Behavioral/Systems/Cognitive

A Stable Sparse Fear Memory Trace in Human Amygdala

Dominik R. Bach, Nikolaus Weiskopf, and Raymond J. Dolan

Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom

Pavlovian fear conditioning is highly conserved across species, providing a powerful model of aversive learning. In rodents, fear memory is stored and reactivated under the influence of the amygdala. There is no evidence for an equivalent mechanism in primates, and an opposite mechanism is proposed whereby primate amygdala contributes only to an initial phase of aversive learning, subsequently ceding fear memory to extra-amygdalar regions. Here, we reexamine this question by exploiting human high-resolution functional magnetic resonance imaging in conjunction with multivariate methods. By assuming a sparse neural coding, we show it is possible, at an individual subject level, to discriminate responses to conditioned (CS+ and CS−) stimuli in both basolateral and centro-cortical amygdala nuclei. The strength of this discrimination increased over time and was tightly coupled to the behavioral expression of fear, consistent with an expression of a stable fear memory trace. These data highlight that the human basolateral and centro-cortical amygdala support initial learning as well more enduring fear memory storage. A sparse neuronal representation for fear, here revealed by multivariate pattern classification, resolves why an enduring memory trace has proven elusive in previous human studies.

Introduction

A substantial literature indicates the rodent amygdala stores a fear memory trace during and after delay conditioning (Maren, 1998, 1999; Zimmerman et al., 2007; Pape and Pare, 2010). Evidence that this is the case for humans is at best ambiguous (Mechias et al., 2010) (Table 1), and a different temporal course is proposed where an enhanced conditioned stimulus (CS+ vs CS−) response during learning habituates after a few CS presentations in human functional magnetic resonance imaging (fMRI) studies (Büchel et al., 1998; Morris et al., 2001; Morris and Dolan, 2004; Marschner et al., 2008). This is diametrically opposite to the generality of rodent data and reinforces the view that the primate and human amygdala is concerned solely with initial learning, and not with storage of a stable CS+/unconditioned stimulus (US) association. One interpretation of these data is that this initial amygdala response reflects deployment of attentional resources toward a CS with uncertain predictive value (Whalen, 1998; Sander et al., 2003).

Functional neuroimaging studies of fear learning in humans have heretofore rested on mass-univariate fMRI methods, reporting a higher mean response to a CS+, presumably elicited by activation of a large neural mass. Yet such findings ignore evidence from rodents that the CS–US association is stored in a small number of sparsely distributed neurons (Reijmers et al., 2007). If these units, once activated, suppress surrounding neurons, then this would provide a pattern of activity not detectable with mass-univariate approaches. Although the resolution of fMRI is orders of magnitude lower than the size of individual neurons, an uneven distribution of active and inactive neurons will generate a biased signal within individual voxels such that some will consistently show more, and others less, signal (Norman et al., 2006; Swisher et al., 2010). This possibility provides a powerful motivation for exploiting multivariate analysis, an approach commonplace for pattern decoding in vision and memory research. Here, our interest lay not in decoding but in detecting multivariate patterns of responses that could reflect activity of sparsely distributed neurons that we predict encode a learnt CS–US association.

We hypothesized a temporally extended BOLD response pattern for CS+ relative to a CS−, a pattern reflecting the amygdala's role in learning and expression of conditioned fear. As we were interested in a mechanism that is invariant across individuals, we analyzed individual subjects and focused our approach on showing replicability across these individuals rather than exploiting a conventional approach of group analysis. Thus, we report data from six healthy persons, scanned with isotropic 1.5 mm high-resolution fMRI (see Fig. 2*A*) while engaged in 180 trials of a standard delay conditioning task, where one of two differently colored circles (the CS+) predicted a 50% probability of receiving an electric shock (US) 3.5 s after CS onset, and the other one (the CS−) predicted the absence of electric shock (Fig. 1). Anticipatory skin conductance responses (Bach et al., 2010a) (aSCR) served as our behavioral index of aversive learning. Note that we only analyzed trials without a US to avoid confounds associated with an overlap in CS and US responses.

Materials and Methods

The study was approved by a local ethics committee. Seven healthy right-handed individuals (five male; two female; mean age ± SD, 23.1 ± 3.5 years) without history of psychiatric or neurological disease took part in a standard delay conditioning paradigm during functional imaging; one

Received March 25, 2011; revised May 10, 2011; accepted May 19, 2011.

Author contributions: D.R.B. and R.J.D. designed research; D.R.B. performed research; N.W. contributed unpublished reagents/analytic tools; D.R.B. analyzed data; D.R.B., N.W., and R.J.D. wrote the paper.

This work was supported by a program grant and strategic award from the Wellcome Trust and a Max Planck Research Award from the Max Planck Gesellschaft to R.J.D. We thank Zoltan Nagy, Karl Friston, Guillaume Flandin, and John Ashburner for methodological support. Some of the methods used here were developed in collaboration with Jean Daunizeau and Tim Behrens.

Correspondence should be addressed to Dominik R. Bach, Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London WC1N 3BG, United Kingdom. E-mail: d.bach@fil.ion.ucl.ac.uk.

DOI:10.1523/JNEUROSCI.1524-11.2011

 0270-6474/11/319383-07$15.00/0

Table 1. Overview of fMRI delay conditioning experiments

Author	CS	US	CS+ > CS−	Habituation	SCR recorded	aSCR
LaBar et al., 1998	Colored shapes	Electric shock	✘✘*	✘✘	✘	
Büchel et al., 1998	Neutral faces	Loud noise	✘✘	✓✓	✓	✘✘
Morris et al., 1998	Angry faces	Loud noise	✓✓	✘✘	✓	✘✘
Knight et al., 1999	Colored lights	Electric shock	✘✘	✘✘	✘	
Pine et al., 2001	Colored lights	Pressure pain	✘✘	✘✘*	✘	
Morris et al., 2001	Angry faces	Loud noise	✓✘	✘✓	✘	
Veit et al., 2002	Neutral faces	Pressure pain	✘✘	✘✘*	✓	✘✘
Gottfried et al., 2002	Neutral faces	Unpleasant odor	✘✘	✘✘	✘	
Armony and Dolan, 2002	Angry faces	Loud noise	✓✓	✘✘	✘	
Cheng et al., 2003	Colored lights	Electric shock	✘✘	✘✘	✓	✘✘*
Gottfried and Dolan, 2004	Neutral faces	Unpleasant odor	✘✓	✘✘	✘	
Morris and Dolan, 2004	Neutral faces	Loud noise	✓✘	✘✓	✘	
Knight et al., 2004a	Colored shapes	Electric shock	✘✘	✘✘	✓	✘✘
Knight et al., 2004b	Colored lights	Electric shock	✘✘	✘✘	✓	✘✘
Birbaumer et al., 2005	Neutral faces	Pressure pain	✓✘	✘✘*	✓	✘✘
Knight et al., 2005	Sine tones	Loud noise	✘✘	✘✘	✓	✘✘*
Tabbert et al., 2006	Geometric shapes	Electric shock	✘✓	✘✘	✓	✘✘
Cheng et al., 2006	Colored shapes	Electric shock	✘✘	✘✘	✓	✘✘*
Cheng et al., 2007	Geometric shapes	Electric shock	✘✘	✘✘	✓	✘✘*
Milad et al., 2007	Colored lights	Electric shock	✘✓	✘✘	✓	✘✘
Jensen et al., 2008	Colored shapes	Electric shock	✘✘	✘✘	✓	✘✘
Schiller et al., 2008	Angry faces	Electric shock	✘✘*	✘✘	✓	✘✘
Delgado et al., 2008	Colored shapes	Electric shock	✘✘	✘✘	✓	✘✘
Marschner et al., 2008	Geometric shapes	Electric shock	✘✘	✘✓	✓	✘✘
Knight et al., 2009	Sine tones	Loud noise	✘✓	✘✘	✓	✘✘

Experiments were identified by using the keywords "fMRI" or "BOLD" and "conditioning" in PubMed and iteratively searching references of each identified study. For drug/patient studies, only the control condition is listed here. Several studies are not listed because they did not report results from the initial learning phase (Glascher and Buchel, 2005; Kalisch et al., 2006, 2009) or they were drug/patient studies not reporting results for the (full) placebo/control group (Critchley et al., 2002; Eippert et al., 2008). Also, studies that implicitly used an instructed learning paradigm (Jensen et al., 2003) are not included, or only the condition not involving instructed fear is listed (Tabbert et al., 2006). In a study investigating emotion regulation strategies, only the main effects across the different strategies are listed (Delgado et al., 2008). In order to control for false positives, results are only shown when they survived small volume or whole-brain correction for multiple testing at least at $p < 0.05$ (false-defense ratio or family-wise error).

We indicate, for both hemispheres separately, whether amygdala activity was reported for the contrasts CS+ > CS−, differential habituation, and, if SCR was recorded, for a correlation with anticipatory SCRs. ✘, No activation reported; ✓, activation reported.

*, Authors reported activation at an uncorrected level.

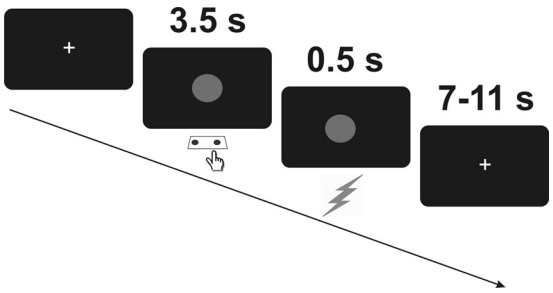

Figure 1. Experimental timeline for one of 180 trials. The CS+ or CS− appears after a fixation cross for 4 s, during which the participant indicates the color of the CS with a key press. In punished CS+ trials, it coterminates with an electric shock of 500 ms duration. The intertrial interval is randomly varied between 7 and 11 s.

male subject was excluded from fMRI analysis due to movement artifacts, thus we provide data from $n = 6$ individuals. An eighth participant had a panic attack during functional scanning and did not complete the study. CS stimuli were a blue and an orange circle, and US a discomforting electric shock. CS color was balanced across participants; CS were presented in the center of a screen/mirror system on a black background. Participants were tasked to indicate the color with a left/right button press on a keypad held in the right hand. CS stayed on the screen for 4 s; in punished trials, a US followed 3.5 s after CS onset as a 500 Hz train of electrical pulses (square wave, individual pulse duration: 1 ms; total duration: 500 ms, 400 V; mean current ± SD: 0.22 ± 0.06 mA) via a pin-shaped electrode attached to the left forearm, and coterminated with the CS. The intertrial interval was randomly determined on each trial to be 7, 9, or 11 s. Between trials, a white fixation cross was shown on the screen. There were 180 trials: 90 presentations for each CS+ and CS−; CS+ was followed by US on 45 trials. The experiment was divided into four blocks of 45 trials each. Only CS+ trials without US were analyzed to avoid contamination of CS and US responses in BOLD and SCR. Skin conductance was recorded from the left second and third finger (Bach et al., 2010b) and anticipatory sympathetic responses were extracted in a model-based approach (Bach et al., 2010a). Heart rate and respiration were monitored using a pulse oxymeter and breathing belt (Hutton et al., 2011).

At the beginning of each experiment, we acquired field maps with a double echo gradient-echo fast low-angle shot (FLASH) sequence (TE, 10.0 and 12.46 ms; TR, 1020 ms; matrix size, 64 × 64) using 64 slices covering the whole head (voxel size, 3 × 3 × 3 mm). We acquired T2*-weighted single-shot gradient-echo echo-planar images (EPI) in oblique transverse orientation [flip angle α, 90°; bandwidth (BW), 1953 Hz/pixel; phase-encoding (PE) direction, anterior–posterior; bandwidth in PE direction, 14 Hz/pixel; TE, 30 ms; asymmetric echo shifted forward by 26 PE lines; effective TR, 3000 ms). The manufacturer's standard automatic 3D-shim procedure was performed before functional scanning. Each EPI volume contained 30 contiguous slices of 1.5 mm thickness (field of view, 192 × 192 mm^2; matrix size, 128 × 128). BOLD sensitivity losses in the orbitofrontal cortex and the amygdala due to susceptibility artifacts were intrinsically minimized by the high spatial resolution and by applying a z shim gradient moment of −0.4 mT/m·ms, a slice tilt of −30°, and a positive PE gradient polarity (Weiskopf et al., 2006, 2007a). We acquired four sessions of 202 volumes with on-line image reconstruction and real-time image quality assurance (Weiskopf et al., 2007b). The first four volumes per session were discarded to allow for T1 equilibration, and each session was concluded by 10 volumes without stimulus presentation. For coregistration of functional images, we acquired a T1-weighted image using a 3D FLASH sequence (isotropic spatial resolution, 1 mm; α, 18°; TR, 9.0 ms; TE, 3.5 ms).

High-resolution 3D modified driven equilibrium Fourier transformation (MDEFT) anatomical images were acquired on a 3 T Trio whole-body scanner (Siemens). Two hundred twenty-four sagittal partitions

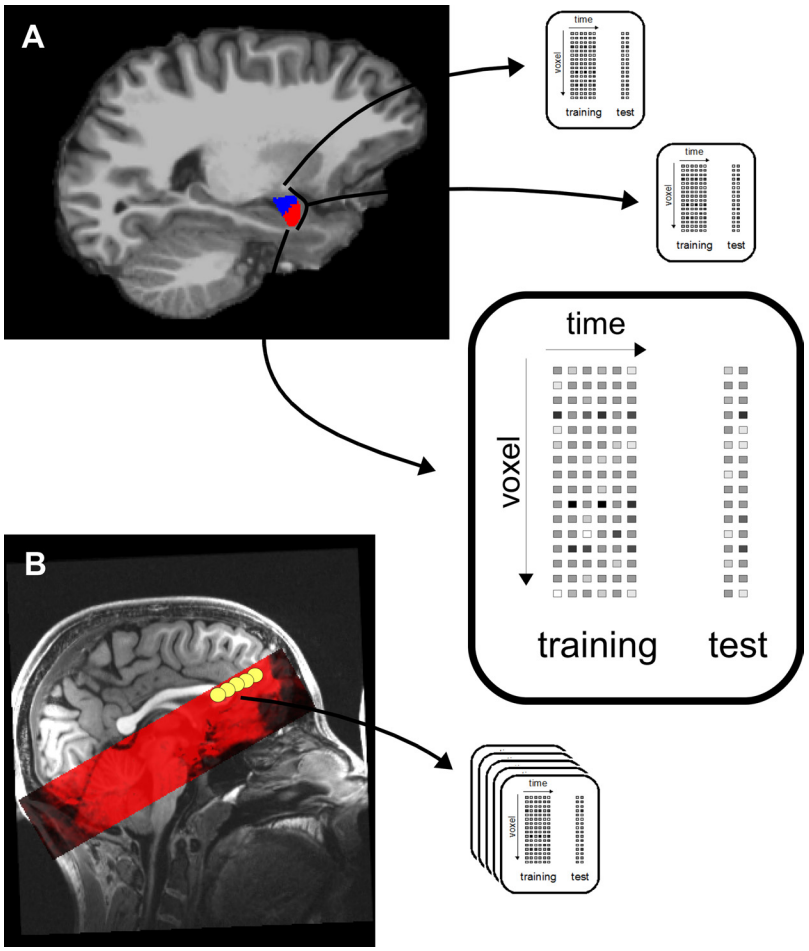

Figure 2. For functional imaging, 30 oblique transverse slices were centered over the amygdala; here overlaid on the 0.77 mm high-resolution T1-weighted image (***B***). The region of interest, manually segmented on high-resolution T1-weighted images, was parcellated into basolateral (blue) and centro-cortical (red) nucleus group, based on anatomical connectivity profiles determined by probabilistic tractography of diffusion weighted images (***A***). Data from each voxel and each trial of the regions of interest were entered into multivariate analysis. For searchlight analysis, data were extracted from a moving sphere and results were mapped onto the center of the sphere.

Table 2. Multivariate analysis results for each individual and region of interest

	Full amygdala			Deep			Superficial		
	L	R	B	L	R	B	L	R	B
CS+ versus CS−									
Participant 1	*		*			*			
Participant 2	*		**				*		
Participant 3							*		
Participant 4		*	*	*				*	
Participant 5	*	*	**				*		*
Participant 6	*		*	**	*	*	*		
Time effect CS+ versus time effect CS−									
Participant 1	*			*		**			
Participant 2		**	**					**	**
Participant 3									
Participant 4	**		**	**		**			
Participant 5									
Participant 6	*		**	*		**			
aSCR, controlling for CS+ versus CS−									
Participant 1	*		*	**		*			
Participant 2	**		*				**		**
Participant 3	**	*	**	**	*	*		*	*
Participant 4			**	**		*			*
Participant 5	**	**	**		**	**	**	**	**
Participant 6					*				

Regions of interest were full amygdala and deep and superficial nucleus groups, each for the left (L) and right (R) hemispheres and combined for both (B) hemispheres. *$p < 0.05$ and **$p < 0.01$ for the permutation test.

were acquired twice with an image matrix of 304 × 288 (read × phase) and twofold oversampling in read direction (head/foot direction) to prevent aliasing (isotropic spatial resolution, 0.77 mm; α, 16°; TR, 7.92 ms; TE, 2.48 ms; TI, 910 ms; BW, 196 Hz/pixel). Special radio frequency excitation pulses were used to compensate for B1 inhomogeneities of the transmit coil in superior/inferior and anterior/posterior directions (Deichmann et al., 2004). Images were reconstructed by performing a standard 3D Fourier transform, followed by modulus calculation. No data filtering was applied in *k* space or in the image domain. The two images were realigned and averaged offline using SPM8 functions.

To provide an accurate definition of our region of interest, the amygdala was manually delineated on T1-weighted images using Anatomist (http://www.brainvisa.info) as described previously (Bach et al., 2011). The inferior/posterior and anterior/superior boundary, in general clearly visible on at least a few sagittal slices, were marked; the most posterior points were the posterior nuclei bordering the ventral horn of the anterior extent of the lateral (temporal) ventricle and white matter; the inferior boundary separating amygdala from hippocampus and lateral ventricle; and the anterior boundary separating amygdala from white matter, entorhinal cortex, gyrus ambiens, and uncus. We then proceeded from posterior to anterior in coronal slices, using the sagitally marked boundaries, hippocampus, optical tract, and sulcus semiannularis as guiding landmarks. Each slice was compared against schematic tables of an anatomical atlas (Mai et al., 2008). Particular care was taken not to include the peduncle of the lentiform nucleus, hippocampal tissue, or periamygdaloid tissue between lateral amygdala and white matter of the temporal lobe. Amygdala boundaries were then straightened in sagittal slices and once more controlled in coronal slices. Mask boundaries were automatically smoothed using the SPM8 (http://www.fil.ion.ucl.ac.uk/spm) functions spm_erode and spm_dilate.

The manually segmented amygdala region of interest was parcellated into two nucleus groups (Fig. 2*A*) based on their connectivity profile, as described previously (Bach et al., 2011). In brief, we acquired diffusion-weighted images (Nagy et al., 2007), corrected susceptibility-induced distortion (Andersson et al., 2003), and computed fiber tracts from the amygdala to lateral orbitofrontal cortex and temporal pole (Behrens et al., 2003, 2007). The connectivity profile for each amygdala voxel was fed into an automatic *k*-means clustering procedure that generated two spatially contiguous clusters, one of which connected stronger to the lateral orbitofrontal cortex, the other to the temporal pole. Both anatomical location of the clusters and their connectivity profile suggest that they correspond to the basolateral and centro-cortical nucleus group.

Functional images were analyzed in a standard preprocessing pipeline in SPM8. EPI images were generated off-line from the complex *k*-space raw data using a generalized reconstruction method based on the measured EPI *k*-space trajectory to minimize ghosting. They were then corrected for geometric distortions caused by susceptibility-induced field inhomogeneities. A combined approach was used that corrects for both static distortions and changes in these distortions due to head motion (Andersson et al., 2001; Hutton et al., 2002). The static distortions were calculated for each subject from a field map that was estimated from the double-echo FLASH images using the FieldMap toolbox as implemented in SPM8. Using these parameters, the EPI images were then realigned and unwarped, a procedure that allows the measured static distortions to be

included in the estimation of distortion changes associated with head motion. The motion-corrected images were then coregistered to an unwarped EPI image with 90 slices and whole-brain coverage, which was in turn aligned to the individual's anatomical 1-mm-resolution 3D FLASH image using a 12-parameter affine transformation; masks for the regions of interest generated from 0.77-mm-resolution T1-weighted images were coregistered to the 3D FLASH image as well.

For multivariate analysis, we estimated the response amplitude per trial (Schurger et al., 2010) by using, for each trial, one regressor for CS onset and another for US onset, without differentiating between the different CS. No time derivative was used, and as regressors of no interest we added cardiac phase (10 regressors), respiratory phase (6 regressors), and respiratory volume (1 regressor), similar to the method described by Hutton et al. (2011). This model is necessarily underspecified such that parameter estimates have little precision; however, this will increase noise in the estimates and is agnostic to the difference between a CS+ not followed by a US, and a CS−, hence the resulting parameter estimates will be unbiased with respect to our hypotheses. For each region of interest, parameter estimates for all CS not followed by US were extracted and *z*-transformed voxelwise. Because there is evidence for a time-dependency of BOLD responses in fear conditioning, we collapsed data from all sessions and used a leave-one-out scheme wherein every third CS− and every third CS+ were separated as test dataset, and the rest of the data were used as training data for feature selection and support vector machines or support vector regression (SVR), respectively. The null performance for this interleaved data scheme was determined empirically (see below). Target variables were the dichotomous contrast CS−/CS+, the *z*-transformed aSCR estimate orthogonalized to this contrast, and the time effect of CS+ versus CS−. To assess the latter, a simple linear regression model was constructed for each voxel, with time as predictor and responses to the CS− as predicted variable (independently for training and test dataset). Using parameters from this model, we made a prediction for CS+ responses from time, and subtracted these from the actual CS+ responses. Those residuals from the CS+ responses were used as data points, and the time points of the CS+ as target variable. Feature selection was based on the univariate relation between each voxel and the target variable; the 300 voxels with the highest explained variance on a univariate basis were extracted (if there were <300 voxels in a region of interest, all voxels were used). Three hundred features is an arbitrary threshold; using other feature numbers between 200 and 400 yielded similar results. Feature selection was based on the training dataset and surplus features were removed both from training and test dataset. The training data were fed into a linear support vector classification (C-SVC) or support vector regression (ε-SVR), as implemented in LIBSVM (Chang and Lin, 2001). The ensuing model was used to predict the test data from the target. The critical variables were the classification accuracy for SVC and explained variance in SVR. Because we did not want to make any assumptions about the distribution of cross-validation performance under the null hypothesis, we used a randomization test where the target variable was randomly permuted 1000 times within training and test dataset, and the classification procedure repeated each time.

For all group analyses, we extracted the performance (i.e., classification or explained variance) above chance level and tested this against zero with one-sample *t* tests. Region-of-interest analysis was based on the performance for each individual hemisphere and each participant. Similarly, the difference between nucleus groups was tested with paired *t* tests. However, responses might be particularly pronounced in a very small amygdala area, and this might be missed by focusing on regions of interest, such as the two amygdala nucleus groups. Hence, we applied a searchlight within the amygdala (Kriegeskorte et al., 2006). For each voxel within the amygdala, a sphere of 5 mm radius was constructed around it, and all voxels within this sphere were used for multivariate analysis. In a searchlight approach, a high number of random permutations is not feasible, as we were not interested in an exact test and simply sought to determine the null performance; this was achieved with 100 random permutations. The performance above chance level was mapped on the center voxel of the sphere. To locate responses across the group, we extracted the performance peak across the whole amygdala and mean performance for the two regions of interest. For all analyses, peaks were evenly distributed between both nucleus groups and mean performance did not differ between them.

Similar searchlight analysis was applied to the whole field of view with 10 random permutations to approximate the null performance. The resulting images were smoothed with an isotropic 8 mm full width at half maximum Gaussian filter to accommodate interindividual differences in anatomy and aligned using DARTEL (Ashburner, 2007), based on segmented (Ashburner and Friston, 2005) high-resolution T1-weighted images. A second-level test was then performed on these images with a voxel-level threshold of $p < 0.001$. As we regard this as an exploratory approach, we present results at a cluster-level false-defense ratio of $q < 0.05$. Peak coordinates were then affine transformed to MNI space.

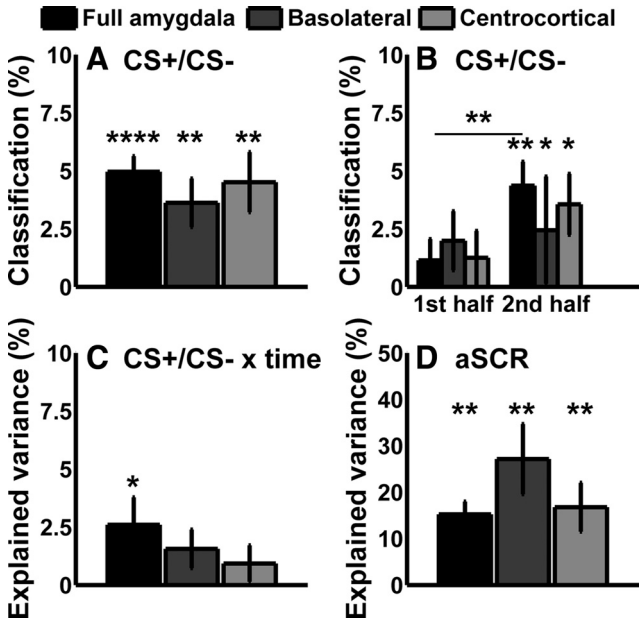

Figure 3. ***A***, Classification above chance for the contrast CS+ and CS− separately for the full amygdala and both amygdala nucleus groups. ***B***, Same analysis, separated for first and second halves of the experiment. ***C***, Explained variance above chance for the difference in linear time effects between CS+ and CS−. ***D***, Explained variance above chance for an association of amygdala neural responses with aSCR. *$p < 0.05$, **$p < 0.01$, ****$p < 0.0001$.

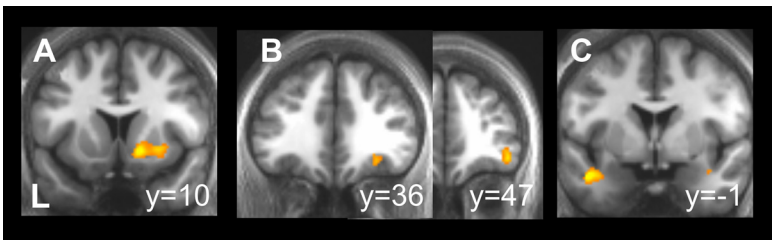

Figure 4. Brain areas outside of our region of interest that show an association with the target variable, detected by one-sample *t* tests of smoothed cross-validation performance in a 5 mm radius searchlight analysis. Results shown here are significant at $p < 0.001$ at the voxel level, whole-brain corrected at a false defense ratio of $q < 0.05$. See Table 3 for additional information. ***A***, ***B***, Pattern separation for CS−/CS+ in right putamen and middle frontal gyrus. ***C***, Linear pattern evolution in left (L) medial temporal gyrus.

Results

Fear conditioning was successful, as indicated by higher aSCR to the CS+ than to the CS− across the whole experiment (Wilcox-

Table 3. Brain areas that consistently exhibit a multivariate BOLD signal pattern across participants, detected by one-sample *t* test of smoothed cross-validation performance in a 5 mm radius searchlight analysis

Brain regions	Brodmann area of local maxima	Hemi-sphere	Voxel count	Peak voxel *t* score	Montreal Neurological Institute brain template coordinates of local maxima
CS+ <> CS−					
Cerebellum		Bilateral	532	25.6	8, −50, −36; 5, −43, −36; −11, −44, −36
Middle frontal gyrus	11 and 47	Right	320	9.52	43, 48, −11; 35, 30, −11; 28, 36, −12
Putamen		Right	289	11.69	15, 11, −8; 27, 11, −7
Time effect CS+ <> CS−					
Cerebellum		Left	942	16.58	−5, −37, −32; −7, −56, −39; −12, −73, −36
Middle temporal gyrus	20 & 21	Left	450	31.91	−47, 1, −32; −41, −9, −36; −44, 7, −26

Results are significant at $p < 0.001$ at the voxel level, whole-brain corrected at the cluster level for a false defense ratio of $q < 0.05$.

on's signed rank test, $p < 0.05$). Participants were at least partially aware of the association as they rated the US likelihood higher after a CS+ than after a CS− when debriefed *post hoc* (signed rank test, $p < 0.05$).

On a single-subject level, we found a significant difference (permutation test, $p < 0.05$) in BOLD pattern in the amygdala between CS+ and CS− in five of six individuals (Table 2). Classification performance across individuals was above chance at $p < 0.0001$ (Fig. 3*A*). Similar results were found for basolateral and centro-cortical amygdala analyzed independently (both at $p < 0.01$ across individuals). There was no difference between the two nuclear groups, as confirmed with a searchlight approach. Hence, our analysis was sensitive enough to detect a pattern of BOLD responses that differentiates between CS+ and CS− presentations within and across both principal amygdala nuclear groups.

To refute a hypothesis that the amygdala is only involved in initial learning, we repeated our analysis separately for the first and second halves of the experiment (90 trials each) (Fig. 3*B*). We found that classification was better in the second than in the first half of the experiment ($p < 0.01$), consistent with an interpretation that the amygdala encodes a pattern of response that is different in relation to a CS+ and CS− and where this difference becomes more pronounced with time. This suggests an evolution of the strength of aversive memory over the timescale of our experiment.

A temporal evolution of a neuronal pattern should be detectable (as a first approximation) as a linear change over time, and the pattern change should distinguish between a CS+ and a CS− as two separate patterns emerge. Hence, we exploited support vector regression to detect linearly changing patterns where the temporal profile of this change differs between CS+ and CS− (Fig. 3*C*). A differential pattern change, consistent with our predictions, was seen in three of six individuals, as well as being evident also at the group level ($p < 0.05$ level). This effect was not different between the two principal nuclear groups.

The behavioral expression of fear fluctuates from trial to trial, as quantified here in aSCR. The expression of fear memory should be linked to a differential pattern in the amygdala, if indeed this pattern reflects the establishment of a stable memory representation. To address this, we next estimated multivariate responses associated with aSCR after transformation to ensure independence from the CS+/CS− contrast (Fig. 3*D*). We found a significant association in five of six individuals (group level, $p < 0.01$), with the same responses being replicated for both nuclear groups and no difference evident between these loci. The explained variance was much higher than for the simple contrasts but variance across individuals was also larger, consistent with the observations that trial-by-trial estimates of expressed fear are inherently noisy.

Our results indicate that a fear memory trace is encoded in the human amygdala as a sparsely distributed stable representation. This finding might seem inconsistent with a recent primate study showing that posttraining amygdala lesions do not destroy fear-potentiated startle in rhesus monkeys (Antoniadis et al., 2007), suggesting that additional fear memory traces are expressed in other brain structures. To address this apparent anomaly, we applied a multivariate searchlight technique (Kriegeskorte et al., 2006) across the whole-brain volume. This approach is also able to demonstrate the specificity of our effect. We replicated the expression of specific CS+/CS− patterns in the amygdala (199 voxel, $T_{PEAK} = 13.6$, peaks at 28, −8, −14; 33, −18, −11; 41, −16, −14 mm). Notably, extra-amygdala responses to the CS+/CS− contrast were seen in cerebellum, putamen, and middle frontal gyrus (Fig. 4, Table 3). An additional responses cluster in the left hippocampus and parahippocampal gyrus was too small to survive correction for multiple comparison ($q = 0.09$; 133 voxel; $T_{PEAK} = 10.0$; peaks at −38, −22, −16; −27, −21, −9 mm). These regions represent candidate areas that could support access to fear memory when an amygdala representation is not available (Antoniadis et al., 2007, 2009). Note that our limited field of view may preclude detection of additional areas encoding such patterns.

Discussion

Our finding that BOLD response patterns in the amygdala of individual subjects differ for CS+ and CS− across 180 CS presentations challenges the view that within primates the amygdala only accounts for initial learning (Büchel et al., 1998; Morris et al., 2001; Morris and Dolan, 2004; Marschner et al., 2008) or is involved solely in detecting uncertain predictive value (Whalen, 1998). Crucially, CS+/CS− differentiation was better in the second half of the experiment, consistent with an expectation that a fear memory trace evolves over time. Even if a memory trace is not exclusively stored in the amygdala, as indicated by findings that amygdala lesions can leave fear memory intact (Antoniadis et al., 2007, 2009), our data point to the amygdala being a key component of a network activated when a CS+ is presented, in effect leading to reactivation of fear memory.

The differential CS+/CS− encoding we describe is significant for anatomically distinct amygdala subregions, suggesting that different parts of the amygdala generate this pattern. This fits both a parallel model of amygdala function, where basolateral and centromedial nuclei both store the CS+-US association (Paré et al., 2004; Balleine and Killcross, 2006), and a classic serial model, where the basolateral amygdala stores the association and the centromedial nucleus serves as an output relay (LeDoux, 2000; Pape and Pare, 2010). However, our results are a challenge to a model of fear learning that implicates only neuronal units in basolateral amygdala (Koo et al., 2004).

Our experiment lasted ~45 min. An important question is how long the human amygdala stores fear memory in the absence

of CS presentations. This is difficult to assess with noninvasive methods that require a large number of experimental events. Indeed, after a significant lapse of time, when a CS–US association is possibly attenuated, both reinforcement and non-reinforcement are likely to engage further learning processes. This points to a need to develop new assessment strategies that are sufficiently sensitive to harvest data from a very few experimental trials.

Distributed neural representations are key to understanding phenomena such as encoding of fine visual features or detailed memory. Our findings highlight the importance of multivariate fMRI methods in detecting such neural patterns in the acquisition and storage of memory. This point is reinforced by a surprising consideration: the majority of previous human fMRI studies using mass-univariate approaches fail to find stronger BOLD responses to a predictor of an aversive outcome (CS+) (Table 1). Our results indicate that these approaches are not sufficiently sensitive to a neural pattern that encodes fear memory (Reijmers et al., 2007), and that a rapid habituating neural mass response previously reported (Büchel et al., 1998; Morris et al., 2001; Morris and Dolan, 2004; Marschner et al., 2008) might reflect an entirely different mechanism. While we acknowledge our small sample size, we believe the strength of our approach is its ability to show replicability across individuals, which is what one might expect for a fundamental and highly conserved learning mechanism.

Exploiting multivariate approaches allows a more direct comparison between single-unit recordings in rodents and human brain responses. Our data appear to show similarities and differences between rodent and primate learning. The demonstration of a stable and strengthening fear memory trace in the human amygdala refutes a diametrically opposite assumption regarding Pavlovian fear conditioning in primates, including humans. Instead, we highlight a convergence in neuronal events supporting Pavlovian learning in human and rodent amygdala. In keeping with most, but not all (Koo et al., 2004), rodent findings, we demonstrate involvement of both main amygdala nuclei in fear learning. In contrast, the likelihood of species differences is supported by evidence that fear memory in rodents may not involve structures outside the amygdala. Numerous studies show that both conditioned freezing (Maren, 1998, 1999; Zimmerman et al., 2007) and fear-potentiated startle (Campeau and Davis, 1995) are disrupted when the rodent amygdala is removed posttraining. In monkeys, however, fear-potentiated startle is not disrupted with posttraining amygdala lesions (Antoniadis et al., 2007), although they prevent the establishment of new fear memory (Antoniadis et al., 2009). Our demonstration of an extra-amygdala contribution to fear memory is in keeping with the latter observation. We do note, however, that our standard human delay conditioning paradigm involves other demands than those posed in rodent research and that such factors might explain differences in the neural structures that support fear memory. To summarize, fear learning in humans rests on a sparse neural code, a finding that highlights convergence and divergence with findings from other species.

References

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modeling geometric deformations in EPI time series. Neuroimage 13:903–919.

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870–888.

Antoniadis EA, Winslow JT, Davis M, Amaral DG (2007) Role of the primate amygdala in fear-potentiated startle: effects of chronic lesions in the rhesus monkey. J Neurosci 27:7386–7396.

Antoniadis EA, Winslow JT, Davis M, Amaral DG (2009) The nonhuman primate amygdala is necessary for the acquisition but not the retention of fear-potentiated startle. Biol Psychiatry 65:241–248.

Armony JL, Dolan RJ (2002) Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. Neuropsychologia 40:817–826.

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113.

Ashburner J, Friston KJ (2005) Unified segmentation. Neuroimage 26:839–851.

Bach DR, Daunizeau J, Friston KJ, Dolan RJ (2010a) Dynamic causal modelling of anticipatory skin conductance responses. Biol Psychol 85:163–170.

Bach DR, Flandin G, Friston KJ, Dolan RJ (2010b) Modelling event-related skin conductance responses. Int J Psychophysiol 75:349–356.

Bach DR, Behrens TE, Garrido L, Weiskopf N, Dolan RJ (2011) Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. J Neurosci 31:618–623.

Balleine BW, Killcross S (2006) Parallel incentive processing: an integrated view of amygdala function. Trends Neurosci 29:272–279.

Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM (2003) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 50:1077–1088.

Behrens TE, Berg HJ, Jbabdi S, Rushworth MF, Woolrich MW (2007) Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage 34:144–155.

Birbaumer N, Veit R, Lotze M, Erb M, Hermann C, Grodd W, Flor H (2005) Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. Arch Gen Psychiatry 62:799–805.

Büchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron 20:947–957.

Campeau S, Davis M (1995) Involvement of the central nucleus and basolateral complex of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. J Neurosci 15:2301–2311.

Chang CC, Lin CL (2001) LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Cheng DT, Knight DC, Smith CN, Stein EA, Helmstetter FJ (2003) Functional MRI of human amygdala activity during Pavlovian fear conditioning: stimulus processing versus response expression. Behav Neurosci 117:3–10.

Cheng DT, Knight DC, Smith CN, Helmstetter FJ (2006) Human amygdala activity during the expression of fear responses. Behav Neurosci 120:1187–1195.

Cheng DT, Richards J, Helmstetter FJ (2007) Activity in the human amygdala corresponds to early, rather than late period autonomic responses to a signal for shock. Learn Mem 14:485–490.

Critchley HD, Mathias CJ, Dolan RJ (2002) Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. Neuron 33:653–663.

Deichmann R, Schwarzbauer C, Turner R (2004) Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. Neuroimage 21:757–767.

Delgado MR, Nearing KI, LeDoux JE, Phelps EA (2008) Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. Neuron 59:829–838.

Eippert F, Bingel U, Schoell E, Yacubian J, Büchel C (2008) Blockade of endogenous opioid neurotransmission enhances acquisition of conditioned fear in humans. J Neurosci 28:5465–5472.

Gläscher J, Büchel C (2005) Formal learning theory dissociates brain regions with different temporal integration. Neuron 47:295–306.

Gottfried JA, Dolan RJ (2004) Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. Nat Neurosci 7:1144–1152.

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

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. Neuroimage 16:217–240.

Hutton C, Josephs O, Stadler J, Featherstone E, Reid A, Speck O, Bernarding

J, Weiskopf N (2011) The impact of physiological noise correction on fMRI at 7T. Neuroimage 57:101–112.
Jensen J, McIntosh AR, Crawley AP, Mikulis DJ, Remington G, Kapur S (2003) Direct activation of the ventral striatum in anticipation of aversive stimuli. Neuron 40:1251–1257.
Jensen J, Willeit M, Zipursky RB, Savina I, Smith AJ, Menon M, Crawley AP, Kapur S (2008) The formation of abnormal associations in schizophrenia: neural and behavioral evidence. Neuropsychopharmacology 33:473–479.
Kalisch R, Korenfeld E, Stephan KE, Weiskopf N, Seymour B, Dolan RJ (2006) Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. J Neurosci 26:9503–9511.
Kalisch R, Holt B, Petrovic P, De Martino B, Klöppel S, Büchel C, Dolan RJ (2009) The NMDA agonist D-cycloserine facilitates fear memory consolidation in humans. Cereb Cortex 19:187–196.
Knight DC, Smith CN, Stein EA, Helmstetter FJ (1999) Functional MRI of human Pavlovian fear conditioning: patterns of activation as a function of learning. Neuroreport 10:3665–3670.
Knight DC, Cheng DT, Smith CN, Stein EA, Helmstetter FJ (2004a) Neural substrates mediating human delay and trace fear conditioning. J Neurosci 24:218–228.
Knight DC, Smith CN, Cheng DT, Stein EA, Helmstetter FJ (2004b) Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. Cogn Affect Behav Neurosci 4:317–325.
Knight DC, Nguyen HT, Bandettini PA (2005) The role of the human amygdala in the production of conditioned fear responses. Neuroimage 26:1193–1200.
Knight DC, Waters NS, Bandettini PA (2009) Neural substrates of explicit and implicit fear memory. Neuroimage 45:208–214.
Koo JW, Han JS, Kim JJ (2004) Selective neurotoxic lesions of basolateral and central nuclei of the amygdala produce differential effects on fear conditioning. J Neurosci 24:7654–7662.
Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.
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.
LeDoux JE (2000) Emotion circuits in the brain. Annu Rev Neurosci 23:155–184.
Mai JK, Paxinos G, Voss T (2008) Atlas of the human brain. New York: Elsevier.
Maren S (1998) Overtraining does not mitigate contextual fear conditioning deficits produced by neurotoxic lesions of the basolateral amygdala. J Neurosci 18:3088–3097.
Maren S (1999) Neurotoxic basolateral amygdala lesions impair learning and memory but not the performance of conditional fear in rats. J Neurosci 19:8696–8703.
Marschner A, Kalisch R, Vervliet B, Vansteenwegen D, Büchel C (2008) Dissociable roles for the hippocampus and the amygdala in human cued versus context fear conditioning. J Neurosci 28:9030–9036.
Mechias ML, Etkin A, Kalisch R (2010) A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. Neuroimage 49:1760–1768.
Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL (2007) Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. Biol Psychiatry 62:446–454.
Morris JS, Dolan RJ (2004) Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. Neuroimage 22:372–380.
Morris JS, Ohman A, Dolan RJ (1998) Conscious and unconscious emotional learning in the human amygdala. Nature 393:467–470.
Morris JS, Buchel C, Dolan RJ (2001) Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. Neuroimage 13:1044–1052.
Nagy Z, Weiskopf N, Alexander DC, Deichmann R (2007) A method for improving the performance of gradient systems for diffusion-weighted MRI. Magn Reson Med 58:763–768.
Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430.
Pape HC, Pare D (2010) Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. Physiol Rev 90:419–463.
Paré D, Quirk GJ, LeDoux JE (2004) New vistas on amygdala networks in conditioned fear. J Neurophysiol 92:1–9.
Pine DS, Fyer A, Grun J, Phelps EA, Szeszko PR, Koda V, Li W, Ardekani B, Maguire EA, Burgess N, Bilder RM (2001) Methods for developmental studies of fear conditioning circuitry. Biol Psychiatry 50:225–228.
Reijmers LG, Perkins BL, Matsuo N, Mayford M (2007) Localization of a stable neural correlate of associative memory. Science 317:1230–1233.
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.
Schurger A, Pereira F, Treisman A, Cohen JD (2010) Reproducibility distinguishes conscious from nonconscious neural representations. Science 327:97–99.
Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon CH, Kim SG, Tong F (2010) Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. J Neurosci 30:325–330.
Tabbert K, Stark R, Kirsch P, Vaitl D (2006) Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. Neuroimage 32:761–770.
Veit R, Flor H, Erb M, Hermann C, Lotze M, Grodd W, Birbaumer N (2002) Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. Neurosci Lett 328:233–236.
Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. Neuroimage 33:493–504.
Weiskopf N, Hutton C, Josephs O, Turner R, Deichmann R (2007a) Optimized EPI for fMRI studies of the orbitofrontal cortex: compensation of susceptibility-induced gradients in the readout direction. MAGMA 20:39–49.
Weiskopf N, Sitaram R, Josephs O, Veit R, Scharnowski F, Goebel R, Birbaumer N, Deichmann R, Mathiak K (2007b) Real-time functional magnetic resonance imaging: methods and applications. Magn Reson Imaging 25:989–1003.
Whalen PJ (1998) Fear, vigilance and ambiguity: initial neuroimaging studies of the human amygdala. Curr Dir Psychol Sci 7:177–188.
Zimmerman JM, Rabinak CA, McLachlan IG, Maren S (2007) The central nucleus of the amygdala is essential for acquiring and expressing conditional fear after overtraining. Learn Mem 14:634–644.