studying the development of this effect, we not only observed that it was related to the strength of previous conditioned responses, but also that predictive compared with nonpredictive conditioned stimuli received more overt attention, as measured by fMRI-concurrent eye tracking, and that this went along with enhanced amygdala responses. We furthermore observed that prefrontal regions play a role in the development of the blocking effect: ventromedial prefrontal cortex (subgenual anterior cingulate) only exhibited responses when conditioned stimuli had to be established as nonpredictive for an outcome, whereas dorsolateral prefrontal cortex also showed responses when conditioned stimuli had to be established as predictive. Most importantly, dorsolateral prefrontal cortex connectivity to amygdala flexibly switched between positive and negative coupling, depending on the requirements posed by predictive relationships. Together, our findings highlight the role of predictive value in explaining amygdala responses and identify mechanisms that shape these responses in human fear conditioning.

Introduction

In contrast to early accounts of classical conditioning (Pavlov, 1927), which assumed that the temporal contiguity of conditioned stimulus (CS) and unconditioned stimulus (US) drives learning, current theories highlight the predictive or informational relationship between CS and US (Rescorla and Wagner, 1972; Mackintosh, 1975; Pearce and Hall, 1980). This is most clearly demonstrated by the blocking paradigm (Kamin, 1968): when stimulus X is presented together with stimulus A and their presentation is paired with the US, learning to X will be strongly reduced if A has previously been paired with the US. Thus, despite being contiguous with the US, learning to X is blocked because A has already been established as a reliable predictor of the US, and X is therefore redundant.

The blocking effect can be observed in various domains of human learning, such as causal learning (Dickinson et al., 1984), eyeblink conditioning (Martin and Levey, 1991), and spatial learning (Prados, 2011), and has also informed neurobiological

accounts of learning in animals (Kim et al., 1998). In appetitive learning, neural signatures of reward processing, such as responses of dopamine neurons in the monkey midbrain (Waelti et al., 2001) and blood oxygenation level-dependent (BOLD) responses in the human ventral striatum (Tobler et al., 2006) have been found to be sensitive to blocking effects. In aversive learning, such a precise neurobiological characterization has not been established: blocking effects have been demonstrated behaviorally in humans (Hinchey et al., 1995) and have been challenged pharmacologically in rats (McNally et al., 2004), but there is no direct evidence that responses of the amygdala—a structure centrally implicated in conditioned fear (Davis and Whalen, 2001; Maren and Quirk, 2004; Öhman, 2009; Johansen et al., 2011)—are sensitive to blocking effects. Such a characterization would be valuable though, as it would highlight the amygdala's sensitivity to the informational content or predictive value of conditioned stimuli in aversive learning.

We therefore set out to probe responses of the human amygdala by using an aversive blocking paradigm (Fig. 1) in combination with functional magnetic resonance imaging (fMRI). In contrast to previous fMRI studies that demonstrated amygdala involvement in the acquisition of conditioned fear (Büchel et al., 1998; LaBar et al., 1998; Bach et al., 2011) or in the tracking of varying contingencies in aversive learning paradigms (Gläscher and Büchel, 2005; Schiller et al., 2008; Li et al., 2011), our paradigm allowed us to manipulate the predictive relationship between CS and US while keeping the contiguity (i.e., pairing) the same. We expected that amygdala responses would be sensitive to blocking effects (Fanselow, 1998), but also tested for such effects in cortical regions implicated in the

Received March 11, 2012; revised July 19, 2012; accepted July 30, 2012.

Author contributions: F.E. and C.B. designed research; F.E. performed research; F.E., M.G., and C.B. analyzed data; F.E., M.G., and C.B. wrote the paper.

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB TRR58, TP B03). We thank Wolfram Schultz and Philippe N. Tobler for allowing us to use and providing us with the visual stimuli they have used in previous studies on the blocking effect. We also thank Michael Rose for technical help during pilot studies; Timo Krämer, Katrin Müller, and Kathrin Wendt for help with MR data acquisition; and Corinna Klinge for helpful comments on previous versions of this manuscript.

Correspondence should be addressed to Falk Eippert, Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. E-mail: f.eippert@uke.uni-hamburg.de.
DOI:10.1523/JNEUROSCI.1210-12.2012

Copyright © 2012 the authors 0270-6474/12/3213164-13\$15.00/0

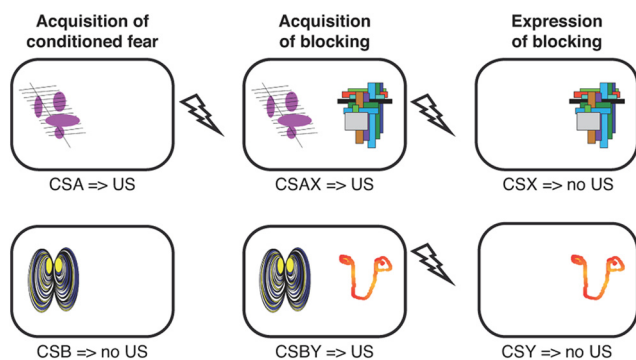


Figure 1. Experimental paradigm, consisting of three phases. Abstract visual stimuli (duration 5 s) served as CSs, and electrical stimulation at the end of the CS served as USs. The structure of the paradigm is similar to Tobler et al. (2006), who studied blocking in appetitive learning. In the first phase (acquisition of conditioned fear), CSA and CSB were presented pseudorandomly, and presentation of CSA was paired with US administration on each trial. In the second phase (acquisition of blocking), compound stimuli were pseudorandomly presented, which consisted of (1) the shock-predictive stimulus CSA plus an additional new stimulus CSX and (2) the non-predictive stimulus CSB plus an additional new stimulus CSY; both compound stimuli were paired with the US on each trial. According to the logic underlying the blocking effect, participants should learn that CSX is redundant, as CSA already predicts the US. This was tested in the third phase (expression of blocking), where CSX and CSY were pseudorandomly presented alone (i.e., as single stimuli; and were not paired with the US). Note that these stimuli differed strongly in their predictive value regarding US administration (CSX, low; CSY, high), although they had received exactly the same pairing with the US. During the second and third phase, trial types of the previous phases were also presented to maintain previous associations. Participants had to perform a reaction time task on each trial and were also asked at several time points during the experiment to give ratings regarding the fear/stress/tension the CS elicited in them; shock-expectancy ratings were acquired after the experiment. In addition to fMRI data, we also collected skin-conductance, heart-rate, and eye-tracking data on each trial.

acquisition of conditioned fear [dorsal anterior cingulate cortex (dACC) and anterior insula (AI); Sehlmeier et al., 2009]. Importantly, our paradigm also allowed us to investigate the development of the blocking effect, where we not only tested for attentional mechanisms (Beesley and Le Pelley, 2011), but also for involvement of prefrontal regions, given their role in associative learning (Sehlmeier et al., 2009) and in shaping amygdala responses (Phelps, 2009).

Materials and Methods

Participants. Data from 32 healthy male participants (mean age: 26.4 years; age range: 19–37 years) were analyzed in this study; 34 participants were originally scanned, but data from 2 participants were excluded, as they were not contingency aware. Trait anxiety scores (31.5 ± 1.2 ; State-Trait Anxiety Inventory; Spielberger et al., 1970) and social desirability scores [13.0 ± 0.70 ; Social Desirability Scale; Crowne and Marlowe, 1960; Lück and Timaeus, 1969 (abridged German version)] of our sample differed less than 1 SD from the mean of the norm sample. The Ethics Committee of the Medical Board in Hamburg approved the study, and all participants gave written informed consent.

Study design. We used a within-subject blocking design (Rescorla, 1981) (Fig. 1) that has been used in several previous studies [e.g., eyeblink conditioning (Martin and Levey, 1991) and appetitive conditioning (Waelti et al., 2001; Tobler et al., 2006)]. The paradigm consisted of three phases, in each of which participants were presented with visual stimuli that served as CSs and painful electrical stimuli that served as USs. The CSs we used were four abstract colored forms on a white background, which had been used in previous studies on the blocking effect (Waelti et al., 2001; Tobler et al., 2006). Each of the four visual stimuli was assigned to one of the four CS types (CSA, CSB, CSX, CSY; see below), with the assignment of picture to condition being randomized across participants. Electrical stimulation was performed using a DS7A stimulator (Digitimer), and consisted of a train of three 2 ms square-wave pulses (separated by 50 ms), which were delivered to the back of the right hand

via a surface electrode with platinum pin (WASP electrode, Specialty Developments). The intensity was adjusted individually for each participant before the experiment, with the aim of reaching a moderate level of pain [corresponding to a rating of 5 on a numerical rating scale (NRS) ranging from 0 (“no pain”) to 10 (“worst pain you would be willing to tolerate in this experiment”)]. Each trial consisted of the presentation of the CS for 5 s, the delivery or omission of the US at 4.7 s after CS start, and an intertrial interval uniformly varying between 5 and 10 s, during which a white crosshair was displayed.

In the first phase of the experiment, only CSA (100% reinforcement) and CSB (0% reinforcement) were presented 10 times each, in order for participants to learn that CSA predicted the US, whereas CSB did not. In the second phase of the experiment, CSA and CSX were presented together 10 times, as were CSB and CSY. Both compounds (CSAX and CSBY) received 100% reinforcement. CSA (100% reinforcement) and CSB (0% reinforcement) were also presented separately in this phase (each five times) to maintain the original associations (i.e., CSA => US, CSB => no US). It is during this second phase that the blocking effect should develop, as CSA already predicts the US (due to their pairing in the previous phase), meaning that when CSX is presented together with CSA in the CSAX compound, CSX is essentially redundant, as it has no predictive value for US occurrence. As a control condition, we presented CSY together with CSB in the CSBY compound: in contrast to CSX, CSY is not redundant, as it now (as part of the CSBY compound) comes to predict the US (note that our control condition corresponds to reduced overshadowing; Hinchey et al., 1995; Vandorpe and De Houwer, 2005). Thus, learning to CSX should be blocked, whereas learning to CSY should proceed, although both stimuli are paired with the US exactly the same number of times. This assumption was tested in the third phase of the experiment, where CSX and CSY were presented alone (each 20 times; i.e., as single stimuli with 0% reinforcement). Their presentation was intermixed with presentations of CSA and CSB (5 times each; 100 and 0% reinforcement, respectively), as well as CSAX and CSBY (10 times each, 100% reinforcement for both), again to maintain previous associations. At the end of the experiment, each condition had been presented 20 times. Throughout all phases, the conditions were presented in a pseudorandomized order with the constraint that no condition could occur more than two times consecutively. The location where stimuli were presented was also randomized: while the presentation of a single stimulus (conditions CSA, CSB, CSX, and CSY) could occur in each corner of the screen, the presentation of two stimuli (conditions CSAX and CSBY) would always occur on one side of the screen, with the vertical position of each part of the compound stimuli randomized.

Throughout the experiment participants had to perform a simple reaction time task, which allowed us to assess their level of vigilance: on each trial they had to indicate via a button press whether the CS appeared on the left or right side of the screen. Additionally, they had to rate the level of fear/stress/tension they were experiencing when exposed to each CS. These ratings were given on a visual analog scale (VAS; range, 0–100; anchors “Inexistent” and “Very strong”, respectively) at seven time points over the course of the experiment: at the start of the experiment, in the middle of each phase, between each of the phases, and at the end of the experiment. Participants practiced the reaction time task and the rating during several practice trials, in which no shock was administered, and in which visual stimuli different from the ones in the experiment were used. Participants were told that their response speed in the reaction time task would not have any relation to the delivery of the shock (i.e., no instrumental contingency) and were furthermore asked to fixate the crosshair during the intertrial interval and to look at the pictures when these were presented. After the paradigm ended—but while they were still in the scanner—participants’ awareness of the contingency was assessed. They were presented with each of the stimuli seen during the experiment and were asked to indicate on a VAS whether they had received a shock when a specific stimulus was presented (VAS range, 0–100; anchors “Certain, that no shock” and “Certain, that shock”, respectively; additionally “Don’t know” was presented in the middle, i.e., at VAS value of 50).

To achieve a reliable blocking effect, we followed the suggestions of previous behavioral studies that singled out possible contributing factors

(see also Shanks, 2010). First, we aimed to induce an elemental instead of a configural mode of processing in our participants (Melchers et al., 2008) by maximizing the spatial distance of the CSs that were presented together in the compound trials, as suggested by previous causal learning studies (Glautier, 2002; Livesey and Boakes, 2004). Second, we aimed to make the transition between the different phases rather smooth, as suggested by a previous conditioning study (Hinchey et al., 1995), to which the following three factors were thought to contribute: (1) the fear ratings that occurred between the phases also occurred within each phase; (2) conditions from the previous phase were intermixed with the new conditions; and (3) each phase started with the presentation of old conditions before new conditions were introduced. Finally, we also subjected our participants to additivity and submaximality manipulations, which have been shown to support a robust blocking effect in both conditioning and causal learning paradigms (Mitchell and Lovibond, 2002; Beckers et al., 2005): before the start of the actual fMRI experiment, participants were presented with two visual stimuli (different from those used in the experiment), each of which was separately presented for 5 s and terminated with a painful shock (the strength of which corresponded to a rating of 5 on the NRS). Importantly, after each of the two stimuli had been presented once, they were presented together and participants received a shock that was perceived as twice as painful; this was the only time this level of shock was administered [its strength was determined during the earlier individual stimulation adjustment, where—after the current level for a rating of 5 was reached (9.6 ± 1.4 mA)—we increased the stimulation and asked participants to indicate when the stimulation would be perceived to be as approximately twice as painful (18.9 ± 3.4 mA)]. Due to these three trials, participants should thus have learned that (1) the outcome of conditioned stimuli is additive (i.e., if each of two single stimuli leads to a moderate shock, their combined presentation as a compound stimulus will lead to a strong shock; additivity manipulation) and (2) that shocks stronger than level 5 could be administered during the experiment (submaximality manipulation). During the actual experiment, participants can thus conclude that CSX is not predictive for the US, as the US paired with compound CSAX is of the same moderate strength as the US paired with CSA (Beckers et al., 2005). While additivity and submaximality effects are intertwined in our manipulation, it has been shown that they exert a separable positive influence on blocking in causal learning (Beckers et al., 2005). It is important to note that we do not wish to make any claims on whether a causal inferential reasoning or an associative learning account is the most parsimonious explanation for the blocking effect (Shanks, 2010): while a causal inferential reasoning interpretation could be suspected based on our description of the additivity manipulation (Beckers et al., 2006), it is important to realize that associative learning models are also able to explain additivity effects in blocking (Haselgrove, 2010).

Data acquisition. We used Presentation software (Neurobehavioral Systems) for stimulus presentation and recording of reaction times and ratings. Skin conductance responses (SCRs) were acquired with MRI-safe electrodes (2700 CLEARTRACE2, ConMed) attached to the hypothenar eminence of the participants' left hand, using a CED 2502, a CED micro1401 mkII, and Spike2 software (all equipment by Cambridge Electronic Design). Due to technical problems, we were not able to acquire SCR data from six participants. The stimuli were projected onto a translucent screen at the end of the scanner bore using a video projector, and participants viewed this screen via a dichroic reflector that allowed infrared light to pass while reflecting the visible light from the video projector. Eye-tracking data were acquired using a 60 Hz MRI-compatible eye-tracking camera (MRC Systems) that was hidden behind the mirror and mounted directly in front of the right eye of the participant. Eye movements were recorded using the software iView X (Sensor-Motoric Instruments).

fMRI data were acquired on a 3 tesla system (Magnetom Trio, Siemens) equipped with a 32-channel head coil. Forty-two transversal slices (slice thickness, 2 mm; gap, 1 mm) were acquired in each volume (repetition time, 2.48 s; echo time, 25 ms; flip angle, 80°; field of view, 208×208 mm; in-plane resolution, 2×2 mm; GRAPPA with PAT-factor 2) using T2*-weighted echoplanar imaging (EPI). Slice orientation was tilted by 30° from transverse to coronal to optimize signal in

susceptibility-prone frontal regions (Deichmann et al., 2003). The first five volumes were discarded to allow for T1 saturation. After having obtained the functional data, we also acquired high-resolution ($1 \times 1 \times 1$ mm voxel size) T1-weighted images for each participant using an MP-RAGE sequence (in one participant, the high-resolution image could not be acquired due to technical problems). To allow for retrospective physiological noise correction of fMRI data, we also acquired pulse and respiration data, using the vendor-supplied pulse sensor and respiratory bellows; data from the pulse sensor were also used for investigating heart-rate (HR) changes. Due to technical problems, we were not able to acquire pulse and respiration data from four participants.

Behavioral data analysis. Behavioral data analysis was performed with Matlab (Mathworks) and SPSS (IBM), using a threshold of $p = 0.05$, one tailed in cases of directed hypotheses. Note that only the first 10 presentations of each CS were considered in all the trial-wise analyses (i.e., eye-tracking, SCR, HR, and fMRI data), as this had two advantages: (1) responses to the CS could not be influenced by other trial types in the following phases (i.e., no contamination of responses to CSA and CSB by presentation of CSAX and CSBY); and (2) extinction had not progressed very far for CSX and CSY in phase 3.

Fear ratings to CSA and CSB were analyzed with a 2×7 repeated-measures ANOVA with factors condition (CSA, CSB) and time (seven time points); the same analysis was performed for ratings of CSY and CSX. In the case of nonsphericity (assessed with Mauchly's test of sphericity), Greenhouse–Geisser corrected results are reported. *Post hoc* paired *t* tests were then used to compare responses to CSA and CSB (CSY and CSX) at each of the seven time points, the results of which were subjected to Bonferroni correction for seven tests. Shock-expectancy ratings of CSA and CSB (CSY and CSX) were analyzed with a paired *t* test (while this test is rather circular for CSA and CSB, as the only two non-contingency-aware participants were excluded from all analyses, we nevertheless report it, so that the reader can gauge the size of the effect). We used Pearson's *r* to test for a correlation between the strength of conditioning ($CSA > CSB$) and the strength of blocking ($CSY > CSX$) across participants; both fear ratings and shock-expectancy ratings were assessed. While shock-expectancy ratings were given just once after the experiment—and thus are a sensible index of conditioning and blocking—fear ratings were given at seven time points during the experiment and thus had to be collapsed to create sensible conditioning and blocking indices for this correlational analysis. For conditioning, this entailed subtracting the first fear rating (which was given before the start of the experiment, when CSA and CSB had not been presented yet) from the average of the second to the seventh rating for both CSA and CSB and then calculating the difference between these indices for CSA and CSB. For blocking, this entailed subtracting the average of the first to the third rating (which were given before CSX and CSY were presented either as part of the compound stimuli or by themselves) from the average of the fourth to the seventh rating for both CSX and CSY and then calculating the difference between these indices for CSY and CSX. To exclude a confounding influence of demand characteristics/social desirability on all rating results, we also tested for a correlation between participants' ratings and their social desirability scores using Pearson's *r*. Reaction time data were averaged for each condition and then compared between corresponding conditions (CSA–CSB, CSAX–CSBY, CSX–CSY) using paired *t* tests.

After the identification of blinks from the pupillary recordings, eye-tracking position data were parsed using a dispersion-based algorithm for detecting fixations and saccades (Salvucci and Goldberg, 2000). Fixations were coded when the maximum dispersion within an 80 ms moving window did not exceed 50 pixels on the screen. Eye-tracking data from seven participants could not be analyzed due to technical problems. Two types of analysis were performed on the eye-tracking data: one investigated the saccadic latency to each CS; and the other investigated the fixation duration for each CS. The region of interest (ROI) for both analyses was defined by a square of 125×125 pixels around the center of the CS (CS size, 125×125 pixels). For both analyses, the fixation position in the last 100 ms before CS onset was used as baseline. When this fixation fell outside a square of 75×75 pixels around the center of the screen or when saccades of $>1^\circ$ of visual angle were made in the baseline

period, the respective trial was discarded. Similarly, we excluded saccades from the analysis that occurred earlier than 150 ms or later than 1000 ms after CS onset, as these are highly unlikely to be related to initial stimulus detection. With respect to the fixation data, we calculated the cumulative fixation duration on the CS within a time window ranging from 150 to 5000 ms after stimulus onset. Overall, we were able to analyze data from 57.5% of the trials for the saccadic latency analysis and 78.2% of the trials for the fixation duration analysis. Saccadic latency and fixation duration data were averaged for each condition and then compared between corresponding conditions (CSA–CSB, CSAX–CSBY, CSX–CSY), using paired *t* tests. More importantly, we also investigated the fixation duration for the different parts of the compound stimuli, in that we asked whether participants would fixate longer on the predictive part of the compound stimulus (CSA in CSAX and CSY in CSBY). This was first tested separately for CSAX and CSBY using paired *t* tests for each within-compound test (CSA vs CSX in CSAX and CSY vs CSB in CSBY), but in a second step we also formally tested for an interaction, using a 2×2 repeated-measures ANOVA. As these tests are especially affected by trial exclusions (due to the small number of trials when splitting up 10 events into two conditions), we wanted to ensure that the results of our statistical tests were reliable and therefore also analyzed responses across all 20 trials of each condition. Responses were, however, highly similar (i.e., all the reported results were also significant when including all 20 trials; *t* test for compound CSAX, $t_{(24)} = 2.61$, $p = 0.008$; *t* test for compound CSBY, $t_{(24)} = 2.59$, $p = 0.008$; interaction, $F_{(1,24)} = 7.88$, $p = 0.010$).

SCR data were downsampled to 100 Hz and visually inspected for artifacts. To remove possible artifacts related to electrical stimulation, data were interpolated between 4600 and 5000 ms after the onset of each CS. We filtered the data using a 1 Hz low-pass filter and used standard procedures for analysis (Dawson et al., 2007), as follows: (1) on each trial, the maximum response was determined in a latency window of 1–4 s after the start of each CS and was compared against the mean of a 1 s baseline period before each CS; (2) responses below the minimal response criterion of 0.05 μ S were scored as zero; and (3) data were log-transformed and averaged for each condition. SCR to CSA and CSB (CSY and CSX) were then compared using a paired *t* test.

HR data—as obtained from the pulse sensor—from 1 s before CS onset to the end of CS presentation were subjected to a real-time scaling procedure (Velden and Wölk, 1987), which resulted in one HR value per second [in beats per minute (BPM)]. Similar to the SCR analysis, we then compared the maximum in the CS interval to the response during the prestimulus baseline (i.e., 1 s before CS onset) and averaged the responses for each condition. Note that although conditioned HR responses can be both deceleratory and acceleratory, we were specifically interested in the latter as only these have been associated with an increased state of fear (Hodes et al., 1985; Hamm and Vaitl, 1996). Since acceleratory responses have been shown to grow stronger over the course of acquisition (Hare and Blevings, 1975; Fredrikson, 1981), we investigated not only the first 10 trials of acquisition, but also the second 10 trials, where such an effect would be expected to occur. Similar to analyses of SCR, HR changes in response to CSA and CSB (CSY and CSX) were compared, using paired *t* tests. Similar to the rating analyses, we also tested for a correlation between strength of conditioning (CSA > CSB) and blocking (CSY > CSX) in HR data, using Pearson's *r*; this analysis was not conducted for SCR data as they showed no significant blocking effect. In a *post hoc* HR analysis, we also repeated the main comparisons (CSA vs CSB, CSY vs CSX) for deceleratory responses, by comparing the minimum in the CS interval to the response during the prestimulus baseline.

fMRI data analysis. fMRI data processing and statistical analyses were performed using statistical parametric mapping (SPM8; Wellcome Department of Imaging Neuroscience, London, UK). Data processing consisted of slice timing (correction for differences in slice acquisition time), realignment (rigid body motion correction) and unwarping (accounting for susceptibility by movement interactions), coregistration (between EPI images and the skull-stripped T1 image), spatial normalization using the DARTEL toolbox, and smoothing. We created two sets of functional images: one was obtained by resampling the data to $2 \times 2 \times 2$ mm voxels during normalization and smoothing with an 8 mm full-width at half-maximum (FWHM) isotropic Gaussian kernel, whereas the other one

was obtained by resampling the data from a limited field of view (x , $-40:40$ mm; y , $-50:10$ mm; z , $-60:10$ mm; all coordinates in Montreal Neurological Institute (MNI) space) to $1 \times 1 \times 1$ mm voxels and smoothing with a 4 mm FWHM kernel. While the former analysis was used for the examination of cortical regions, the latter analysis was performed to investigate subcortical effects with greater precision (Eippert et al., 2009; Gamer et al., 2010). Data were also subjected to high-pass filtering (cutoff period: 128 s) and correction for temporal autocorrelations (based on a first-order autoregressive model).

Data analysis was performed using a general linear model (GLM) approach. The first-level design matrix of each participant included separate regressors for each of the six conditions (CSA, CSB, CSAX, CSBY, CSX, CSY), which were modeled as delta functions at CS onset and were convolved with the canonical hemodynamic response function. We decided to focus our analysis on the well studied enhancement of CS-related amygdala responses in aversive learning (Quirk et al., 1995; LaBar et al., 1998; Paton et al., 2006) and did not investigate US-related responses, as evidence for the amygdala coding prediction errors in aversive learning is scarce (Belova et al., 2007; Delgado et al., 2008b; Li et al., 2011; McNally et al., 2011). Nevertheless, we wanted to account for the variance induced by US-related responses and therefore created similar regressors at US onset. As mentioned before, only responses to the first 10 CSs were investigated, which is why all conditions were split into two regressors, one coding the first 10 presentations and the other coding the last 10 presentations. Note that the correlation [in terms of cosine (\cos) and Pearson's *r*] between the CS and US regressors did not exceed 0.21 in any case and participant, with the maximal values being $\cos = 0.14$ and $r = 0.09$ for CSA, $\cos = 0.14$ and $r = 0.10$ for CSB, $\cos = 0.21$ and $r = 0.17$ for CSAX, $\cos = 0.19$ and $r = 0.15$ for CSBY, $\cos = 0.20$ and $r = 0.17$ for CSX, and $\cos = 0.19$ and $r = 0.15$ for CSY. All these regressors were also parametrically modulated using an exponentially decaying function to capture the variance induced by responses that develop over trials, but as our data did not exhibit the previously reported decay of amygdala responses (CSA vs CSB) (Büchel et al., 1998), we did not investigate time-dependent effects any further. Rating periods were also modeled, with each button press during rating constituting one event. After model estimation, first-level contrast images from each participant were used for second-level group analyses, which were performed using one-sample *t* tests. Contrasts created at the first level tested for responses to (1) CSA versus CSB, (2) CSY versus CSX, (3) CSAX versus CSA, (4) CSBY versus CSB, and (5) (CSBY vs CSB) versus (CSAX vs CSA). Similar to the behavioral data, we also tested for a relationship between amygdala activation during conditioning (CSA > CSB) and blocking (CSY > CSX) and therefore used the strength of amygdala activation during conditioning from each participant as a covariate in a second-level design investigating blocking. We also wanted to investigate whether there would be regions activating to both CSAX versus CSA and CSBY versus CSB and therefore used a conjunction analysis (testing against the conjunction null; Nichols et al., 2005) at the second level in a full-factorial model.

For each participant, we also created another first-level model that contained the same onset regressors as described previously, but incorporated participants' eye-tracking data as parametric modulators. More specifically, for each of the CSAX and CSBY trials we quantified how much longer participants focused on the predictive part (CSA in CSAX and CSY in CSBY) than the nonpredictive part (CSX in CSAX and CSB in CSBY) of each compound. These differences were then used to parametrically weight the regressors for CSAX and CSBY, and the respective contrasts were raised to and tested at the second-level using one-sample *t* tests.

Finally, we used psychophysiological interaction analyses (PPIs) (Friston et al., 1997) to test for coupling between the prefrontal cortex and the amygdala during the blocking acquisition phase, where the new learning underlying the blocking effect occurs. In these analyses, the first-level design matrix of each participant consisted of the following three regressors: (1) the time course of the seed region; (2) the psychological variable (i.e., a linear combination of onset regressors); and (3) the product of the former two. Only the parameter estimate of the interaction term (i.e., the product of the time course and the psychological variable) was raised to the second level, where a one-sample *t* test was carried out. In each of

these analyses (four in total), the psychological variable was represented by the difference between CSBY and CSAX, and the time course was obtained from one of the four dorsolateral prefrontal cortex (dlPFC) peaks identified in the conjunction GLM analysis. As PPIs test for differential connectivity and results could in principle just be driven by effects in one condition, we also performed “single-condition PPIs,” to quantify the unique contributions of CSAX and CSBY to the observed results.

All first-level analyses also contained participant-specific confound regressors in the design matrix, which were used to model variance of no interest: one regressor modeled the CSF time course, another set of regressors modeled cardiac and respiratory effects (using the method developed by Deckers et al., 2006), and a final set of regressors modeled images contaminated by movement (using an adaptive velocity cutoff with a criterion of either 0.2 mm/TR or 0.4 mm/TR; Gläscher et al., 2010).

In all analyses, we used an initial voxelwise height threshold of $p = 0.005$ and an extent threshold of $k = 10$ voxels. Correction for multiple comparisons using Gaussian random field theory was based on probabilistic maps of our target regions (obtained from the Harvard-Oxford atlas and thresholded at 50%): amygdala, dlPFC (approximated by the middle frontal gyrus mask), and ventromedial PFC (vmPFC)/subgenual anterior cingulate cortex (sgACC) (approximated by the subcallosal cortex mask, which is somewhat overly conservative as it also includes more lateral parts). While the terminology used to describe this latter region is rather inconsistent across studies, note that we were specifically interested in the subgenual anterior cingulate cortex, as it shows a very prominent involvement in various processes of human fear control (Phelps et al., 2004; Delgado et al., 2008a; Klumbers et al., 2010; Nili et al., 2010). No probabilistic maps exist for dACC and AI, which we therefore approximated by spheres (10 mm radius) centered on coordinates from a recent meta-analysis of fear conditioning (Mechias et al., 2010). For the sake of brevity, we only report the four most significant peaks in each region.

Cortical responses were additionally investigated with a threshold-free ROI-based approach, in which the values from all voxels in the respective ROIs were averaged per participant and condition (using *rfxplot*: Gläscher, 2009) and then compared between conditions at the group level using a threshold of $p = 0.05$ (one tailed in all cases except for the sgACC, which sometimes exhibits decreases below the baseline); note that this kind of test does not require correction for multiple comparisons across voxels as only the average of all voxels in the ROI is subjected to statistical testing. This approach gives a summary measure of a region's involvement and can be seen as complementary to the above-described peak-based analysis, as it is unbiased (positive as well as negative voxels contribute) and more sensitive to spatially extended responses; additionally, it allows plotting of parameter estimates without being sensitive to issues of circularity.

Results

Control of confounds

False or missed responses in the reaction time task at the beginning of each trial occurred on only 1.7% of the trials, suggesting good task compliance. There were no significant differences in reaction times between the corresponding conditions (i.e., CSA-CSB, CSAX-CSBY, CSX-CSY), which indicates similar levels of vigilance; the same relationship held for saccadic latencies to the target as well as target fixation durations (Table 1). With regard to the rating effects reported below, we found no significantly positive correlation of any effect with social desirability scores (all $p > 0.356$), making it unlikely that these results were driven by possible demand characteristics.

Acquisition of conditioned fear

We first tested whether the acquisition of conditioned fear was successful on the behavioral level, as this is a prerequisite for the blocking effect to occur. A 2×7 repeated-measures ANOVA of the fear ratings to CSA (reinforced stimulus) and CSB (nonreinforced stimulus) revealed a significant main effect of condition

Table 1. Control of confounds

	RT	SL	FD
CSA	626 ± 21	309 ± 17	3065 ± 257
CSB	630 ± 20	299 ± 15	3045 ± 231
CSA vs CSB			
<i>t</i>	−0.43	0.73	0.30
<i>p</i>	0.673	0.473	0.768
CSAX	660 ± 22	340 ± 20	2677 ± 247
CSBY	646 ± 22	322 ± 18	2667 ± 245
CSAX vs CSBY			
<i>t</i>	1.68	1.31	0.15
<i>p</i>	0.103	0.204	0.881
CSX	618 ± 21	280 ± 11	2924 ± 250
CSY	611 ± 19	287 ± 13	2853 ± 262
CSX vs CSY			
<i>t</i>	0.83	−0.70	0.86
<i>p</i>	0.413	0.496	0.400

Shown are the mean ± SEM for the three pairwise condition comparisons, as well as the corresponding *t* values ($df = 31$ for RT; $df = 24$ for SL and FD) and *p* values. RT, Reaction time (ms); SL, saccadic latency (ms); FD, fixation duration (ms). As these analyses served to investigate participants' compliance and vigilance throughout the experiment, all trials were included.

($F_{(1,31)} = 232.94$, $p < 0.001$), a significant main effect of time ($F_{(3,0,93.2)} = 38.86$, $p < 0.001$), as well as a significant time-by-condition interaction ($F_{(3,3,100.8)} = 32.71$, $p < 0.001$). *Post hoc* tests demonstrated no significant differences between CSA and CSB before the start of the paradigm ($p = 0.074$ uncorrected), but significantly higher ratings of fear for CSA than CSB at each time point after the start of the paradigm (all $p < 0.001$ uncorrected; all effects survive Bonferroni correction for seven time points; Fig. 2a). The shock-expectancy ratings obtained after the experiment were significantly higher for CSA than CSB ($t_{(31)} = 35.56$, $p < 0.001$; Fig. 2b), as were SCR ($t_{(25)} = 3.39$, $p = 0.001$; Fig. 2c); acceleratory HR changes were somewhat more complex and only showed such a pattern in the last 10 acquisition trials (first 10 trials: $t_{(27)} = -0.53$, $p = 0.700$; last 10 trials: $t_{(27)} = 1.55$, $p = 0.067$, trend-level; Fig. 2d). In a *post hoc* analysis, we also investigated deceleratory HR changes but did not find any significant differences between CSA and CSB, neither in the first 10 trials (CSA, -2.51 BPM; CSB, -2.40 BPM; $t_{(27)} = -0.28$, $p = 0.78$, two tailed) nor in the last 10 trials (CSA, -2.31 BPM; CSB, -2.42 BPM; $t_{(27)} = 0.29$, $p = 0.78$, two tailed).

Moving on to the neural level, the threshold-free ROI-based analysis of cortical regions (which does not require correction for tests across multiple voxels; see Materials and Methods) showed significantly stronger BOLD responses to CSA than CSB in all three candidate regions [dACC: $t_{(31)} = 2.84$, $p = 0.004$; right AI (rAI): $t_{(31)} = 3.05$, $p = 0.003$; left AI (lAI): $t_{(31)} = 2.88$, $p = 0.004$; Fig. 2e]; the same pattern of results was observed in the peak-based analysis (Table 2). Most importantly, BOLD responses in the amygdala were also significantly stronger for CSA than CSB (MNI coordinates: 29, -7 , -17 ; $t_{(31)} = 4.51$; $p < 0.001$ uncorrected; $p = 0.018$ corrected; Fig. 2f). Together, these results demonstrate that the acquisition of conditioned fear was successful, which allowed us to test for possible blocking effects on the behavioral and neural level.

Expression of blocking

A 2×7 repeated-measures ANOVA of the fear ratings to CSY (non-blocked stimulus) and CSX (blocked stimulus) revealed a significant main effect of condition ($F_{(1,31)} = 13.67$, $p = 0.001$), a significant main effect of time ($F_{(2,6,79.8)} = 40.39$, $p < 0.001$), as well as a significant time-by-condition interaction ($F_{(3,5,108.8)} = 5.99$, $p < 0.001$). *Post hoc* tests demonstrated no significant differences between CSY

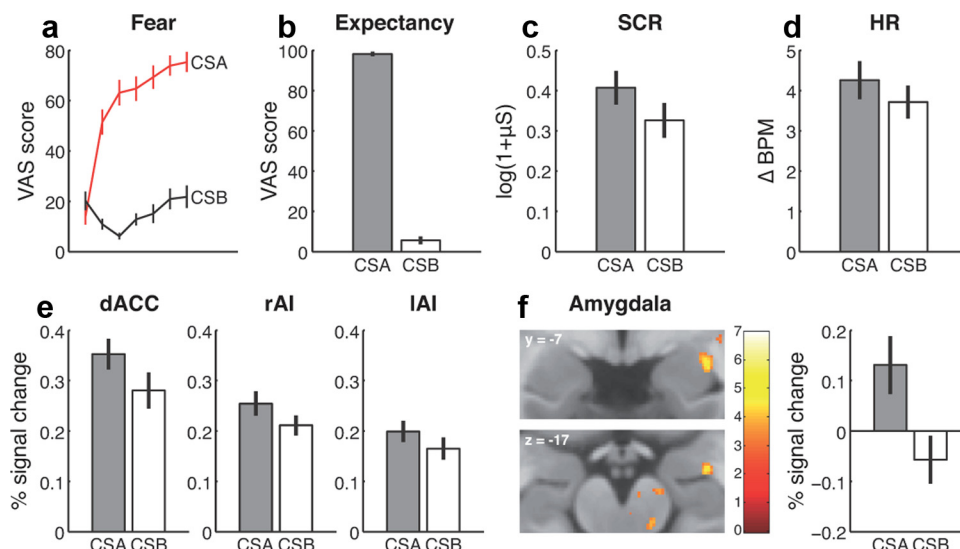


Figure 2. Behavioral and neural responses during the acquisition of conditioned fear (CSA vs CSB). **a–d**, Significant behavioral effects were observed in fear ratings (**a**), which were obtained at seven time points during the experiment, shock-expectancy ratings (**b**), which were obtained after the experiment, SCRs (**c**), which were obtained on each trial and averaged, and HR responses (**d**), which were obtained on each trial and averaged (trend-level). **e, f**, Significantly enhanced responses were also observed on the neural level, as evidenced in dACC, rAI, and IAI (**e**), as well as amygdala (**f**). Coronal and transverse brain sections depict the location of the amygdala response, and descriptive parameter estimates show the amygdala response separately for each condition. The activation map is thresholded at the initial uncorrected level of $p < 0.005$ for visualization purposes and displayed on the average structural image of all participants; the color bar denotes t values. All error bars reflect the SEM.

Table 2. Peak-based analysis results for dACC and AI

Region	x	y	z	t	p
CSA > CSB					
dACC	−4	10	42	4.39	0.007
	4	10	34	4.29	0.009
	−4	10	32	4.29	0.009
	−4	8	38	4.19	0.011
rAI	38	26	−4	3.99	0.017
IAI	−28	26	0	4.30	0.009
	−32	26	−2	4.22	0.011
CSY > CSX					

CSAX > CSA					

CSBY > CSB					
dACC	6	16	44	5.07	0.001
	−6	8	36	4.46	0.005
	8	12	34	4.44	0.006
	−4	10	42	4.37	0.007
rAI	36	24	−6	4.73	0.003
IAI	−30	24	−2	4.76	0.003
Interaction					
dACC	−6	12	30	3.98	0.016
	−4	12	44	3.75	0.027
rAI	32	22	−8	4.09	0.013
	38	26	−4	3.99	0.016
IAI	−28	24	−2	3.63	0.034

Contrasts are listed according to the appearance in the Results section; the interaction contrast refers to [(CSBY > CSB) > (CSAX > CSA)]. Coordinates are denoted by x, y, z (in mm; MNI space). The strength of activation is expressed in t scores ($df = 31$), and corrected p values are reported. Peaks are listed according to increasing p values for each region. ---, No significant responses.

and CSX before the start of the second phase (all $p > 0.201$ uncorrected), but significantly higher ratings for CSY than CSX at each time point thereafter (all $p < 0.002$ uncorrected, all effects survive Bonferroni correction for seven time points; Fig. 3a). The shock-expectancy ratings obtained after the experiment were significantly higher for CSY than CSX ($t_{(31)} = 1.80, p = 0.041$; Fig. 3b). Acceleratory HR changes were also significantly stronger for CSY than CSX

($t_{(27)} = 2.12, p = 0.022$; Fig. 3c), but SCR did not show a significant difference ($t_{(25)} = 0.65, p = 0.260$); closer inspection of the SCR data indicated that this might be due to the low level of responding present at this late time point of the experiment (participants only showed responses to CSX or CSY on 34% of the trials). A *post hoc* analysis of deceleratory HR changes did not show a significant difference (CSX, -2.88 BPM; CSY, -2.55 BPM; $t_{(27)} = 0.97; p = 0.34$, two-tailed).

On the neural level, we tested for significantly stronger responses to CSY than CSX in the three cortical structures of interest (dACC, rAI, and IAI), as well as the amygdala. The ROI-based analysis indicated no significant difference between CSY and CSX in all three cortical regions (dACC: $t_{(31)} = 0.45, p = 0.328$; rAI: $t_{(31)} = 0.40, p = 0.347$; IAI: $t_{(31)} = 0.82, p = 0.210$), which was mirrored by no significant results in the peak-based analysis. In contrast to the cortical regions, the amygdala showed significantly stronger BOLD responses to the nonblocked stimulus CSY than to the blocked stimulus CSX (MNI coordinates: $-29, -3, -22$; $t_{(31)} = 4.11; p < 0.001$ uncorrected; $p = 0.033$ corrected; Fig. 3d). In a *post hoc* analysis, we also tested for an overlap of conditioning effects (difference of CSA vs CSB) and blocking effects (difference of CSY vs CSX) by using a conjunction analysis (Nichols et al., 2005) in a full-factorial model. We did not observe a significant overlap of these two effects at a corrected level, but when lowering the threshold to $p < 0.05$ uncorrected, we observed some overlap between conditioning-related and blocking-related activations in the amygdala (MNI coordinates: $-19, -2, -14$; $t_{(124)} = 2.06; p = 0.021$ uncorrected; MNI coordinates: $-28, -7, -18$; $t_{(124)} = 1.68; p = 0.047$ uncorrected).

Having demonstrated blocking effects on the behavioral (fear ratings, shock-expectancy ratings, acceleratory HR changes) and neural level (amygdala responses), we next asked whether the strength of conditioning (difference of CSA vs CSB) would be positively related to the strength of blocking (difference of CSY vs CSX) across participants; note that using differential effects should abolish any unspecific effects that might vary across participants (e.g., tonic differences in heart rate) and could thus bias this analysis. While no significantly positive relationship could be

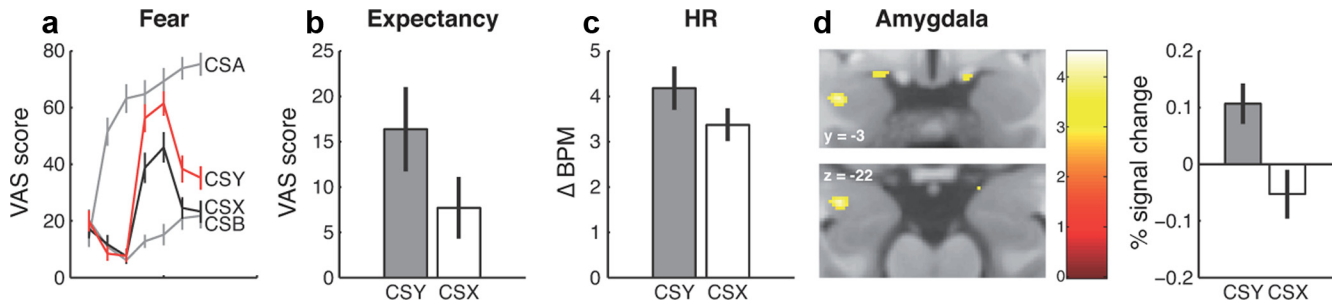


Figure 3. Behavioral and neural responses during blocking of conditioned fear (CSY vs CSX). *a–c*, Significant behavioral blocking effects were observed in fear ratings (for comparison, responses to CSA and CSB are depicted in gray) (*a*), shock-expectancy ratings (*b*), and HR responses (*c*). *d*, A significant blocking effect was also observed in the amygdala. Coronal and transverse brain sections depict the location of the amygdala response, and descriptive parameter estimates show the amygdala response separately for each condition. The activation map is thresholded at the initial uncorrected level of $p < 0.005$ for visualization purposes and displayed on the average structural image of all participants; the color bar denotes t values. All error bars reflect the SEM.

observed for shock-expectancy ratings ($r = 0.10$, $p = 0.293$), we observed a positive relationship for both fear ratings ($r = 0.31$, $p = 0.042$) and HR changes ($r = 0.51$, $p = 0.003$). Furthermore, the strength of amygdala BOLD responses during conditioning predicted the strength of both right (MNI coordinates: 23, -12, -12; $t_{(30)} = 5.27$; $p < 0.001$ uncorrected; $p = 0.003$ corrected; MNI coordinates: 19, -6, -12; $t_{(30)} = 4.31$; $p < 0.001$ uncorrected; $p = 0.028$ corrected) and left amygdala BOLD responses during blocking (MNI coordinates: -22, -4, -14; $t_{(30)} = 4.66$; $p < 0.001$ uncorrected; $p = 0.010$ corrected; MNI coordinates: -19, -8, -19; $t_{(30)} = 4.24$; $p < 0.001$ uncorrected; $p = 0.027$ corrected), i.e., the greater the difference between CSA and CSB, the greater the difference between CSY and CSX.

Acquisition of blocking

The design of our study also allowed us to investigate the neurobiological mechanisms underlying the acquisition of the blocking effect (i.e., responses to the compound stimuli CSAX and CSBY). We expected that responses in our structures of interest (dACC, rAI, IAI, and amygdala) would not increase from CSA to CSAX, as CSX has no additional predictive value. Conversely, we expected an increase in responses from CSB, which predicted the absence of shock, to CSBY, which predicts shock.

Results from the ROI-based analyses in dACC, rAI, and IAI confirmed these predictions: when we tested for stronger responses to CSAX than CSA, we did not observe any significant increases (dACC: $t_{(31)} = 1.07$, $p = 0.147$; rAI: $t_{(31)} = 0.45$, $p = 0.329$; IAI: $t_{(31)} = 0.07$, $p = 0.473$; Fig. 4*a,b*). Conversely, when testing for stronger responses to CSBY than CSB, we observed significant response increases in all three regions (dACC: $t_{(31)} = 3.42$, $p = 0.001$; rAI: $t_{(31)} = 3.18$, $p = 0.002$; IAI: $t_{(31)} = 2.34$, $p = 0.013$; Fig. 4*a,b*). To formally test for a significantly stronger increase from CSB to CSBY than from CSA to CSAX, we also performed an interaction analysis, which revealed a significant interaction in all three regions (dACC: $F_{(1,31)} = 8.95$, $p = 0.005$; rAI: $F_{(1,31)} = 10.56$, $p = 0.003$; IAI: $F_{(1,31)} = 7.87$, $p = 0.009$). Results from the peak-based analyses showed a similar pattern, with significant BOLD responses (or their absence) in all three regions in the respective contrasts (Table 2; Fig.

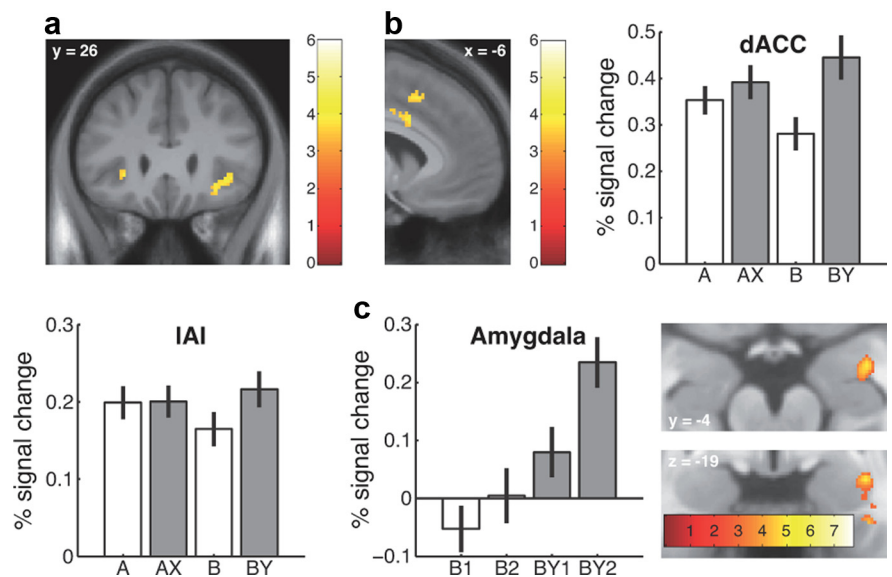


Figure 4. Neural responses during the development of the blocking effect. *a, b*, A significant interaction (i.e., a stronger increase from CSB to CSBY than from CSA to CSAX) was observed in IAI and rAI (*a*) and dACC (*b*). The coronal and sagittal brain sections reflect the location of the responses in the peak-based analysis, whereas the parameter estimates were obtained from the ROI-based analysis (parameter estimates for rAI are not shown). The cortical activation maps are thresholded at $p < 0.001$, uncorrected, for visualization purposes and are displayed on the average structural image of all participants. *c*, Amygdala responses (coronal and transverse section) stem from the contrast CSBY-CSB over all 20 stimuli. As shown in the descriptive plot of the parameter estimates, amygdala responses to CSBY developed slowly (B1, first 10 trials of CSB; B2, second 10 trials of CSB; BY1, first 10 trials of CSBY; BY2, second 10 trials of CSBY). The amygdala activation map is thresholded at the initial uncorrected level of $p < 0.005$ for visualization purposes and displayed on the average structural image of all participants. All color bars denote t values and error bars reflect the SEM.

4*a,b*). Similar to the cortical regions, amygdala responses did not show any significant increase from CSA to CSAX, but unexpectedly failed to show a significant increase from CSB to CSBY. A closer inspection of the data revealed that this was due to amygdala responses being rather slow in tracking the change in reinforcement: when investigating responses across all 20 stimuli (instead of only the first 10), there was a highly significant increase from CSB to CSBY (MNI coordinates: 31, -4, -19; $t_{(31)} = 4.97$; $p < 0.001$ uncorrected; $p = 0.005$ corrected; Fig. 4*c*) and a descriptive plot of the parameter estimates shows that the difference is already existent during the first 10 stimuli, but grows much stronger during the second 10 stimuli (Fig. 4*c*). Most importantly however, and mirroring the cortical regions, the interaction was significant in the amygdala during the first 10 stimuli (MNI coordinates: 30, -5, -18; $t_{(31)} = 4.72$; $p < 0.001$ uncorrected; $p = 0.011$ corrected).

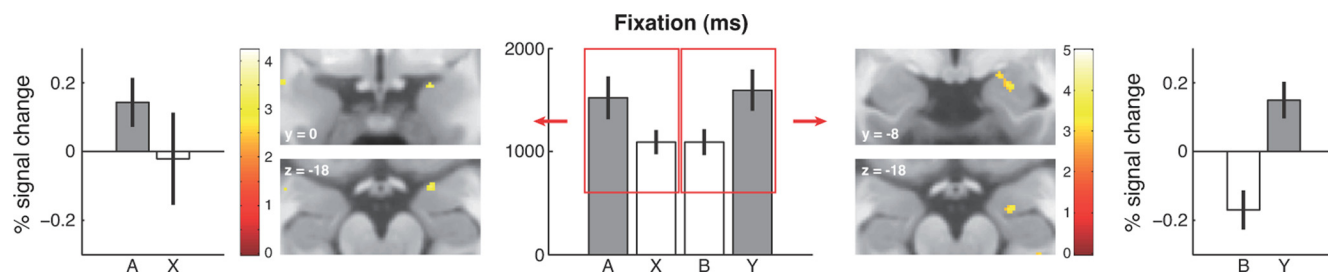


Figure 5. Attentional effects during the development of the blocking effect. Eye-tracking data (middle bar graph) acquired during the compound trials (CSAX, CSBY) showed that participants fixated significantly longer on the predictive (CSA in CSAX, CSY in CSBY) than on the nonpredictive part (CSX in CSAX, CSB in CSBY) of the compound stimuli. When using these trial-wise differences in fixation duration to inform the imaging analysis (as represented by the red rectangles and arrows), significantly stronger amygdala responses were observed on those trials where participants fixated longer on CSA than CSX (left), and on CSY than CSB (right). Descriptive parameter estimates were also obtained for each of these trial types (by using *rfxplot*'s regressor splitting capability). Activation maps are thresholded at the initial uncorrected level of $p < 0.005$ for visualization purposes and are displayed on the average structural image of all participants; the color bars denote t values. All error bars reflect the SEM.

While the above-described pattern of responses to the compound stimuli supports our hypotheses, we were also interested in isolating the responses associated with the individual parts of the compounds (e.g., CSA and CSX in compound CSAX), as they have different predictive values. Furthermore, our data had shown that responses to CSY (nonblocked cue) were significantly more pronounced than to CSX (blocked cue) and this difference must develop during the blocking acquisition phase. In a first step, we therefore analyzed the eye-tracking data and observed that participants fixated significantly longer on the predictive than the nonpredictive part of the compound stimuli (i.e., CSA in CSAX: $t_{(24)} = 1.91$; $p = 0.034$; and CSY in CSBY: $t_{(24)} = 2.20$; $p = 0.019$); this effect also held when formally testing for an interaction ($F_{(1,24)} = 4.92$; $p = 0.036$; Fig. 5). Next, we asked whether this higher level of overt attention to the predictive stimulus would also be related to the strength of amygdala activation (i.e., whether we would observe stronger amygdala BOLD responses in trials where participants fixated more on the predictive part than on the nonpredictive part of the compound stimulus). We therefore parametrically modulated each of the compound onset regressors (CSAX and CSBY) with the trial-wise difference in fixation duration between the parts of the compound stimuli. As expected, the amygdala showed significantly stronger responses when participants fixated more on CSA than on CSX (MNI coordinates: 19, 0, -18; $t_{(22)} = 3.34$; $p = 0.001$ uncorrected; Fig. 5) and more on CSY than CSB (MNI coordinates: 21, -8, -18; $t_{(22)} = 3.64$; $p = 0.001$ uncorrected; MNI coordinates: 18, -6, -14; $t_{(22)} = 3.35$; $p = 0.001$ uncorrected; Fig. 5), although these effects did not survive correction for multiple comparisons. A descriptive plot of the parameter estimates associated with each type of event showed that there was virtually no amygdala response on trials during which participants fixated more on CSX than CSA, and that there was a negative amygdala response on trials during which participants fixated more on CSB than CSY (Fig. 5); BOLD responses were similarly strong on trials during which participants fixated more on CSA (than CSX) and CSY (than CSB) in both compounds.

Finally, we investigated the involvement of prefrontal regions (dlPFC and vmPFC) in the acquisition of blocking. When comparing BOLD responses between CSAX and CSA in the ROI-based analysis, we observed significantly stronger right dlPFC (rdlPFC) responses to CSAX [rdlPFC: $t_{(31)} = 3.08$, $p = 0.002$; left dlPFC (ldlPFC): $t_{(31)} = 0.43$, $p = 0.335$; Fig. 6a]. Interestingly, the vmPFC also showed a significant difference between CSAX and CSA, but in the opposite direction, with parameter estimates indicating a clearly negative response under CSAX [right vmPFC

(rvmPFC): $t_{(31)} = -2.17$, $p = 0.038$; left vmPFC (lvmPFC): $t_{(31)} = -1.97$, $p = 0.058$, trend-level; Fig. 6b). The comparison of CSBY and CSB showed significantly stronger responses to CSBY in the dlPFC (rdlPFC: $t_{(31)} = 3.30$, $p = 0.001$; ldlPFC: $t_{(31)} = 2.72$, $p = 0.006$; Fig. 6a), but virtually no changes in vmPFC (rvmPFC: $t_{(31)} = -0.54$, $p = 0.591$; lvmPFC: $t_{(31)} = -0.75$, $p = 0.461$; Fig. 6b). The results of the peak-based analyses mirrored these effects (Table 3). Next, we asked whether there would be responses that are common to the change from CSA to CSAX and CSB to CSBY, due to the new learning that occurs in both conditions. Using a conjunction analysis [(CSAX > CSA) and (CSBY > CSB)], we observed significant responses in the right dlPFC, but not the left dlPFC or the vmPFC (Table 3; Fig. 7).

We then investigated whether the right dlPFC, despite showing significant responses in both comparisons (i.e., CSAX > CSA and CSBY > CSB), would exhibit differential connectivity to the amygdala in the blocking acquisition phase. We expected a significantly more positive coupling between dlPFC and amygdala on CSBY than CSAX trials, due to the different type of learning that should occur: while CSY (as part of CSBY) has to be established as predictive for the US and should thus engage the amygdala, CSX (as part of CSAX) has to be established as nonpredictive for the US. Using effective connectivity (PPI) analyses—based on the four rdlPFC peaks from the conjunction analysis—we observed that the rdlPFC indeed showed a significantly more positive coupling with the amygdala under CSBY than CSAX (Table 4), with the most posterior dlPFC peak showing the strongest coupling; at an uncorrected level of significance, this effect was observed for all four rdlPFC peaks. To investigate the specificity of this finding, we also tested the reverse contrast (i.e., stronger coupling under CSAX than CSBY), but did not find any effect in the amygdala, not even at an uncorrected threshold. We also wanted to confer a certain degree of functional relevance on these findings, and therefore asked whether the influence of rdlPFC would occur in the amygdala region observed during the acquisition of conditioned fear (right dorsal lateral amygdala) or in the amygdala region observed during the blocking of conditioned fear (left ventral lateral amygdala; these search regions were created by thresholding the respective maps (CSA vs CSB, CSY vs CSX) at the initial threshold of $p = 0.005$ and $k = 10$, multiplying the resulting images with the amygdala masks and binarizing the result). We only observed significant results in the blocking-related region (MNI coordinates: -28, -5, -23; $t_{(31)} = 3.30$; $p = 0.001$ uncorrected; $p = 0.020$ corrected), but not in the acquisition-related region. We then tested for the individual contribution of each condition to this result and observed a significantly positive coupling in the blocking-related region under CSBY (MNI coordinates: -29,

−5, −24; $t_{(31)} = 3.68$; $p < 0.001$ uncorrected; $p = 0.008$ corrected; Fig. 7) and a significantly negative coupling under CSAX (MNI coordinates: −31, −2, −21; $t_{(31)} = 2.97$; $p = 0.003$ uncorrected; $p = 0.043$ corrected; Fig. 7).

Discussion

In this study, we demonstrated that behavioral and amygdala responses show significant blocking effects in the context of aversive learning. We furthermore observed that attentional factors are involved in the development of this effect and that prefrontal regions play a significant role by flexibly changing their coupling to the amygdala, according to the type of learning that occurs.

On the behavioral level, we observed significant blocking effects in fear ratings and shock-expectancy ratings. With respect to autonomic responses, acceleratory HR changes reflected blocking, but no such effect was observed for deceleratory HR changes and SCR. The absence of a blocking effect in SCR was unexpected and contrasts with previous behavioral studies (Hinchey et al., 1995; Mitchell and Lovibond, 2002; but see Davey and Singh, 1988; Lovibond et al., 1988). A potential explanation for the difference in results between our study and the ones by Hinchey et al. (1995) and Mitchell and Lovibond (2002) might be the longer duration of our paradigm, as participants showed a very low level of electrodermal response during this late phase of the experiment, possibly obscuring a blocking effect in SCR. Another autonomic measure of learning, HR acceleration, did, however, show significantly stronger responses to the non-blocked stimulus (CSY) than to the blocked stimulus (CSX). This is especially relevant, because conditioned HR acceleration is a sensitive index of the affective component in aversive learning (Öhman, 2009), as it has been selectively associated with an increased state of fear, as represented for example by emotional ratings (Hodes et al., 1985; Hamm and Vaitl, 1996) and fear-potentiated startle (Hamm and Vaitl, 1996), and can also occur without awareness of the CS–US relationship (Hamm and Vaitl, 1996). A *post hoc* analysis of deceleratory HR responses did not show a blocking effect, but also failed to show a basic effect of conditioning. It is also important to note that we observed a blocking effect in shock-expectancy ratings, as these were given after the experiment, when substantial extinction had already occurred to CSX and CSY (Martin and Levey, 1991). While this speaks for the strength of the blocking effect, fear ratings indicated that blocking was not absolute, which is in line with the original finding that blocking is not necessarily an all-or-none effect (Kamin, 1968). Importantly, at the end of the experiment, fear ratings to CSX and CSB settled at the same level, significantly below those to CSY. Together, these data demonstrate that blocking effects in humans can be observed in multiple response systems, ranging from the representation of declarative knowledge of the contingency (expectancy ratings) over the subjective affective impact of conditioned stimuli (fear ratings) to autonomic nervous system responses (acceleratory HR changes).

Mirroring the behavioral results, we observed significantly lower amygdala responses to blocked stimuli (CSX) than nonblocked stimuli (CSY). This highlights that temporal contiguity is clearly not sufficient in explaining amygdala responses during aversive learning,

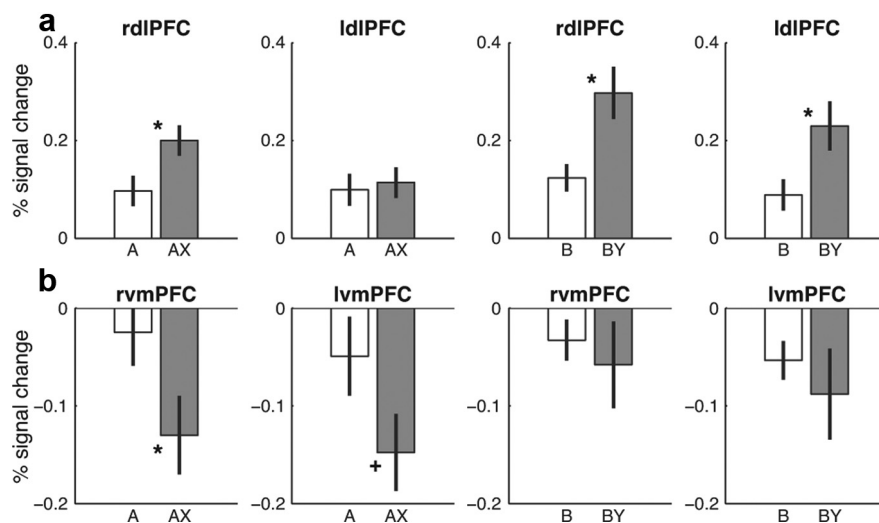


Figure 6. Prefrontal responses during the development of the blocking effect. *a, b*, Responses from the ROI-based analysis in rdIPFC and ldIPFC (*a*), and in rvmPFC and lvmPFC (*b*). Responses were compared between CSAX and CSA trials, and between CSBY and CSB trials. While dlPFC responses were significantly stronger for CSAX–CSA and CSBY–CSB, vmPFC responses were significantly stronger only for CSAX–CSA; importantly, the responses in vmPFC were deactivations. Symbols denote significance; *significant at $p = 0.05$; + $p = 0.058$, trend level. Error bars reflect the SEM.

Table 3. Peak-based analysis results for dlPFC and vmPFC

Region	x	y	z	t	p
CSAX > CSA					
rdIPFC	42	14	34	5.26	0.001
	50	16	36	4.58	0.006
	50	32	30	4.47	0.008
	36	18	50	4.40	0.010
CSAX < CSA					
vmPFC ^a	6	24	−12	3.08	0.095
CSBY > CSB					
ldIPFC	−44	8	46	4.23	0.014
	−42	12	48	3.98	0.026
	−48	30	24	3.88	0.032
	−46	20	34	3.86	0.034
rdIPFC	50	30	26	6.40	0.000
	42	16	32	5.78	0.000
	42	4	52	5.20	0.001
	52	20	34	4.92	0.002
Conjunction					
rdIPFC	50	30	26	4.10	0.005
	40	14	32	3.94	0.008
	46	16	34	3.56	0.026
	52	20	34	3.55	0.026

Contrasts are listed according to the appearance in the Results section; the conjunction contrast refers to [(CSAX > CSA) and (CSBY > CSB)] and tests against the conjunction-null (Nichols et al., 2005). Coordinates are denoted by x, y, z (in mm; MNI space). The strength of activation is expressed in t scores ($df = 31$), and corrected p values are reported. Peaks are listed according to increasing p values for each region. rdIPFC, Right dorsolateral prefrontal cortex; ldIPFC, left dorsolateral prefrontal cortex.

^aNote that the vmPFC response does not survive correction for multiple comparisons.

but that a notion of predictive value or informational content has to be invoked (Kamin, 1968; see also Egger and Miller, 1962). It also ties in with recent animal studies demonstrating amygdalar dopamine transmission being involved in aversive blocking (Iordanova, 2010) as well as contingency-related but not contiguity-related amygdala responses being involved in appetitive learning (Bermudez and Schultz, 2010). We also note that the diminished responses to the blocked cue occurred in the lateral part of the amygdala, at a location that might correspond to the lateral nucleus, which is the main site of plasticity in fear conditioning (Quirk et al., 1995; Repa et al., 2001; for review, see Maren and Quirk, 2004). While

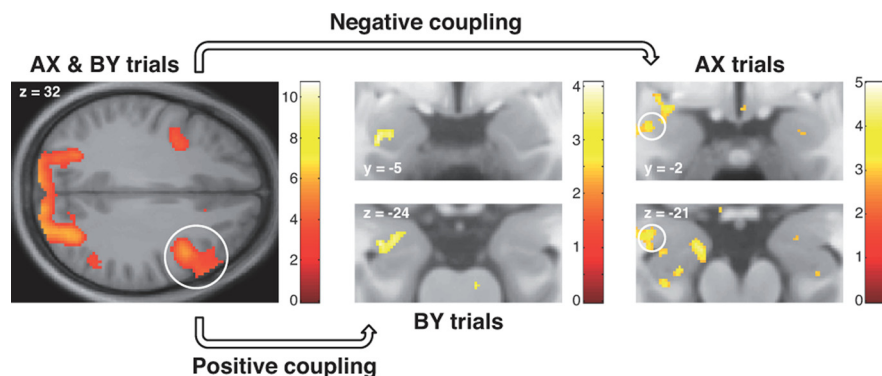


Figure 7. dIPFC–amygdala connectivity during the development of the blocking effect. The PPI analysis originated from a peak in dIPFC (inside white circle), obtained from the conjunction analysis ((CSAX > CSA) and (CSBY > CSB)); left transverse section). dIPFC–amygdala coupling is significantly positive on BY trials (middle coronal and transverse sections) and significantly negative on AX trials (right coronal and transverse sections). Importantly, the amygdala peaks obtained from this connectivity analysis fall into the amygdala region where blocking effects are observed (represented by the white circles). Activation maps are thresholded at the initial uncorrected level of $p < 0.005$ for visualization purposes and displayed on the average structural image of all participants; the color bars denote t values.

Table 4. PPI analysis results

	x	y	z	t	p_{uncor}	p_{cor}
dIPFC peak 1 (50, 30, 26)	–26	–2	–20	3.67	0.000	0.084
	–25	–7	–23	3.42	0.001	0.138
	29	1	–23	3.91	0.000	0.063
	22	–12	–12	3.07	0.002	0.300
dIPFC peak 2 (40, 14, 32)	–26	–6	–23	4.21	0.000	0.027
	–15	–7	–21	4.13	0.000	0.031
	–28	–4	–27	3.43	0.001	0.134
	–24	–5	–18	3.01	0.003	0.279
	16	–4	–21	4.01	0.000	0.050
	22	–6	–18	3.30	0.001	0.203
dIPFC peak 3 (46, 16, 34)	27	–2	–28	3.14	0.002	0.269
	–15	–7	–20	3.41	0.001	0.139
	–27	–10	–17	3.21	0.002	0.198
	–23	–6	–17	3.08	0.002	0.248
	–15	–2	–19	2.75	0.005	0.404
dIPFC peak 4 (52, 20, 34)	15	–5	–21	2.96	0.003	0.349
	–26	–11	–17	3.49	0.001	0.117

Significant peaks within the amygdala, as resulting from the PPI analyses that used the conjunction dIPFC (dorsolateral prefrontal cortex, MNI coordinates listed in brackets) peaks as seed regions. Coordinates are denoted by x, y, z (in mm; MNI space). The strength of activation is expressed in t scores ($df = 31$), and both uncorrected (p_{uncor}) and corrected p values (p_{cor}) are reported; peaks that survive correction are in bold typeface. Peaks are listed according to increasing p values, with left amygdala peaks listed before right amygdala peaks.

the amygdala might thus provide a good starting point in investigating aversive blocking, it would also be interesting to investigate the interplay between the amygdala and the periaqueductal gray as well as ventral striatum, since opioidergic neurotransmission in these structures contributes to blocking in aversive learning (Iordanova et al., 2006; McNally and Cole, 2006). A further observation we made was that the strength of the blocking effect in both behavioral (fear ratings and HR acceleration) and amygdala responses correlated with the amount of previous learning, replicating one of the original behavioral findings of Kamin (1968) and again underscoring the role of predictive value in learning. Contrary to our hypotheses, the blocking effect was only expressed in the amygdala, but not in AI or dACC, which is surprising given that responses of these regions to the compound stimuli were perfectly in line with the development of blocking. This would suggest that previous observations of AI and dACC responses during fear conditioning (Sehlmeyer et al., 2009) only reflect CS–US contiguity processing. However, there is also an alternative explanation,

in line with a previous report on differences in temporal integration between the amygdala and cortical regions (Gläscher and Büchel, 2005): while the amygdala was slow in increasing its responses from CSB to CSBY, but showed extended responses to CSY versus CSX, cortical responses were faster in changing their allegiance, and might also have extinguished more quickly and could thus have been missed in our analysis of CSY versus CSX.

When investigating the development of the blocking effect, we observed that participants fixated significantly longer on the predictive than on the nonpredictive part of each compound, suggesting that the predictive history of cues influences attentional processes. This finding is consistent with previous reports on enhanced levels of overt attention to predictive cues, as observed in blocking procedures in the field of causal learning (Kruschke et al., 2005; Wills et al., 2007; Beesley and Le Pelley, 2011; for a more general perspective, see Hogarth et al., 2010), and extends this pattern of learning-induced gazing behavior to the domain of aversive learning. Although participants fixated significantly longer on the predictive parts of the compound, they still showed fixations to the nonpredictive parts, demonstrating that blocking is not due to a failure of stimulus detection (Kamin, 1968; Waelti et al., 2001). The observed attentional bias toward the predictive part of each compound was paralleled by enhanced amygdala responses, which is not only interesting with respect to the amygdala's role in attentional processes (Holland and Gallagher, 1999; Holland et al., 2000), but even more so with regard to its involvement in directing gaze to relevant features of the visual field, as demonstrated by both lesion (Adolphs et al., 2005) and imaging studies (Gamer and Büchel, 2009). These latter studies focused on facial features, but, considering the amygdala's more general role in detecting and responding to relevant events (Anderson and Phelps, 2001; Sander et al., 2003; Klinge et al., 2010), it is likely that such a process also occurs for more abstract stimuli that have been endowed with relevance via associative learning processes.

Alternatively, the observed amygdala responses might be secondary to gaze changes and could reflect an approximation to the predictive value of the fixated-upon individual cues (Gläscher and Büchel, 2005). A further observation during the development of the blocking effect was the involvement of the dIPFC, which showed significant activations in both investigated comparisons (CSAX vs CSA and CSBY vs CSB). It is tempting to interpret these responses as relevant for the new learning that has to occur in this phase: while CSX (as part of CSAX) has to be established as nonpredictive for the US, CSY (as part of CSBY) has to be established as predictive for the US. Previous studies would support such an interpretation, linking dIPFC responses to accurate contingency tracking (Carter et al., 2006) and to unsigned prediction error signals as observed in causal learning studies (Fletcher et al., 2001; Turner et al., 2004). Yet, confounding factors such as stimulus novelty and stimulus complexity were also present on compound trials and could thus account for the increase in dIPFC activity (Turner et al., 2004). To dissociate learning-related from confound-related dIPFC signals, we investigated whether the dIPFC would show differential connectivity to the amygdala, depending on the requirements of the predictive relation-

ships: we expected a significantly stronger dlPFC–amygdala coupling on BY than on AX trials, as on the former trials excitatory learning takes place, whereas on the latter trials inhibitory learning could take place (as CSX has to be established as redundant). dlPFC–amygdala coupling—originating from a dlPFC region showing positive responses on both AX and BY trials—indeed flexibly switched between positive and negative. Most importantly, the observation that dlPFC–amygdala coupling was significantly negative on CSAX trials suggests that blocking is not a passive process, but that establishing CSX as redundant involves an active inhibitory process, possibly mediated by prefrontal influences on the amygdala. As acquisition-related and blocking-related amygdala responses occurred in different parts of the amygdala (acquisition, right dorsal lateral amygdala; blocking, left ventral lateral amygdala), we investigated which part of the amygdala was targeted by the dlPFC and observed that it was the blocking-related amygdala region. Since the trials in which we observed dlPFC–amygdala coupling preceded the trials in which we observed the significant blocking effect, we speculate that the dlPFC-mediated shaping of amygdala responses plays a major role in establishing the blocking effect.

Because direct dlPFC–amygdala projections are rather sparse (Freese and Amaral, 2009), it has been suggested that the dlPFC exerts its regulatory effects on the amygdala via the vmPFC (Phelps, 2009), and there is some evidence supporting this account (Delgado et al., 2008a). Directly testing this assumption would require the specification and comparison of different models of effective connectivity (Stephan and Friston, 2010), which is beyond the scope of the present study. However, we indeed observed that vmPFC (subgenual anterior cingulate) was involved in the development of blocking, but its responses—which were characterized by a significant deactivation—were limited to the comparison of CSAX versus CSA. This selective response is especially interesting given this region's involvement in the control of conditioned fear in animals (Rosenkranz et al., 2003; Vidal-Gonzalez et al., 2006) and in both automatic and voluntary forms of fear regulation in humans (Phelps et al., 2004; Delgado et al., 2008a; Klumpp et al., 2010; Nili et al., 2010), but obviously it leaves open the question via which pathway the dlPFC-mediated upregulation of amygdala responses—as observed on BY trials—is realized.

Some limitations of the present study are also worth mentioning. First, although we acquired eye-tracking data, we could not analyze changes in pupil diameter (which frequently show significant conditioning-induced effects; Reinhard and Lachnit, 2002; Kluge et al., 2011) because our stimuli were not equiluminant. Second, the 100% reinforcement design used in combination with a 5 s interval between CS onset and US onset led to a small correlation between CS and US regressors, which might confound the comparison of CSA versus CSB to some degree. Third, the neurobiological results obtained here might not generalize to blocking effects in other domains (such as causal learning), where similar behavioral measures are used (expectancy ratings), but which likely have a different neurobiological implementation. Fourth, it is important to keep in mind that our study used different behavioral measures than studies that investigated blocking in animals (Kamin, 1968; Beckers et al., 2006), and it would therefore be important for future studies to take an approach that can be implemented in both animal and human studies of blocking, such as measuring fear-potentiated startle in aversive learning.

In conclusion, we could show that enhanced behavioral and amygdala responses to conditioned stimuli are highly sensitive to

blocking effects and can thus not be explained on the grounds of mere temporal contiguity, without referring to the informational content or predictive value of conditioned stimuli. Such a mechanism is of obvious advantage in a constantly changing environment, and our data suggest that prefrontal regions contribute to it by flexibly changing their coupling with the amygdala, based on the requirements posed by predictive relationships. Together, these data shed new light on the specific contributions of the human amygdala to aversive learning.

References

- Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR (2005) A mechanism for impaired fear recognition after amygdala damage. *Nature* 433:68–72.
- Anderson AK, Phelps EA (2001) Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature* 411:305–309.
- Bach DR, Weiskopf N, Dolan RJ (2011) A stable sparse fear memory trace in human amygdala. *J Neurosci* 31:9383–9389.
- Beckers T, De Houwer J, Pineño O, Miller RR (2005) Outcome additivity and outcome maximality influence cue competition in human causal learning. *J Exp Psychol Learn Mem Cogn* 31:238–249.
- Beckers T, Miller RR, De Houwer J, Urushihara K (2006) Reasoning rats: forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *J Exp Psychol Gen* 135:92–102.
- Beesley T, Le Pelley ME (2011) The influence of blocking on overt attention and associability in human learning. *J Exp Psychol Anim Behav Process* 37:114–120.
- Belova MA, Paton JJ, Morrison SE, Salzman CD (2007) Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55:970–984.
- Bermudez MA, Schultz W (2010) Responses of amygdala neurons to positive reward-predicting stimuli depend on background reward (contingency) rather than stimulus-reward pairing (contiguity). *J Neurophysiol* 103:1158–1170.
- Büchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* 20:947–957.
- Carter RM, O'Doherty JP, Seymour B, Koch C, Dolan RJ (2006) Contingency awareness in human aversive conditioning involves the middle frontal gyrus. *Neuroimage* 29:1007–1012.
- Crowne DP, Marlowe D (1960) A new scale of social desirability independent of psychopathology. *J Consult Psychol* 24:349–354.
- Davey GC, Singh J (1988) The Kamin “blocking” effect and electrodermal conditioning in humans. *J Psychophysiol* 2:17–25.
- Davis M, Whalen PJ (2001) The amygdala: vigilance and emotion. *Mol Psychiatry* 6:13–34.
- Dawson ME, Schell AM, Filion DL (2007) The electrodermal system. In: *Handbook of psychophysiology* (Cacioppo JT, Tassinary LG, Berntson GG, eds), pp 159–182. Cambridge, UK: Cambridge UP.
- Deckers RH, van Gelderen P, Ries M, Barret O, Duyn JH, Ikonomidou VN, Fukunaga M, Glover GH, de Zwart JA (2006) An adaptive filter for suppression of cardiac and respiratory noise in MRI time series data. *Neuroimage* 33:1072–1081.
- Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* 19:430–441.
- Delgado MR, Nearing KI, Ledoux JE, Phelps EA (2008a) Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron* 59:829–838.
- Delgado MR, Li J, Schiller D, Phelps EA (2008b) The role of the striatum in aversive learning and aversive prediction errors. *Philos Trans R Soc Lond B Biol Sci* 363:3787–3800.
- Dickinson A, Shanks D, Evenden J (1984) Judgement of act-outcome contingency: the role of selective attribution. *Q J Exp Psychol A* 36:29–50.
- Egger MD, Miller NE (1962) Secondary reinforcement in rats as a function of information value and reliability of the stimulus. *J Exp Psychol* 64:97–104.
- Eippert F, Bingel U, Schoell ED, Yacubian J, Klinger R, Lorenz J, Büchel C (2009) Activation of the opioidergic descending pain control system underlies placebo analgesia. *Neuron* 63:533–543.
- Fanselow MS (1998) Pavlovian conditioning, negative feedback, and blocking: mechanisms that regulate association formation. *Neuron* 20:625–627.

- Fletcher PC, Anderson JM, Shanks DR, Honey R, Carpenter TA, Donovan T, Papadakis N, Bullmore ET (2001) Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat Neurosci* 4:1043–1048.
- Fredrikson M (1981) Orienting and defensive reactions to phobic and conditioned fear stimuli in phobics and normals. *Psychophysiology* 18:456–465.
- Freese JL, Amaral DG (2009) Neuroanatomy of the primate amygdala. In: *The human amygdala* (Whalen PJ, Phelps EA, eds), pp 3–42. New York: Guilford Press.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229.
- Gamer M, Büchel C (2009) Amygdala activation predicts gaze toward fearful eyes. *J Neurosci* 29:9123–9126.
- Gamer M, Zurowski B, Büchel C (2010) Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc Natl Acad Sci U S A* 107:9400–9405.
- Gläscher J (2009) Visualization of group inference data in functional neuroimaging. *Neuroinformatics* 7:73–82.
- Gläscher J, Büchel C (2005) Formal learning theory dissociates brain regions with different temporal integration. *Neuron* 47:295–306.
- Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585–595.
- Glautier S (2002) Spatial separation of target and competitor cues enhances blocking of human causality judgements. *Q J Exp Psychol B* 55:121–135.
- Hamm AO, Vaitl D (1996) Affective learning: awareness and aversion. *Psychophysiology* 33:698–710.
- Hare RD, Blevings G (1975) Conditioned orienting and defensive responses. *Psychophysiology* 12:289–297.
- Haselgrove M (2010) Reasoning rats or associative animals? A common-element analysis of the effects of additive and subadditive pretraining on blocking. *J Exp Psychol Anim Behav Process* 36:296–306.
- Hinchey J, Lovibond PF, Ter-Horst KM (1995) Blocking in human electrodermal conditioning. *Q J Exp Psychol B* 48:2–12.
- Hodes RL, Cook EW 3rd, Lang PJ (1985) Individual differences in autonomic response: conditioned association or conditioned fear? *Psychophysiology* 22:545–560.
- Hogarth L, Dickinson A, Duka T (2010) Selective attention to conditioned stimuli in human discrimination learning: untangling the effects of outcome prediction, valence, arousal, and uncertainty. In: *Attention and associative learning* (Mitchell CJ, Le Pelley ME, eds), pp 71–98. Oxford, UK: Oxford UP.
- Holland PC, Gallagher M (1999) Amygdala circuitry in attentional and representational processes. *Trends Cogn Sci* 3:65–73.
- Holland PC, Han JS, Gallagher M (2000) Lesions of the amygdala central nucleus alter performance on a selective attention task. *J Neurosci* 20:6701–6706.
- Iordanova MD (2010) Dopamine transmission in the amygdala modulates surprise in an aversive blocking paradigm. *Behav Neurosci* 124:780–788.
- Iordanova MD, McNally GP, Westbrook RF (2006) Opioid receptors in the nucleus accumbens regulate attentional learning in the blocking paradigm. *J Neurosci* 26:4036–4045.
- Johansen JP, Cain CK, Ostroff LE, LeDoux JE (2011) Molecular mechanisms of fear learning and memory. *Cell* 147:509–524.
- Kamin LJ (1968) “Attention-like” processes in classical conditioning. In: *Miami symposium on the prediction of behavior, 1967: aversive stimulation* (Jones MR, ed), pp 9–31. Coral Gables, FL: University of Miami.
- Kim JJ, Krupa DJ, Thompson RF (1998) Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science* 279:570–573.
- Klinge C, Röder B, Büchel C (2010) Increased amygdala activation to emotional auditory stimuli in the blind. *Brain* 133:1729–1736.
- Kluge C, Bauer M, Leff AP, Heinze HJ, Dolan RJ, Driver J (2011) Plasticity of human auditory-evoked fields induced by shock conditioning and contingency reversal. *Proc Natl Acad Sci U S A* 108:12545–12550.
- Klumpers F, Raemackers MA, Ruigrok AN, Hermans EJ, Kenemans JL, Baas JM (2010) Prefrontal mechanisms of fear reduction after threat offset. *Biol Psychiatry* 68:1031–1038.
- Kruschke JK, Kappenman ES, Hetrick WP (2005) Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *J Exp Psychol Learn Mem Cogn* 31:830–845.
- LaBar KS, Gatenby JC, Gore JC, LeDoux JE, Phelps EA (1998) Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* 20:937–945.
- Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci* 14:1250–1252.
- Livesey EJ, Boakes RA (2004) Outcome additivity, elemental processing and blocking in human causality judgements. *Q J Exp Psychol B* 57:361–379.
- Lovibond PF, Siddle DA, Bond N (1988) Insensitivity to stimulus validity in human pavlovian conditioning. *Q J Exp Psychol B* 40:377–410.
- Lück HE, Timaeus E (1969) Scales for the measurement of manifest anxiety (MAS) and social desirability (SDS-E and SDS-CM). *Diagnostica* 15:134–141.
- Mackintosh NJ (1975) A theory of attention: variations in the associability of stimuli with reinforcement. *Psychol Rev* 82:276–298.
- Maren S, Quirk GJ (2004) Neuronal signalling of fear memory. *Nat Rev Neurosci* 5:844–852.
- Martin I, Levey AB (1991) Blocking observed in human eyelid conditioning. *Q J Exp Psychol B* 43:233–256.
- McNally GP, Cole S (2006) Opioid receptors in the midbrain periaqueductal gray regulate prediction errors during pavlovian fear conditioning. *Behav Neurosci* 120:313–323.
- McNally GP, Pigg M, Weidemann G (2004) Blocking, unblocking, and overexpectation of fear: a role for opioid receptors in the regulation of Pavlovian association formation. *Behav Neurosci* 118:111–120.
- McNally GP, Johansen JP, Blair HT (2011) Placing prediction into the fear circuit. *Trends Neurosci* 34:283–292.
- Mechias ML, Etkin A, Kalisch R (2010) A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. *Neuroimage* 49:1760–1768.
- Melchers KG, Shanks DR, Lachnit H (2008) Stimulus coding in human associative learning: flexible representations of parts and wholes. *Behav Processes* 77:413–427.
- Mitchell CJ, Lovibond PF (2002) Backward and forward blocking in human electrodermal conditioning: blocking requires an assumption of outcome additivity. *Q J Exp Psychol B* 55:311–329.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Nili U, Goldberg H, Weizman A, Dudai Y (2010) Fear thou not: activity of frontal and temporal circuits in moments of real-life courage. *Neuron* 66:949–962.
- Öhman A (2009) Human fear conditioning and the amygdala. In: *The human amygdala* (Whalen PJ, Phelps EA, eds), pp 118–154. New York: Guilford Press.
- Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439:865–870.
- Pavlov IP (1927) *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. London: Oxford UP.
- Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87:532–552.
- Phelps EA (2009) The human amygdala and the control of fear. In: *The human amygdala* (Whalen PJ, Phelps EA, eds), pp 204–219. New York: Guilford Press.
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43:897–905.
- Prados J (2011) Blocking and overshadowing in human geometry learning. *J Exp Psychol Anim Behav Process* 37:121–126.
- Quirk GJ, Repa C, LeDoux JE (1995) Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron* 15:1029–1039.
- Reinhard G, Lachnit H (2002) Differential conditioning of anticipatory pupillary dilation responses in humans. *Biol Psychol* 60:51–68.
- Repa JC, Muller J, Apergis J, Desrochers TM, Zhou Y, LeDoux JE (2001) Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nat Neurosci* 4:724–731.
- Rescorla RA (1981) Within-signal learning in autoshaping. *Anim Learn Behav* 9:245–252.
- Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: varia-

- tions in the effectiveness of reinforcement and nonreinforcement. In: *Classical conditioning II: current research and theory* (Black AH, Prokasy WF, eds), pp 64–99. New York: Appleton-Century-Crofts.
- Rosenkranz JA, Moore H, Grace AA (2003) The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J Neurosci* 23:11054–11064.
- Salvucci DD, Goldberg JH (2000) Identifying fixations and saccades in eye-tracking protocols. In: *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications*, pp 71–78. New York: ACM.
- Sander D, Grafman J, Zalla T (2003) The human amygdala: an evolved system for relevance detection. *Rev Neurosci* 14:303–316.
- Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008) From fear to safety and back: reversal of fear in the human brain. *J Neurosci* 28:11517–11525.
- Sehlmeyer C, Schöning S, Zwitserlood P, Pfleiderer B, Kircher T, Arolt V, Konrad C (2009) Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS One* 4:e5865.
- Shanks DR (2010) Learning: from association to cognition. *Annu Rev Psychol* 61:273–301.
- Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G (1970) *State-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Stephan KE, Friston KJ (2010) Analyzing effective connectivity with fMRI. *Wiley Interdiscip Rev Cogn Sci* 1:446–459.
- Tobler PN, O’Doherty JP, Dolan RJ, Schultz W (2006) Human neural learning depends on reward prediction errors in the blocking paradigm. *J Neurophysiol* 95:301–310.
- Turner DC, Aitken MR, Shanks DR, Sahakian BJ, Robbins TW, Schwarzbauer C, Fletcher PC (2004) The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cereb Cortex* 14:872–880.
- Vandorpe S, De Houwer J (2005) A comparison of forward blocking and reduced overshadowing in human causal learning. *Psychon Bull Rev* 12:945–949.
- Velden M, Wölk C (1987) Depicting cardiac activity over real time: a proposal for standardization. *J Psychophysiol* 1:173–175.
- Vidal-Gonzalez I, Vidal-Gonzalez B, Rauch SL, Quirk GJ (2006) Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learn Mem* 13:728–733.
- Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48.
- Wills AJ, Lavric A, Croft GS, Hodgson TL (2007) Predictive learning, prediction errors, and attention: evidence from event-related potentials and eye tracking. *J Cogn Neurosci* 19:843–854.