Annals of Medicine. 2010; 42: 127–134

ORIGINAL ARTICLE

A translational bridge between mouse and human models of learned safety

DANIELA D. POLLAK[1,2], MICHAEL T. ROGAN[1,3], TOBIAS EGNER[4], DAVID L. PEREZ[5,6], TED K. YANAGIHARA[3] & JOY HIRSCH[1,3,7,8]

[1]*Department of Neuroscience, Columbia University, New York, NY, USA,* [2]*Department of Physiology, Center for Physiology and Pharmacology, Medical University of Vienna, Austria,* [3]*Program for Imaging & Cognitive Sciences, PICS, Columbia University, New York, NY, USA,* [4]*Center for Cognitive Neuroscience, Department of Psychology & Neuroscience, Duke University, NC, USA,* [5]*Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA,* [6]*Department of Neurology, Massachusetts General Hospital, Boston, MA, USA,* [7]*Department of Radiology, Columbia University, New York, NY, USA, and* [8]*Department of Psychology, Columbia University, New York, NY, USA*

Abstract

Background. Learned safety is established by negatively correlating the occurrence of a neutral stimulus and a noxious stimulus, which renders the previously neutral stimulus a 'safety signal'. While the neurophysiological and molecular mechanisms have been characterized in mice, it is currently not known how the neural substrates involved compare between mice and people.

Methods. Here we attempt to adapt the original animal protocol to humans and use functional magnetic resonance imaging to examine neural responses to the conditioned stimulus in safety conditioned and fear conditioned subjects. Diffusion tensor imaging (DTI) was used in a parallel group of subjects as a first approach to delineate the underlying neural circuitry.

Results. Learned safety is associated with dampened amygdala and increased dorsolateral prefrontal cortex and caudate responses and paralleled by pupillary constriction. A neural connection between the amygdala and the dorsolateral prefrontal cortex is suggested by DTI.

Conclusion. We present a translational bridge between mouse and human models of learned safety in which cellular and molecular insights from animal experiments are extended to the human neural circuitry. This study provides an example of how animal experiments can be used to inform and target human studies, which in turn can corroborate results obtained in experimental animals.

Key words: *Amygdala, dorsolateral prefrontal cortex, functional magnetic resonance imaging (fMRI), learned safety, translational research*

Introduction

Fear conditioning results from a positive correlation (pairing) of a previously neutral conditioned stimulus (CS) and an aversive unconditioned stimulus (US). During safety conditioning, by contrast, a CS that is negatively correlated (explicitly unpaired) with an aversive US becomes a positive signal (predictor) for safety (1,2). Exposure to unpredictable aversive events in the absence of safety signals that can indicate to an individual when it is appropriate to relax and feel safe can lead to chronic stress and anxiety, eventually contributing to the development of psychopathologies (3). Classical Pavlovian fear conditioning has been most widely used to study the origins of fear and anxiety and the pathophysiology of affective disorders both in humans and in experimental animal models (4). Learned safety can be successfully induced in mice using an explicitly unpaired CS-US

Correspondence: Daniela D. Pollak, Department of Physiology, Center for Physiology and Pharmacology, Medical University of Vienna, Schwarzspanierstrasse 17, A-1090 Vienna, Austria. E-mail: daniela.pollak@meduniwien.ac.at

(Received 30 October 2009; accepted 23 December 2009)

ISSN 0785-3890 print/ISSN 1365-2060 online © 2010 Informa UK Ltd. (Informa Healthcare, Taylor & Francis AS)
DOI: 10.3109/07853890903583666

Key messages

- Learned safety based on a mouse model can be translated to humans and investigated using functional magnetic resonance and diffusion tensor imaging.
- Learned safety in humans is associated with dampened amygdala and increased dorsolateral prefrontal cortex neural activity.
- Animal experiments can be used to inform and target human studies, which in turn can corroborate results obtained in experimental animals.

Abbreviations

EPI	echo-planar imaging
CS	conditioned stimulus
dlPFC	dorsolateral prefrontal cortex
DTI	diffusion tensor imaging
fMRI	functional magnetic resonance imaging
hrf	hemodynamic response function
ROI	region of interest
SPGR	spoiled gradient recalled acquisition in the steady state
SPM	statistical parametric mapping
US	unconditioned stimulus

protocol and has been proposed as an animal model of a behavioral intervention of depression (1,2). In mice, learned safety induces cell biological changes characteristic of the effects of pharmacological antidepressants. Moreover, presentation of the safety signal leads to decreased neuronal activity in the amygdala and increased responses in the caudate (2). Safety learning established through differential conditioning procedures (5) and reversal of fear learning (6) has been studied in humans. However, it is currently not known whether the learned safety model based on the explicit unpairing procedure used in the mouse study to elicit an antidepressant-like effect (1,2) can be translated to humans and how the neural substrates involved compare between mice and people. Here we employed functional magnetic resonance imaging (fMRI) to investigate the neural correlates of learned safety in healthy human volunteers with special emphasis on the neural circuitry level translation of the model. The fMRI analysis was inspired by the mouse studies and directed towards activity profiles in the amygdala and the caudate, regions that we have previously shown to be implicated in learned safety in mice (2). In addition, since learned safety elicits an antidepressant-like response in mice, and disruption of the amygdala-prefrontal emotion regulation circuitry has been proposed as a core mechanism in the pathogenesis of depressive disorders in humans (7), the dorsolateral prefrontal cortex (dlPFC) was also examined as an a-priori region of interest (ROI). We hypothesized that neural responses to learned safety signals would be associated with reduced amygdala activity, paired with enhanced responses in the caudate and dlPFC.

Materials and methods

Participants

Twenty-five healthy volunteers (12 men, 13 women, mean age 28.9 ± 1.99 (SEM) years) gave written informed consent to participate in this study, in accordance with Columbia University's institutional guidelines. All participants had normal or corrected-to-normal vision and were screened by self-report in order to exclude any subjects reporting previous or current neurological or psychiatric conditions, or current psychotropic medication use.

Experimental paradigm and procedure

Unconditioned stimulus (US). The aversive unconditioned stimulus consisted of recordings of human screams presented at a realistic volume (98 dB), i.e. a level that would occur if a person were to actually scream in the subject's close vicinity.

A total of 133 recordings of male and female human screams, each with a uniform rise time (200 ms) and a duration of 1–3 sec, were generated and tested in a group of 15 naive subjects (not participating in the imaging study). In this pilot study the subjects were exposed to the scream recordings at a comparable volume used in the imaging experiment and after each scream were instructed to rate their personal experience of aversiveness elicited by the scream on a scale from 1 (not at all aversive) to 5 (extremely aversive). Average aversiveness scores were calculated based on the ratings given by all the subjects. Sixty screams with the highest average aversiveness ratings were selected and used to construct 20 unique US, each consisting of 3 individual scream recordings (a scream triplet) delivered at an inter-stimulus-interval that brought the total duration of the scream US to 12 seconds. Each triplet was constructed to have the same net aversiveness rating and the same net scream duration. Auditory stimuli were delivered via in-ear pneumatic tubes encased within foam earplugs. To further isolate the subject from scanner noise, the outer ears were completely covered by a 25mm thick heat-sensitive form-fitting foam rubber pad compressed against the sides of the subject's head.

Conditioned stimulus (CS). The CS consisted of an easily visible annulus made of two rings, one slightly brighter than the background, and one slightly darker than the background, centered at the fixation point (a black small gray circle on a blank dark gray background) and was always presented for 20 s. In order to relate to our mouse protocol, we designed two matched training protocols for humans: *safety conditioning* (10 US and 10 explicitly unpaired visual CS), and *fear conditioning* (10 CS, each with a co-terminating US), which were carried out in two different groups of subjects. The safety conditioning group consisted of six male and eight female subjects (mean age 29.4 ± 2.53 (SEM)). The fear conditioning group comprised six male and five female subjects (mean age 26.6 ± 1.06 (SEM)).

The training phases (total duration of 8.3 min) for the two groups were designed as follows. *Safety conditioning*: 10 scream US that were explicitly unpaired with 10 visual CS, such that screams never occurred during the presentation of the CS. The time between each CS and the proximal US ranged from 2 to 44 s (mean 16.63 s). *Fear Conditioning*: 10 scream US were paired with 10 visual CS with the last 12 s of each CS overlapping with a scream US. The time between the offset of a scream US and the onset of the proximal US ranged from 16 to 48 s (mean 33.78 s). Throughout the scanning session a fixation point was presented to the subjects via screen goggles, and room lights were set at low ambient light levels. The training phase was separated from the test phase by a structural scan (spoiled gradient recalled acquisition in the steady state (SPGR)) with a total duration of 12 min, during which the screen displaying the fixation point was replaced by a blank screen and subjects were instructed to rest. During the test phase (total duration of 2.3 min) in both groups, five CS and no US were presented (mean inter-CS interval 9.5 s).

In a pilot experiment, independent individuals not participating in the fMRI study were presented with the CS in the scanning environment in a manner comparable to the presentation of the stimulus in the actual study (i.e. size and contrast intensity of the visual symbol and time of stimulus presentation), and pupillary measurements were recorded. This pilot experiment was carried out to test whether the CS itself could affect pupillary size. However, it was found that presentation of our visual CS only induced initially a slight and transient dilation of the pupil, a typical response to novelty and thus was suitable to be used as conditioned stimulus. We controlled for non-associative effects of CS presentation by comparing the experimental groups, each of which viewed the same CS in the same presentation sequence in the test phase.

Stimuli for the fMRI studies were presented with Presentation software (Neurobehavioral Systems, http://nbs.neuro-bs.com) and displayed on VisuaStim XGA LCD screen goggles (Resonance Technology, Northridge, CA).

Physiological set-up and assessment (pupillary response)

The continuous record of pupil diameter was processed to identify and eliminate blink artifacts using amplitude thresholds, binned to the temporal resolution of the scan (2 s), and convolved with a canonical hemodynamic response function using Matlab codes adapted in our laboratory. Irrelevant drifts in the pupil diameter data over the course of the scan session were removed. The group analysis was performed on the mean pupil diameter across the first scream US and first test CS presentation evaluating the time interval from 0.5 s to 4 s of stimulus delivery, the same time window during which pupillary constriction in response to complex isoluminant visual stimuli had previously been demonstrated (8). To statistically evaluate the response stimulus presentation compared to the corresponding time period before the onset of the stimulus within one group, we carried out one-sample *t* tests using a hypothetical mean of 0. Group differences were evaluated using two-tailed independent sample Student *t* tests. A significance level of $P < 0.05$ was accepted as statistically significant.

Functional analysis

Image acquisition. Images were acquired with a GE 1.5 tesla MRI scanner. Functional data images were acquired along the AC-PC line with a T2*-weighted echo-planar imaging (EPI) sequence of 24 contiguous axial slices repetition time (TR) = 2000 ms, echo time (TE) = 40 ms, flip angle = 90°, field of view (FOV) = 190 × 190 mm) of 4.5 mm thickness and 3 mm in-plane resolution. Structural data were acquired with a high-resolution T1-weighted SPGR scan (TR = 19 ms, TE = 5 ms, flip angle = 20°, FOV = 220 × 220 mm), recording 124 slices at a slice thickness of 1.5 mm and an in-plane resolution of 0.86 × 0.86 mm.

Image preprocessing. All preprocessing and statistical analyses were carried out using Statistical Parametric Mapping 5 (SPM5) (http://www.fil.ion.ucl.ac.uk/spm/software/spm5). Functional images were slice-timing corrected and spatially realigned to the first volume of the first run. For each subject, the structural scan was co-registered to a mean image of the realigned functional scans. The co-registered structural image was then used to calculate transformation parameters for

normalizing the functional images to the Montreal Neurological Institute (MNI) template brain. The normalized functional images were spatially smoothed with a Gaussian kernel of 8 mm^3. The first five scans of each run were discarded prior to further analysis. Vectors of stimulus onsets were created for each of the experimental conditions. These vectors were then convolved with SPM5's canonical hemodynamic response function (hrf) and employed as regressors to model the blood oxygen level dependency (BOLD) responses associated with the task. A 128-s temporal high-pass filter was applied to the data to remove low-frequency artifacts. Temporal autocorrelation in the time series data was estimated using restricted maximum likelihood estimates of variance components using a first-order autoregressive model (AR-1), and the resulting nonsphericity was used to form maximum likelihood estimates of the activations.

Image analysis. To test the hypotheses with respect to regionally specific group effects, the estimates were compared by using two contrasts (increased or decreased BOLD signal in the presence of the CS in safety versus fear conditioned subjects) using a first-order time modulation model in SPM5, to account for session-specific adaptation in brain activity as a result of repeated CS presentation over time. The resulting set of voxel values was tresholded at $P < 0.001$ uncorrected with a spatial extent of five contiguous voxels, in anatomical a-priori ROIs of the amygdala, caudate, and dlPFC. Multiple testing has been corrected for by controlling the family-wise error rate (FWE). To correct for the expected within-session effects over time (i.e. loss of power of the CS to elicit neural responses due to the repeated exposure during the test phase), we integrated a first-order time modulation into our model. The comparisons were made between safety conditioned and fear conditioned subjects. Anatomical ROIs were defined using the Wake Forest University PickAtlas toolbox (http://fmri.wfubmc.edu/cms/software#PickAtlas).

Tractography

Image acquisition. Diffusion tensor images (DTI) were acquired on a 1.5-T GE twin-speed scanner using an 8-channel sense head coil with a single-shot sequence of 55 unique diffusion directions at a b-value = 900 with TE = 7.8 ms and TR = 17000 ms. A single no-diffusion volume (b-value = 0) was acquired and used as a reference to correct for eddy currents and head motion. Isotropic (2.5 mm^3 voxels) diffusion-weighted data were acquired for all subjects. Array size was 128 × 128 in a FOV of 3 × 32 mm. Altogether 58 slices were acquired, and the total scan time was 16 minutes and 32 seconds.

Image analysis. DTI analysis was completed using the FMRIB's software library diffusion toolbox (http://www.fmrib.ox.ac.uk/fsl) (9). A probability of connectivity map was generated for regions of interest as described by Behrens et al. (10). Seed and target masks were generated using the ROI definition function in Marsbar (http://marsbar.sourceforge.net), by taking the co-ordinates of peak fMRI activation for amygdala (i.e. −28, 0, −16) and dlPFC (i.e. −30, 52, 32) and generating a sphere of a radius of 4 mm. Briefly, in native diffusion space, the principal diffusion direction (PDD) of non-isotropic water movement was modeled as a tensor for each voxel in the brain. Complex fiber structure (i.e. crossing or diverging fibers) increases the uncertainty of the PDD estimate. Bayesian statistics were used to generate probability density functions (pdfs) of PDD uncertainty allowing for the detection of non-dominant fiber pathways (11). From these pdfs, 5000 tract-following samples were taken with a maximum curvature threshold of ± 80 degrees. The subjects who participated in the DTI study did not participate in the functional study of learned safety, and, therefore, the normalized clusters of activity served as predictors of the neural circuitry. Connectivity maps were then inspected for each subject. As hypothesized, the largest number of streamlines followed a direct path from the amygdala seed mask to the waypoint (dlPFC) mask in the majority (21 out of 28) of subjects. These thresholded paths were then binarized and added across all subjects, generating a group representation of individual pathways. In Figure 2E, intensity values at each voxel for this group image correspond to the number of subjects with a streamline passing through that voxel. The group image does not correspond to a map of probabilistic connectivity from the seed to the waypoints mask as presented for individual subjects, but instead represents the importance of each voxel to this pathway with respect to all subjects. Given the current investigational status of accepted methods for performing statistical analyses or graphical representation of group tractography data, this method aims to conservatively quantify our results at the group level.

Results

Aiming to provide a translational bridge between the animal and human conditioning procedure, we developed a conditioning protocol along the lines of the mouse model. We used recordings of human screams as US and a neutral visual symbol as CS. CS and US

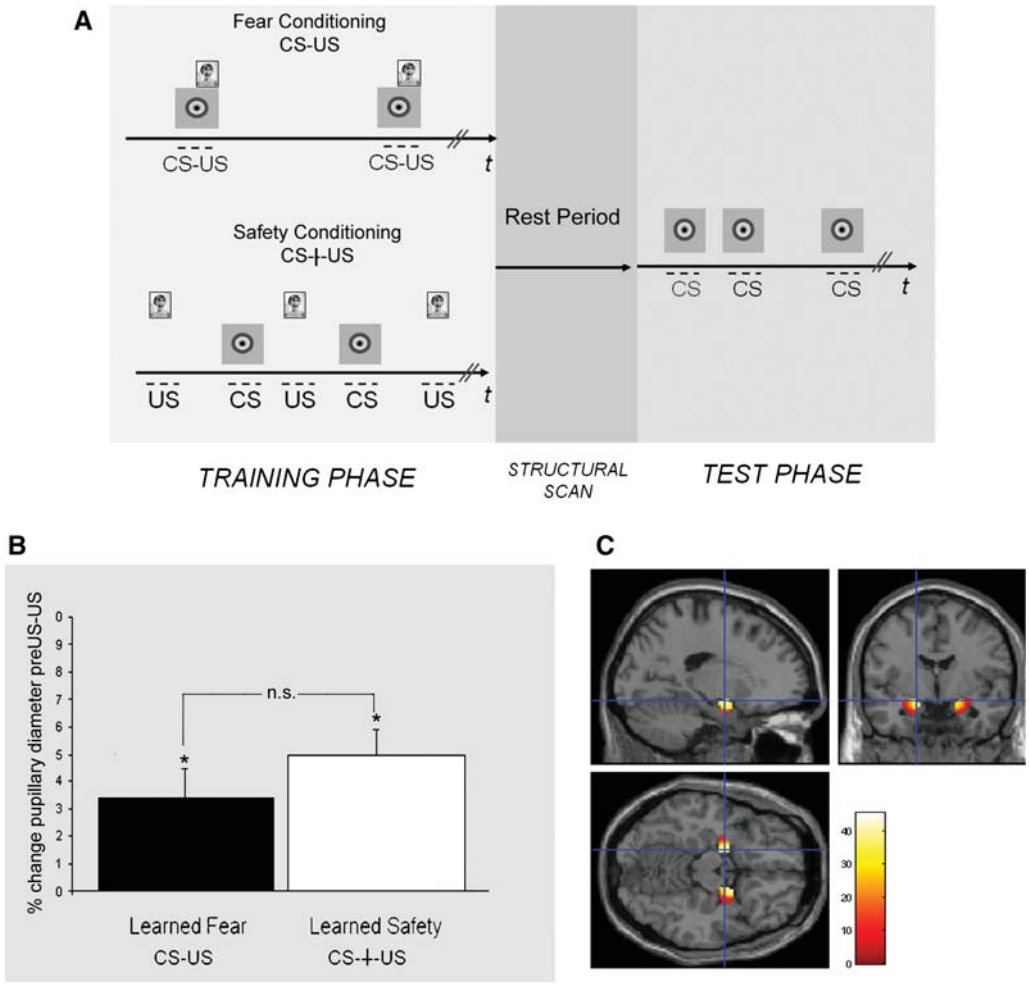

Figure 1. Experimental paradigm and pupillary and amygdala responses to the aversive stimulus. A: Training and test phase: The training phase (left) consisted of several explicitly unpaired (bottom row) or paired (top row) presentations of the CS and the US and was followed by a period of rest (middle) during which the structural MRI images were acquired. The test phase (right) consisted of five presentations of the CS alone. B: Presentation of the screams during the training phase led to significant pupillary dilation in safety and fear conditioning groups ($^{*}P < 0.05$) which was comparable among the two groups ($P > 0.05$). C: Activity in the left and right amygdala in both groups (average effect of condition from both groups) (Montreal Neurological Institute (MNI) co-ordinates: −18, −4, −14, $k = 226$, t score = 5.30, $P < 0.05$, FWE corrected; and 20, 0, −14, $k = 244$, t score = 5.15, $P < 0.05$, FWE corrected).

were explicitly temporally unpaired in the learned safety group and paired for learned fear during the training phase (Figure 1A). We first tested the emotional response elicited by the scream US by evaluating the pupillary diameter and amygdala activity in response to the US during the training phase in both groups. We found that presentation of the US induced comparable pupillary dilation in both groups, validating the aversiveness of the scream US and confirming base-line similarity among the paired and the unpaired training groups (Figure 1B). Both groups showed significant amygdala activation in response to the scream US (Figure 1C). In addition to acquiring functional imaging data, we used measurements of pupillary diameter as a proxy for emotional responses elicited by the CS. Pupillary responses are one of the most widely used and validated psychophysiological measures in applied and basic research of emotional processes, and ample evidence from the literature indicates that the pupil response is a suitable indicator for human Pavlovian conditioning (12). In the test phase we measured the pupillary diameter in the presence of the first test CS in order to provide an independent biophysical measure of the emotional value associated to the conditioned stimulus (8). In the unpaired (i.e. learned safety trained) group we observed pupillary constriction, while subjects of the paired training group (i.e. fear conditioning) showed pupillary dilation in response to the CS (see Figure 2A).

We next analyzed the neural response to the CS presentation during the test phase in our a-priori ROIs. In the amygdala, a cluster of voxels was significantly less activated in the unpaired than in the paired group ($P < 0.05$, FWE corrected) (Figure 2B).

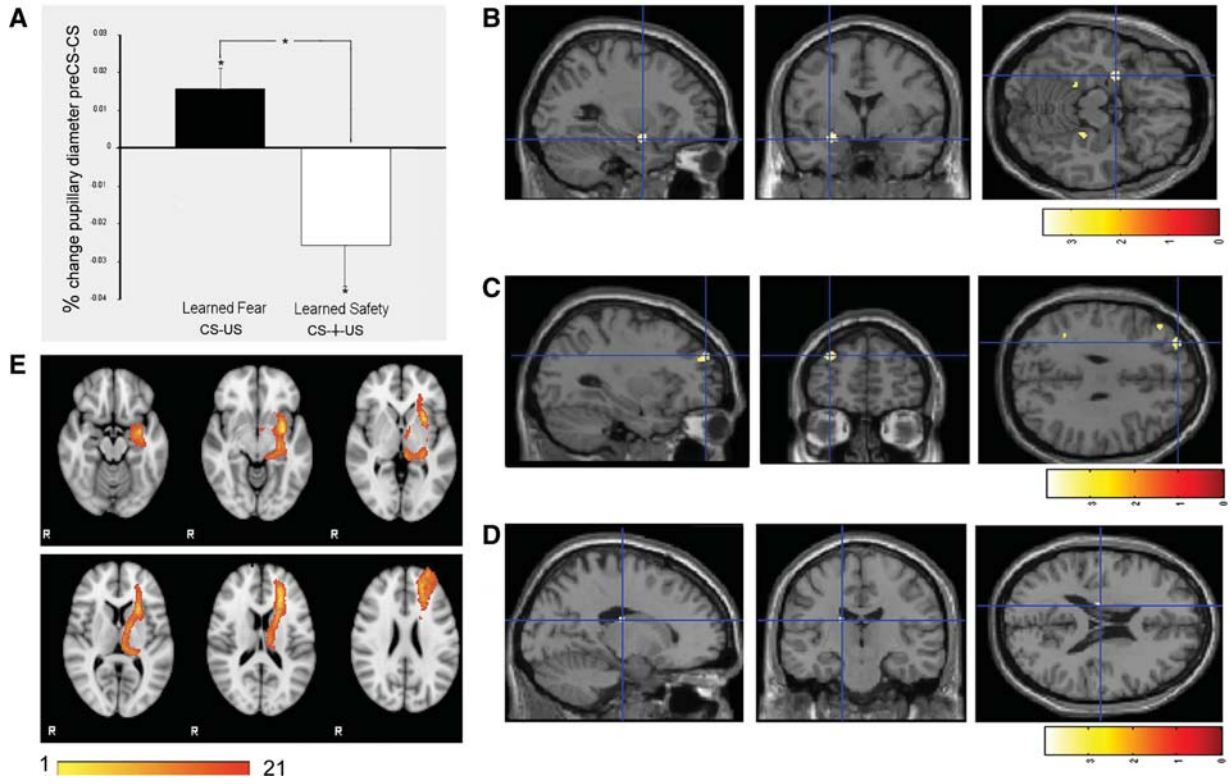

Figure 2. Pupillary and neural responses to the conditioned stimulus in the test phase. A: Percentage change in pupillary diameter during the presentation of the first CS ($^*P < 0.05$) demonstrates pupillary constriction in the learned safety group and pupillary dilation in the learned fear group. B: A cluster of differential activation in the left amygdala between safety and fear trained subjects in response to the CS is shown on a standard brain (Montreal Neurological Institute (MNI) co-ordinates: −28, 0, −16, $k = 8$, t score = 3.56, $P < 0.05$, FWE corrected). For display purposes, activity clusters are shown at $P < 0.005$, uncorrected in B and C. C: Differential activation in the left dlPFC between safety and fear trained groups in response to the CS (MNI: −30, 52, 32, $k = 9$, t score = 3.97, $P < 0.001$ uncorrected). D. Differential activation in the left caudate between safety and fear trained groups in response to the CS (MNI: −16, −18, 20, $k = 12$, t score = 2.11, $P = 0.021$ uncorrected). E. Group-based probabilistic tractography map between the amygdala and the dlPFC.

Conversely, a cluster of voxels in the left dlPFC (corresponding to Brodman area (BA) 9) showed a definite trend for greater activation in the paired than in the unpaired group ($P < 0.001$, uncorrected) (Figure 2C). When we employed a more liberal threshold ($P < 0.05$, uncorrected), we also observed a trend for a cluster of voxels in the left caudate with greater activity in the unpaired than in the paired group (see Figure 2D). Exploratory whole-brain analysis revealed no other differentially activated regions at a threshold corrected for multiple comparisons (for additional regions that displayed differential activity at $P < 0.001$, uncorrected, see Supplemental Table I online).

These data open up the possibility to suggest that as a result of unpaired training presentation of the CS may lead to dlPFC-mediated suppression of amygdalar activity in conditioned subjects, closely resembling data from previous studies of instructed suppression of negative affect (13). Although the notion that dlPFC regions involved in cognitive control are able to modulate amygdalar responses is widely held (13), animal studies have documented only sparse direct connections between the dlPFC and the amygdala (14). This motivated an investigation of connectivity using diffusion tensor imaging (DTI) and probabilistic tractography between the dlPFC region that exhibited enhanced activity, and the region of the amygdala that displayed suppressed responses in the unpaired group. DTI images were acquired from 28 subjects, converted to normalized space, and probabilistic tractography maps were determined for the connection between these ROIs defined by the centroids of activity observed in the fMRI study. The number of streamlines (waytotals) detected was recorded for each subject (Supplemental Table II online). We observed at least one streamline between the amygdala and dlPFC for 21 of the 28 subjects. Of those, more than 10 streamlines were detected in 9 out of 28 subjects, suggestive of a potential direct anatomical connection between these regions (Figure 2E). Because probabilistic tractography results depend significantly on the volume of the seed mask, and our amygdala seed consisted of

only 100 voxels in normalized space, we believe that this result may under-represent the connections.

Discussion

As with rodents, safety has been far less studied in the human than fear. When safety responses and safety learning have been examined, it has been in the context of discriminative learning, a more complex task than the one we used in our mouse studies of safety learning (6,15,16). In a first attempt to translate our findings about the physiological and molecular underpinnings of learned safety in mice to people, we designed a human protocol along the lines of the animal paradigm although there are some important differences between the previously used mouse protocol and the currently used human protocol. Firstly, the conditioning protocol in the mouse consisted of three sessions performed on three different days, while in the human procedure a single training session was performed followed by the test session on the same day. Secondly, the antidepressant-like effect of learned safety observed in mice (1) was not tested in the present human study. Instead we focused our analysis on the neural circuitry in humans and used fMRI as a tool to shed light on associated neural correlates. We targeted our analyses based upon the predictions from the mouse model, assuming that the unpaired CS-US training results in a process of modulation of the emotional assessment processes in the amygdala, possibly mediated by the dlPFC. In fact we found that presentation of the CS following unpaired training (i.e. learned safety) was associated with dampened amygdalar activity, and heightened dlPFC responses. These results suggest that indeed, following unpaired training, the conditioned stimulus may acquire the potential to signal safety and control emotional responsiveness in the amygdala through increased top-down regulation via the dlPFC. This hypothesis is also in line with a wealth of literature proposing an executive control function of the dlPFC over amygdalar emotional reactivity and the emotional regulation (13).

Interestingly, increased and sustained amygdala activity (17,18) under base-line conditions and following emotion induction (19,20) has been repeatedly reported in depressed patients. Moreover, deficient dlPFC activity has been observed in depression (18,21–23) suggesting that sustained emotional reactivity and negativity bias in depressed patients might result indirectly from impaired dlPFC function. However, the dlPFC-amygdala circuitry can also be involved in other processes, such as attention, language, and memory (24,25). Furthermore, the possibility exists that there could be an additional indirect modulation of amygdala activity that could be mediated, for instance, via the ventromedial prefrontal cortex, which has strong connection both to the dlPFC and the amygdala, and has been shown to modulate amygdalar activity during fear reversal and safety learning (6). Given the role of the caudate in reward systems, increased activity in this region is consistent with the safety signal having positive emotional value.

Although previous studies have used skin conductance response (SCR) to assess emotional learning (6,15) we evaluated pupillary diameters as independent psychophysical measure since is SCR is dynamically sensitive to increases of stress and fear, but less sensitive to decreases of stress and fear. Stress and anxiety drive sweat release in the skin, which increases SCR, but reduction of SCR depends essentially upon diffusion and evaporation of sweat, and is therefore not under tight neural control. We observed pupillary constriction in response to the CS in the unpaired group which may reflect reduced fear in the presence of the safety signal since dilation of the pupils is thought to reflect anxiety states (12) whereas pupillary constriction is known to be an effect of anxiolytic drugs, such as benzodiazepines (26) and morphines (27).

Taken together, we conclude that rodent learned safety protocols can be translated to human protocols and models, suggesting homology of the dlPFC-amygdala circuitry potentially leading to a dlPFC-mediated suppression of fear responses in the amygdala (28). Although a potential antidepressant effect of learned safety in humans has not been assessed here, results from the present analysis suggest future experiments in this direction. This study thus presents a novel and useful type of translational research in the field of neuropsychiatric disorders in which behavioral, cellular, and molecular studies in experimental animals are combined with whole-brain systems-level information of neural circuitry by functional neuroimaging.

Limitations

Some conceptual limitations need to be considered for the interpretation of our findings. These constraints arise from our aim to develop a human safety learning paradigm that would allow us to stay as close as possible to the original mouse protocol. We therefore established for the human, as we did for the mouse study (1), a learned safety paradigm and contrasted this to learned fear in a between-group rather than a within-group (differential conditioning) design. As a consequence, we cannot refer

a within-subject control condition when describing the learned safety CS-induced neural activity. Moreover, this might have reduced the statistical power of the analysis.

Acknowledgements

Daniela D. Pollak was supported by the Austrian Academy of Science (Max-Kade-Fellowship) and the Austrian Science Fund (Erwin-Schroedinger-Fellowship). Tobias Egner was supported by a Columbia University fMRI Research Fellowship. Ted Yanagihara was supported by the Medical Scientist Training Program, Columbia University Medical School. Imaging funds were provided by education grants to Joy Hirsch, Director, fMRI Research Center, Columbia University (now PICS, Program for Imaging & Cognitive Sciences) for translational and applied studies. The extremely skillful and dedicated expertise of Stephen Dashnaw in carrying out the imaging experiments is highly appreciated.

Declaration of interest: The authors report no biomedical financial interests or other conflicts of interest.

References

1. Pollak DD, Monje FJ, Zuckerman L, Denny CA, Drew MR, Kandel ER. An animal model of a behavioral intervention for depression. Neuron. 2008;60:149–61.
2. Rogan MT, Leon KS, Perez DL, Kandel ER. Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. Neuron. 2005;46:309–20.
3. Seligman ME. Chronic fear produced by unpredictable electric shock. J Comp Physiol Psychol. 1968;66:402–11.
4. Mineka S, Oehlberg K. The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. Acta Psychologica. 2008;127:567–80.
5. Monk CS, Grillon C, Baas JM, McClure EB, Nelson EE, Zarahn E, et al. A neuroimaging method for the study of threat in adolescents. Dev Psychobiol. 2003;43:359–66.
6. Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA. From fear to safety and back: reversal of fear in the human brain. J Neurosci. 2008;28:11517–25.
7. Siegle GJ, Thompson W, Carter CS, Steinhauer SR, Thase ME. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: related and independent features. Biol Psychiatry. 2007;61:198–209.
8. Conway CA, Jones BC, DeBruine LM, Little AC, Sahraie A. Transient pupil constrictions to faces are sensitive to orientation and species. J Vis. 2008;8:17.1–1.
9. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004;23 Suppl 1:S208–19.
10. Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med. 2003;50:1077–88.
11. Behrens TE, Berg HJ, Jbabdi S, Rushworth MF, Woolrich MW. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage. 2007;34:144–55.
12. Reinhard G, Lachnit H, Konig S. Tracking stimulus processing in Pavlovian pupillary conditioning. Psychophysiology. 2006;43:73–83.
13. Ochsner KN, Gross JJ. The cognitive control of emotion. Trends Cogn Sci. 2005;9:242–9.
14. Barbas H. Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. Neurosci Biobehav Rev. 1995;19:499–510.
15. Grillon C, Falls WA, Ameli R, Davis M. Safety signals and human anxiety: a fear-potentiated startle study. Anxiety. 1994;1:13–21.
16. Lissek S, Rabin SJ, McDowell DJ, Dvir S, Bradford DE, Geraci M, et al. Impaired discriminative fear-conditioning resulting from elevated fear responding to learned safety cues among individuals with panic disorder. Behav Res Ther. 2009;47:111–8.
17. Abercrombie HC, Schaefer SM, Larson CL, Oakes TR, Lindgren KA, Holden JE, et al. Metabolic rate in the right amygdala predicts negative affect in depressed patients. Neuroreport. 1998;9:3301–7.
18. Drevets WC. Prefrontal cortical-amygdalar metabolism in major depression. Ann N Y Acad Sci. 1999;877:614–37.
19. Sheline YI, Barch DM, Donnelly JM, Ollinger JM, Snyder AZ, Mintun MA. Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study. Biol Psychiatry. 2001;50:651–8.
20. Siegle GJ, Steinhauer SR, Thase ME, Stenger VA, Carter CS. Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. Biol Psychiatry. 2002;51:693–707.
21. Davidson RJ. Affective style, psychopathology, and resilience: brain mechanisms and plasticity. Am Psychol. 2000;55:1196–214.
22. Harvey PO, Fossati P, Pochon JB, Levy R, Lebastard G, Lehericy S, et al. Cognitive control and brain resources in major depression: an fMRI study using the n-back task. Neuroimage. 2005;26:860–9.
23. Mayberg HS, Liotti M, Brannan SK, McGinnis S, Mahurin RK, Jerabek PA, et al. Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. Am J Psychiatry. 1999;156:675–82.
24. Drevets WC. Functional anatomical abnormalities in limbic and prefrontal cortical structures in major depression. Prog Brain Res. 2000;126:413–31.
25. Duncan J, Owen AM. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci. 2000;23:475–83.
26. Hou RH, Scaife J, Freeman C, Langley RW, Szabadi E, Bradshaw CM. Relationship between sedation and pupillary function: comparison of diazepam and diphenhydramine. Br J Clin Pharmacol. 2006;61:752–60.
27. Murray RB, Adler MW, Korczyn AD. The pupillary effects of opioids. Life Sci. 1983;33:495–509.
28. Etkin A, Egner T, Peraza DM, Kandel ER, Hirsch J. Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. Neuron. 2006;51:871–82.

Supplemental Table I. Regions with greater activity (at $P < 0.001$, uncorrected) in the safety than in the fear conditioned group (whole brain analysis).

	Co-ordinates				
Region of activation	*x*	*y*	*z*	*k*	*t* score
Supramarginal gyrus	−44	−28	46	44	4.56
Precuneus	16	−74	46	53	4.3
Subgyral	46	−36	−12	23	4.28
Superior frontal gyrus	−30	52	30	52	4.12
Inferior parietal lobule	60	−48	48	67	4.09
Middle frontal gyrus	−46	36	30	16	3.93
Cerebellum	2	−52	−32	35	3.92
Precuneus	−16	−70	42	30	3.89
Inferior parietal	34	−52	40	60	3.82
Inferior occipital gyrus	34	−92	−4	12	3.72
Frontal superior medial	6	34	56	11	3.62
Inferior frontal gyrus	34	36	16	11	3.61
Middle frontal gyrus	34	36	26	7	3.54
Subgyral	−26	−48	36	8	3.42

Supplemental Table II. Number of streamlines (waytotals) for each subject.

Subject	Waytotal
1	0
2	0
3	6
4	1
5	0
6	367
7	0
8	3
9	0
10	1
11	17
12	12
13	4
14	8
15	4
16	8
17	24
18	2
19	0
20	45
21	4
22	14
23	2
24	167
25	1
26	68
27	1
28	114