

# If it bleeds, it leads: separating threat from mere negativity

Kestutis Kveraga,<sup>1,2</sup> Jasmine Boshyan,<sup>1,3</sup> Reginald B. Adams Jr.,<sup>4</sup> Jasmine Mote,<sup>5</sup> Nicole Betz,<sup>6</sup> Noreen Ward,<sup>1</sup> Nouchine Hadjikhani,<sup>1,2</sup> Moshe Bar,<sup>1,2,7,8</sup> and Lisa F. Barrett<sup>1,6,7</sup>

<sup>1</sup>Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA 02129, USA, <sup>2</sup>Department of Radiology, Harvard Medical School, Boston, MA, 02115 USA, <sup>3</sup>Department of Psychology, Brandeis University, 415 South Street, Waltham MA 02453-2728, USA, <sup>4</sup>Department of Psychology, The Pennsylvania State University, University Park, PA, 16802-3103 USA, <sup>5</sup>Department of Psychology, 3210 Tolman Hall, University of California, Berkeley, Berkeley, 94720 CA, USA, <sup>6</sup>Department of Psychology, 125 Nightingale Hall, Northeastern University, Boston, MA, 02115 USA, <sup>7</sup>Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA, 02129 USA, and <sup>8</sup>Gonda Multidisciplinary Brain Research Center, Building 901, Bar-Ilan University, Ramat Gan, 5290002 Israel

**Most theories of emotion hold that negative stimuli are threatening and aversive. Yet in everyday experiences some negative sights (e.g. car wrecks) attract curiosity, whereas others repel (e.g. a weapon pointed in our face). To examine the diversity in negative stimuli, we employed four classes of visual images (Direct Threat, Indirect Threat, Merely Negative and Neutral) in a set of behavioral and functional magnetic resonance imaging studies. Participants reliably discriminated between the images, evaluating Direct Threat stimuli most quickly, and Merely Negative images most slowly. Threat images evoked greater and earlier blood oxygen level-dependent (BOLD) activations in the amygdala and periaqueductal gray, structures implicated in representing and responding to the motivational salience of stimuli. Conversely, the Merely Negative images evoked larger BOLD signal in the parahippocampal, retrosplenial, and medial prefrontal cortices, regions which have been implicated in contextual association processing. Ventrolateral as well as medial and lateral orbitofrontal cortices were activated by both threatening and Merely Negative images. In conclusion, negative visual stimuli can repel or attract scrutiny depending on their current threat potential, which is assessed by dynamic shifts in large-scale brain network activity.**

**Keywords:** affect; emotion; scenes; fMRI; valence; arousal

## INTRODUCTION

Negative stimuli draw our attention. Rapid recognition of threat promotes survival (LeDoux, 2012) and selective pressures faced by early hominids did not favor those who ignored the signs of a nearby predator or failed to learn from the circumstances of a conspecific's demise. Newsmen have exploited this adaptive feature of our brains for their gain, as captured by media maxims such as, 'if it bleeds, it leads' and 'if it burns, it earns'. But do our brains react to all bleeding and burning in the same way?

Previous studies have shown that stimuli are automatically evaluated in terms of their affective valence: that is, whether a stimulus is pleasant, and thus worth approaching, or unpleasant, and best avoided (Duckworth *et al.*, 2002; Barrett, 2006). This determination is the earliest and most critical dimension of stimulus meaning (Ellsworth and Scherer, 2003; Barrett and Bliss-Moreau, 2009). Many studies have shown that affective experiences, words and even facial expressions can be represented along negative-to-positive gradient of stimulus valence along one axis, and arousal (also called engagement or activation) along the other (Russell, 1980; Watson and Tellegen, 1985; Larsen and Diener, 1992; Barrett and Russell, 1999). A recent meta-analysis of 397 functional magnetic resonance imaging (fMRI) studies (including over 6800 participants) demonstrates, however, that there is no pattern of voxels that consistently and specifically represents negative stimuli as a unified class (Lindquist, Saptute, Wager, Weber, Barrett, unpublished data). Although individuals experience a range of stimuli as unpleasant, the brain does not necessarily represent all negative stimuli in the same way. Everyday experiences support this observation. Although

few of us feel compelled to approach a dangerous animal or someone holding a gun, widespread rubbernecking at accident scenes and other sights of misfortune suggests that people can be tugged by the morbid, irresistible impulse to examine negative stimuli of a certain type. Imminent threat and accident scenes are both unpleasant situations, but our response to each can be quite different.

In this study, we examined the hypothesis that threatening and non-threatening negative stimuli would be treated differently by the brain. Specifically, we predicted that the potential for personal harm would be assessed quickly and automatically as an integral part of stimulus meaning. Images representing situations of high personal harm would generate rapid defensive action (LeDoux, 2012)—a 'fight-or-flight' response, with activations in brain regions that regulate such defensive reactions, such as the amygdala and periaqueductal gray (PAG) (Bandler and Carrive, 1988; Zhang *et al.*, 1990; Bandler and Shipley, 1994; Mobbs *et al.*, 2007). This would be particularly true because high personal harm images would be construed as motivationally salient, and motivational salience is important to modulating amygdala response (Cunningham and Brosch, 2012). In contrast, images representing unpleasant situations that are lower in the likelihood of personal harm would be associated with increased activation within the latter regions as well as engage the association-processing regions (the parahippocampal, retrosplenial and medial prefrontal cortices; Bar and Aminoff, 2003; Davachi, 2006; Bar *et al.*, 2008; Peters *et al.*, 2009; Kveraga *et al.*, 2011) to a greater extent. The brain regions involved in representing and regulating unpleasant subjective feelings [orbitofrontal cortex (OFC), ventromedial prefrontal cortex (vmPFC) and lateral PFC; Kringelbach and Rolls, 2004; Ochsner and Gross, 2005; Touroutoglou *et al.*, 2012; Wilson-Mendenhall *et al.*, 2013] should be engaged by all evocative negative images (threatening and Merely Negative) to assess and resolve the presence or absence of threat and its spatial and temporal characteristics. We tested these

Received 8 May 2013; Revised 4 January 2014; Accepted 10 January 2014  
Advance Access publication 3 February 2014

Correspondence should be addressed to Kestutis Kveraga, Martinos Center for Biomedical Imaging, Department of Radiology, Room 2301, Massachusetts General Hospital, Charlestown, MA 02129, USA. E-mail: keastas@nmr.mgh.harvard.edu

hypotheses in two behavioral experiments and one neuroimaging experiment.

**STUDY 1 (BEHAVIORAL)**

**Methods**

**Participants**

Twenty-seven participants from Massachusetts General Hospital (MGH) and the surrounding community took part in the first behavioral experiment and were compensated with \$20 for their participation. Our participants were of varying age, ethnicity and educational backgrounds: the mean age of the participants was 29.8 years (s.d. 8.6, range 22–56). The participants had a mean of 4.4 years (range 0–11) of post-secondary education, and all had normal or corrected-to-normal vision, as verified by pretests using the Ishihara color plates and a Snellen eye chart. All study procedures were approved by the Institutional Review Board of MGH and Partners Healthcare, protocol #2009P001227.

**Stimuli**

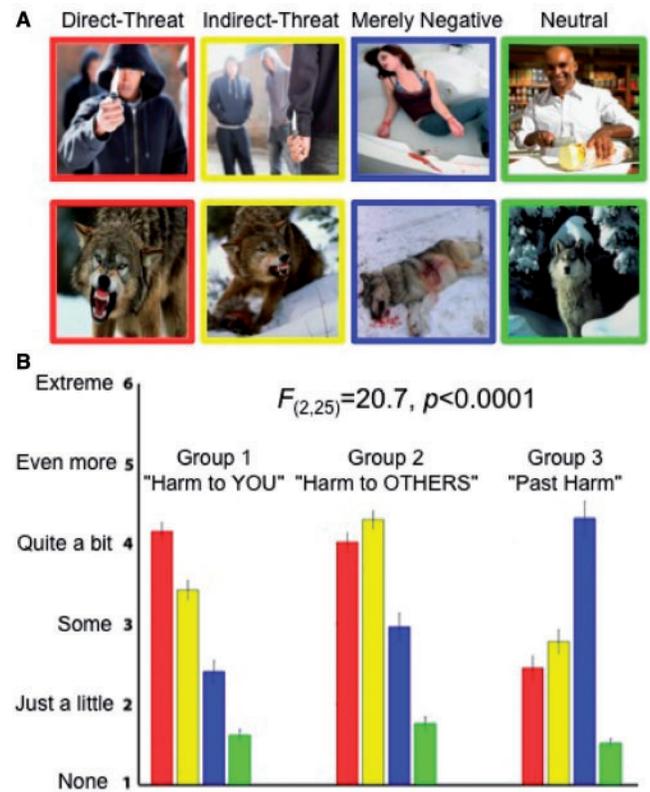
We collected a set of 509 color photo images depicting negative and neutral situations from Internet searches. The stimulus size was standardized to 300 × 300 pixels. Prior to conducting the study, we classified the images into four categories based on our initial impressions: (i) negative, Direct Threat (to the observer); (ii) negative, Indirect Threat (predominantly to others); (iii) Merely Negative, a possible past threat, but not a threat any longer; and (iv) Neutral, not a threat. The images in these four scene categories did not differ in terms of average luminance or spatial frequency: two-dimensional power spectral density estimates using the Welch’s method, implemented in the MATLAB (The Mathworks Inc., Natick, MA) image processing toolbox, revealed no significant differences among the stimulus conditions [one-way analysis of variance (ANOVA)  $F_{(3,396)} = 1.9$ ;  $P = 0.12$ ; Supplementary Figure S1].

**Design and Procedure**

Each group of participants only responded to one of the questions on a scale of 1–6 (ranging from ‘none’ to ‘extreme’). Subjects were not made aware of the other questions, nor were they aware during the debriefing that the stimuli had been assigned a priori into the four categories. Stimuli were presented in a random order, and one of the following three questions was presented below each picture: (i) ‘How much harm might you be about to suffer if this was your view in this scene?’ (Group 1), (ii) ‘How much harm might someone else (not you) be about to suffer in this scene?’ (Group 2), (iii) ‘How much harm might someone else (not you) have already suffered in this scene?’ (Group 3). Participants could view each stimulus as long as they wanted and enter their rating using keys 1–6 on the keyboard. Entering the rating replaced the current stimulus with the next one. The complete instructions and the rating scale are shown in Supplementary Figure S2.

**Results**

Our results revealed that participants were indeed keenly sensitive to the threat potential depicted in the images. Each group classified into the three types (Direct Threat, Indirect Threat and Merely Negative) of affective images in a highly distinct and consistent pattern [mixed-effects ANOVA, question group × scene condition interaction  $F_{(2,6)} = 20.7$ ,  $P < 0.00001$ ], according to the question they were asked about the stimuli, assigning the highest rating to the relevant a priori stimulus condition (Figure 1A). The repeated-measures ANOVAs testing differences in ratings within the groups revealed that each group significantly distinguished the affective (Direct Threat, Indirect Threat and Merely Negative) and Neutral images [Group 1



**Fig. 1** Rating threat in negative and neutral contexts. (A) Examples of stimuli images. The upper row, from left to right, shows examples of images with inanimate potential threat objects in the Direct Threat, Indirect Threat, Merely Negative and Neutral conditions. The lower panel shows examples of the same conditions with dangerous animals. (B) Results of the scene ratings. The three panels show the results from the three rating groups. In the first rating group (left panel), we asked participants to rate all images in terms of potential harm to them personally, designed to discriminate the Direct Threat images (red bars) from the rest. In the second rating group (middle panel) all images were rated in terms of potential harm to someone else, designed to identify the Indirect Threat images (yellow bars). In third rating group (right panel), we asked participants to evaluate the images in terms of past harm, but no present harm, which was designed to discriminate merely negative, but not presently threatening, images (blue bars). All three ratings also implicitly identified Neutral stimuli (green bars), which received the lowest ratings in all three experiments. The images in the different a priori conditions were presented in a randomized mixed sequence and the participants were unaware that the stimuli had been assigned to different conditions. The error bars indicate standard error of the mean. The wording of the questions and the ratings scale used in the three rating groups is shown in Supplementary Figure 2.

$F_{(3,24)} = 26.2$ ,  $P < 0.000002$ ; Group 2  $F_{3,24} = 18.2$ ,  $P < 0.00002$ ; Group 3  $F_{(3,24)} = 48.3$ ,  $P < 0.0000000003$ ]; as well as images within the affective scene categories themselves [Group 1  $F_{(2,16)} = 20.3$ ,  $P < 0.0000002$ ; Group 2  $F_{2,16} = 7.1$ ,  $P < 0.006$ ; Group 3  $F_{(2,16)} = 31.1$ ,  $P < 0.0000003$ ]. The control condition (Neutral images) was rated implicitly and received the lowest threat rating on all three questions [paired-samples  $t$ -test  $t_{(26)} = 16.5$ ,  $P < 0.000001$ ]. *Post hoc* tests revealed that the differences in ratings between each of the conditions were significant, with the only exception in Group 2 (the ‘threat to someone else’ rating), in which the participants did not significantly discount the threat to another person in Direct Threat images (Figure 1A, middle panel). As we aimed to include similar stimuli in all conditions (Figure 1B), our results show that the affective context of a scene plays a critical role in assessing whether the stimulus is deemed a threat. In addition, to ensure that the Direct Threat images were not perceived as more negative or more arousing than the Indirect Threat or Merely Negative images, we performed a second study in which a new cohort of participants rated the valence and arousal of the images (see Supplementary Material).

## STUDY 2 (fMRI)

### Methods

#### Participants

Twenty-one participants (seven males) from the MGH and surrounding communities participated in this fMRI study and were compensated with \$50 for their participation. None of the participants had been in any of the previous behavioral experiments in this study. The age of the participants ranged from 20 to 64 (mean = 30.3, s.d. = 9.9). One participant was excluded from functional data analyses because the structural scans revealed substantial brain atrophy. Thirteen participants were White, five were Asian, one was Black, and two were of mixed ethnic background. All had normal or corrected-to-normal visual acuity, normal color vision, and were right-handed. Thirteen participants spoke English as a native language, and the remaining eight were fluent in it, having spoken English from 10–37 years. The protocol for the study was approved by the Institutional Review Board of MGH and Partners Healthcare, protocol #2009P001227.

#### Stimuli

A subset of 168 stimuli from the large set employed in Study 1 was selected for this study. The images from the different conditions (Direct Threat, Indirect Threat, Negative and Neutral) were thematically matched insofar as possible in terms of scene contexts and situations. The set had 34 Direct Threat images, 37 Indirect Threat images, 30 Negative images and 67 Neutral images (Supplementary Figure S5). To reduce habituation to negative stimuli, we included approximately twice as many Neutral images as any other condition. The spatial frequency content of images used in the fMRI experiment was analyzed by computing a periodogram with overlapping Hamming windows (a power spectral density estimate using the Welch's method in the MATLAB image processing toolbox, averaged over horizontal and vertical dimensions), and no significant differences ( $P > 0.20$ ) were found among the four stimulus conditions (Supplementary Figure S3).

#### Design and procedure

Participants were asked to view each stimulus attentively and report via a button press whether the scene repeated (i.e. a one-back task) or not. Therefore, the subjects pressed a button on each trial. However, they were instructed that this was not a speeded response task and they were to respond after the stimulus offset to minimize head movement artifacts. Approximately 33% of the images repeated exactly once for the one-back task; these repeat trials were included in the general linear model (GLM) as a condition of no interest and not further analyzed. We employed an event-related design in which 25% of the trials were null trials, and two different stimulus presentation orders, each of which was randomly assigned to approximately half of the participants (11 and 10, respectively). The parameters for the stimulus presentation sequences (average efficiency, average variance reduction factor and stimulus counterbalancing) were optimized using the event-related design optimization program 'optseq' (<http://surfer.nmr.mgh.harvard.edu/optseq/>). The trials were 2 s long and the stimulus duration was 1000 ms, with its onset randomly jittered 300–700 ms within each trial. Stimuli were presented via a head coil-mounted mirror and a rear-projection system, and response-box key presses were recorded via a custom-made optical MRI compatible button box. A MacBook laptop running a custom program written in Matlab (version 7.9, Mathworks Inc.) and Psychtoolbox version 3.08 (Peli, 1997), controlled stimulus presentation and manual response collection.

### MRI data acquisition and analysis

Structural and fMRI data for this study were acquired with a Siemens Magnetom Trio Tim 3.0T scanner and a 32-channel head coil. After acquiring localizer, and automatic alignment scout images, a high-resolution structural T1-weighted MPRAGE image was collected in 176 sagittal  $1.0 \times 1.0 \times 1.0$  mm thick slices, with TR (repetition time) = 2530 ms, TE (echo time) = 3.45 ms and flip angle =  $7^\circ$ . Then, after 20 practice trials to familiarize participants with the task, fMRI data for the main experiment were collected in a functional scan lasting 11 min and 12 s, which comprised 336 volumes. Four 'dummy' volumes were acquired at the beginning of the experiment to achieve steady-state magnetization and later discarded. Each volume was acquired in 33 interleaved slices with  $3.1 \times 3.1 \times 3.0$  mm thickness. We used gradient-echoplanal imaging with a TR = 2000 ms, TE = 30 ms and flip angle of  $90^\circ$ . The slice prescription was tilted anteriorly  $\sim 30^\circ$  from the anterior-posterior commissure (AC-PC) line to minimize the susceptibility artifact in the OFC due to tissue-air interface (Kringelbach and Rolls, 2004), as in (Kveraga et al., 2007). The structural and functional images were examined for quality (excessive motion, ghosting and any other artifacts) during both acquisition (using the Siemens Syngo online viewer) and offline. We employed the SPM8 analysis software (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>) for pre-processing and statistical analysis of our imaging data, using the standard processing steps: the functional volumes were realigned to the fifth volume in the series, motion-corrected, aligned with the structural images, normalized to a standard Montreal Neurological Institute (MNI) template, and smoothed with an 8 mm full-width half-maximum (FWHM) isotropic Gaussian kernel. The GLM included regressors for each of the conditions of interest (Direct Threat, Indirect Threat, Negative, Neutral), a condition of no interest (Repeated images) and motion-correction regressors to minimize contributions of remaining motion to the error term. For display purposes, we computed a group average brain from high-resolution anatomical scans of our 20 subjects volumetrically using SPM8, as well as a group average surface using 2D surface alignment techniques implemented in Freesurfer (Fischl et al., 2004).

#### Regions of interest

The main goal of this study was to query how different types of affective information (Threat vs Merely negative) are extracted from image context. Therefore, in the fMRI component of the project, our a priori regions of interest (ROIs) included regions implicated in detecting and responding to threat, affective and scene perception. The regions implicated in detecting and responding to threat whose response we wanted to test in this study included the amygdala, the PAG, and the OFC and vmPFC (Barbas and De Olmos, 1990; Barbas, 1995; Carmichael and Price, 1995; LeDoux, 2000; Ghashghaei and Barbas, 2002; Kondo et al., 2003; Wager et al., 2008; Barrett and Bar, 2009). The regions that have been implicated in scene, context and affective perception were the parahippocampal cortex (PHC), involved in scene processing affective memory, and contextual processing; retrosplenial complex, encompassing the retrosplenial cortex (RSC), posterior cingulate cortex (PCC) extending into the precuneus and the medial prefrontal cortex (mPFC) (Aguirre et al., 1996; Epstein and Kanwisher, 1998; Bar and Aminoff, 2003; Bar et al., 2008; Kveraga et al., 2011). Finally, we wanted to test the ventrolateral prefrontal cortex (vlPFC), which has been implicated in threat reappraisal (Wager et al., 2008) and affective reflection processes (Lieberman et al., 2002b; Lieberman, 2003; Cunningham and Zelazo, 2007). We localized our a priori ROIs informed by coordinates of these ROIs from prior studies: for PAG and amygdala (Mobbs et al., 2007); for PHC, RSC and mPFC (Kveraga et al., 2011); and for vlPFC (Wager et al., 2008).

**Results**

**Whole-brain analysis**

We performed whole-brain analyses to map the activation evoked by each type of affective images by comparing it to that elicited by neutral images. Although for properly balancing Type I and Type II errors in this type of fMRI research (i.e. affective/high-level cognitive studies), the recommended significance level is  $P < 0.005$  uncorrected, with extent threshold  $k = 10$  voxels (Lieberman and Cunningham, 2009), we used a somewhat more conservative threshold with a false discovery rate (FDR)-corrected  $\alpha = 0.05$ . All significant activation foci above this threshold are reported in Table 1. In Figures 2 and 3, the activations are displayed at  $P < 0.001$  ( $T > 3.58$ ), with  $k = 10$ . As predicted, Direct Threat images evoked greater activity in the amygdala and PAG (Figure 2), whereas the Merely Negative images elicited strongest activations in the PHC, RSC proper and mPFC. In addition, Merely Negative images differentially activated the hippocampus and the lateral anterior temporal lobes bilaterally (Figure 2). Indirect Threat images also differentially activated PCC, extending into the RSC and precuneus, lateral occipitoparietal cortex extending into the temporoparietal junction (TPJ) and posterior superior temporal sulcus (pSTS), left PHC, dorsomedial PFC, mPFC, and OFC, and bilateral vLPFC (Table 1).

**ROI analysis**

To provide a more targeted test of our hypotheses, we examined time courses of BOLD activity in our a priori ROIs. To do this, we defined a Visual Stimulation vs Null contrast, including all events in which visual stimuli were displayed, and extracted time courses of activity for all four conditions using the rfxplot toolbox (<http://rfxplot.sourceforge.net>) for SPM. We identified the largest peaks in our a priori ROIs using Affective (Direct Threat, Indirect Threat, Merely Negative) vs Neutral images contrast in each ROI. We then defined a 6 mm sphere around it and, using rfxplot toolbox in SPM8, extracted all voxels from each individual participants functional data in that sphere, whether above threshold or not, using a finite impulse response model for time points from 0 to 16 s from event onset. Therefore, the ROI extraction procedures made no assumptions about the shape of the hemodynamic response, nor was it circular (Kriegeskorte et al., 2009) with regard to the comparison interest (time courses of activity between Direct Threat, Indirect Threat and Merely Negative images), because there was no a priori statistical bias for one (or more) of those conditions to show a higher activity level than the other(s). The extracted time courses were baseline corrected and are displayed in Figure 3A.

As we had predicted, the Direct Threat images elicited the largest early response in the amygdala and the PAG, with a smaller response to Indirect Threat images, and only a late peak evoked by Merely Negative images (Figure 3A, right). Moreover, in the PHC and in RSC proper, this activity pattern was reversed, with Merely Negative images now eliciting the largest response, followed by Indirect Threat images and then Direct Threat images (Figure 3A, left). Finally, all unpleasant images initially evoked responses in mPFC and left lateral OFC ROIs. Later in the timecourse, the threat-related response declined in these regions, whereas the Merely Negative images-related activity continuing to rise to a late peak (Figure 3A, bottom row). The overall differences and overlapping patterns of activity for the three types of affective images are displayed in Figure 3B.

**DISCUSSION**

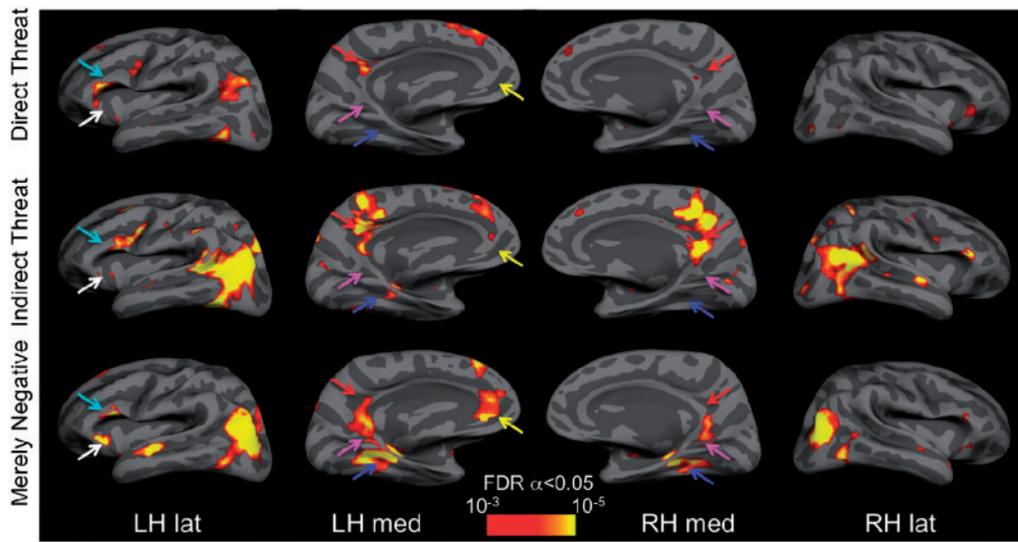
Although people have subjectively unpleasant affective experiences to many different events and objects in the world, the brain does not represent all negative stimuli identically. In these studies, we found that observers were exquisitely sensitive to the threat potential depicted

**Table 1** BOLD activations

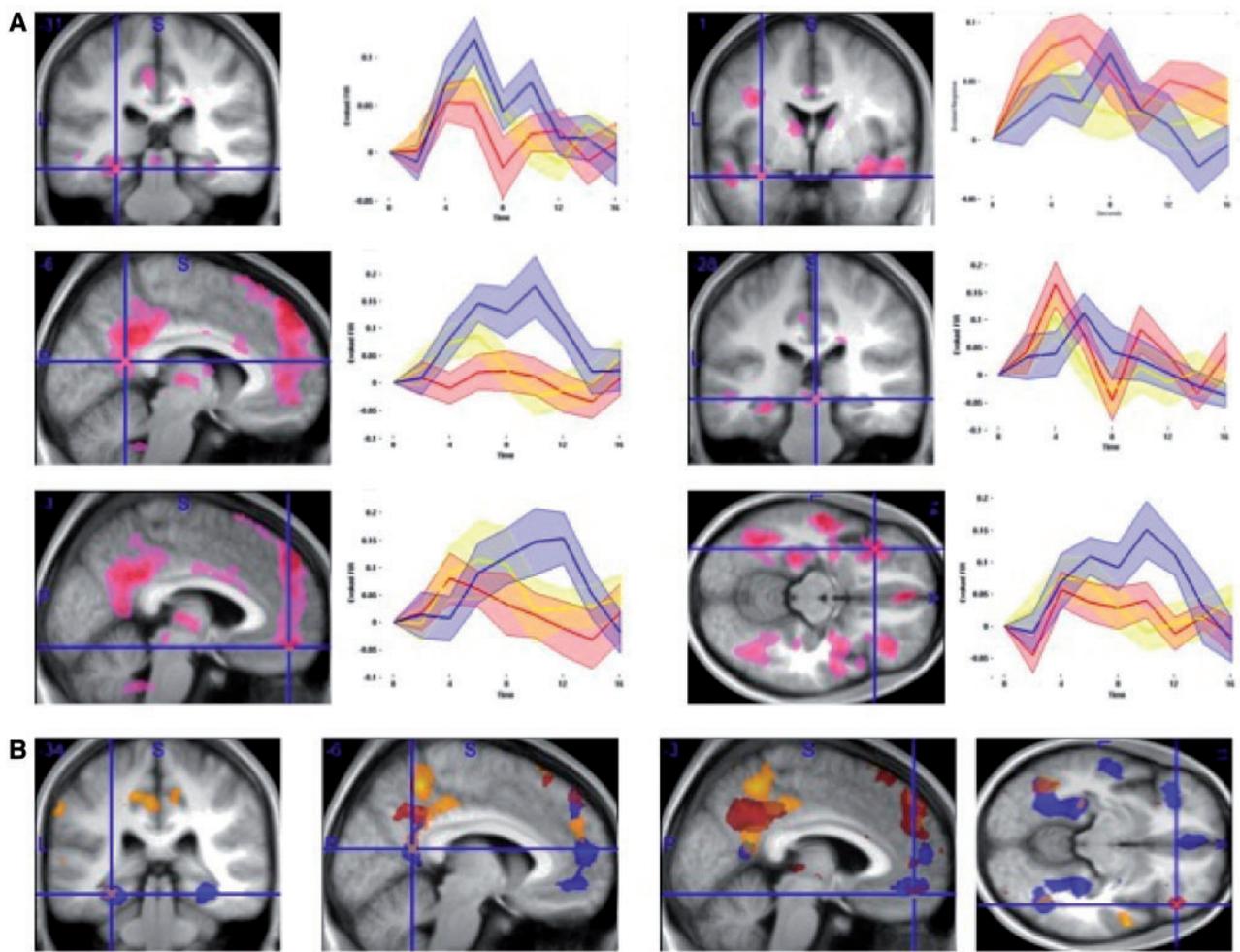
Clusters	k	FDR P	T	Z	MNI coordinates		
					x	y	z
<b>Merely Negative</b>							
L EBA	590	0.00	7.9	5.2	-45	-85	10
R EBA	338	0.00	6.4	4.6	48	-79	10
L PHC	355	0.00	7.2	4.2	-30	-34	-17
R PHC	376	0.00	5.4	4.2	33	-34	-14
L ATL	151	0.00	5.7	4.3	-54	-10	-17
R ATL	34	0.01	4.6	3.7	45	20	-29
L IOFC	181	0.00	5.6	4.7	-42	29	-8
L vLPFC	24	0.03	4.4	3.6	-48	29	10
L DMPFC	64	0.00	5.4	4.1	-6	50	49
L MOFC/vmPFC	178	0.00	6.1	4.5	-3	41	-14
L mPFC	— <sup>a</sup>	0.01	4.8	3.8	-6	56	10
<b>Direct Threat</b>							
R cerebellum	37	0.04	5.5	4.2	15	-85	-38
L PAG	3	0.05	3.7	3.2	-3	-28	-8
R PAG	4	0.04	3.8	3.3	6	-28	-8
L Amyg	26	0.05	3.9	3.3	-30	-1	-20
R Amyg	29	0.06	3.6	3.1	36	-1	-17
L HC	12	0.04	4.3	3.6	-33	-16	-11
L PMC	1189	0.04	6.2	4.5	-30	5	37
L LPC	373	0.04	5.1	4.0	-45	-73	40
L PCC	311	0.04	5.1	4.0	-12	-49	31
R FG	133	0.04	4.8	3.9	42	-70	-11
L FG	168	0.05	4.2	3.5	-39	-67	-11
R lat OFC	86	0.05	4.2	3.5	45	38	-14
LR med OFC	17	0.06	3.6	3.1	0	50	-14
LR dorsal ACC	81	0.05	4.2	3.5	0	11	25
L thalamus	24	0.05	3.7	3.2	-12	-1	10
L VLPFC	— <sup>a</sup>	0.05	5.7	4.3	-54	32	4
R VLPFC	86	0.05	4.2	3.5	57	38	10
<b>Indirect Threat</b>							
L EBA	950	0.00	7.3	5.0	-45	-67	7
L pSTS/TPJ	— <sup>a</sup>	0.00	5.8	4.4	-54	-46	7
R pSTS/TPJ	554	0.00	5.9	4.4	42	-46	19
R PMC	231	0.00	7.8	5.2	39	14	22
L precuneus	— <sup>a</sup>	0.01	5.9	4.4	-6	-49	55
R PCC/precuneus	759	0.00	5.9	4.4	3	-58	25
L PMC	453	0.00	6.8	4.8	-42	5	25
L DMPFC	117	0.01	4.7	3.8	-9	53	22
L VLPFC	22	0.01	4.1	3.4	-42	29	-5
R PMC	42	0.01	5.5	4.2	54	-1	46
R ATL	23	0.01	5.0	3.9	51	-1	-11
L globus pallidus	26	0.01	5.5	4.2	-3	-7	-2
R cerebellum	24	0.01	4.8	3.8	15	-61	-20
R occipital pole	24	0.00	4.1	3.4	24	-94	-2

k, cluster extent in 3 mm<sup>2</sup> voxels; ACC, anterior cingulate cortex; Amyg, amygdala; ATL, anterior temporal lobe; DMPFC, dorsomedial prefrontal cortex; EBA, extrastriate body area; HC, hippocampus; MOFC, medial orbitofrontal cortex; LOFC, lateral orbitofrontal cortex; LPC, lateral parietal cortex; PMC, premotor cortex. <sup>a</sup>This cluster is part of a larger cluster immediately above.

(or absent) in affectively evocative images, and produced different behavioral and neural response patterns to images that contained direct or indirect threats, and those that were merely negative. The behavioral responses to our stimuli reflected keen awareness of the nature of the situations depicted in the images, and the fMRI findings (in a different cohort) revealed differences in activation patterns consistent with the putative functions of the regions involved in these large-scale brain networks. Specifically, the Direct Threat images received the highest ratings in response to the question about potential impending harm to one's person (Group 1, Figure 1B left). Furthermore, even when a new cohort of participants was not explicitly asked to rate the threat potential of the images, but rather performed an orthogonal one-back memory task, the Direct Threat images evoked



**Fig. 2** Statistical parametric maps for the Direct Threat, Indirect Threat and Merely Negative conditions shown on the inflated group average brain. The BOLD activations evoked by the Direct Threat > Neutral images contrast (top row), Indirect Threat > Neutral images contrast (middle row) and Merely Negative > Neutral images contrast (bottom row). The maps were obtained by computing group ( $N = 20$ ) contrasts in SPM8, aligning them with, and overlaying them on, the group average brain surface made with Freesurfer. All SPMS are corrected for multiple comparisons with the  $FDR \alpha = 0.05$  (which varies contrast to contrast), but displayed at the same, more stringent threshold. The arrows indicate some of the ROIs with notable activation differences for the different scene categories: PHC (blue), RSC (magenta), PCC (red), OFC/vmPFC (white), vIPFC (cyan). Images in the different categories did not differ in their spatial frequency content (Supplementary Figure S1).



**Fig. 3** ROI analyses. (A) The ROI foci from which activity was extracted based on Affective images (Direct Threat, Indirect Threat, Merely Negative) vs Neutral images contrast (see ROI analysis in Methods for details). Top left: PHC, middle left: RSC, bottom left: mPFC, top right: right amygdala, middle right: PAG, bottom right: left OFC. (B) The statistical parametric maps for Direct Threat (red colormap), Indirect Threat (yellow colormap) and Merely Negative (blue colormap) vs Neutral images contrasts.

the strongest responses and early peaks of activation in the amygdala and the PAG, regions that have been implicated in detecting threat and eliciting defensive responses (LeDoux, 2012). Conversely, the Merely Negative images received the highest ratings in response to the question about past harm to someone else (Group 3, Figure 1B right), and evoked the largest activations in the PHC, RSC and mPFC, regions implicated in contextual associative processing, mentalizing and reflection (Bar and Aminoff, 2003; Buckner and Carroll, 2007; Bar *et al.*, 2008; Kveraga *et al.*, 2011; Aminoff *et al.*, 2013). The intermediate situations, in which there was an impending harm to someone else, but not necessarily to one's person, evoked the highest ratings on the question about impending harm to someone else (Group 2, Figure 1B, middle), and intermediate neural responses which typically fell between those evoked by Direct Threat and Merely Negative images (Figure 3A).

The Direct Threat images could be thought of as clear visual depictions of impending personal threat and activated the amygdala and the PAG, as well as a host of other cortical and subcortical regions (Table 1, Figure 3) to the largest extent. However, the Indirect Threat images, in which the target of the threat can be more ambiguous, also activate these structures early, albeit with a slighter smaller amplitude, and the Merely Negative images evoke a delayed peak (Figure 3A). It is important to stress at this point that latency differences in BOLD responses have been demonstrated for years (e.g. Duff *et al.*, 2007; Fuhrmann Alpert *et al.*, 2007), including those between visual and affect-processing regions (Sabatinelli *et al.*, 2009), and differences in the amygdala specifically between participants (spider phobics and controls), and between phobogenic and control stimuli (Larson *et al.*, 2006).

The amygdala is a complex agglomeration of nuclei which plays a broad role in representing unique and relevant stimuli, and thus is particularly important for affect and novelty (LeDoux, 2000; Pessoa and Adolphs, 2010; Weierich *et al.*, 2010; Moriguchi *et al.*, 2011), including identifying motivationally salient visual cues (such as potential threats), resolving stimulus ambiguity (Whalen, 2007), responding to novelty (Moriguchi *et al.*, 2011) and producing contextually appropriate behavior around danger stimuli in humans (Feinstein *et al.*, 2011) and macaques (Bliss-Moreau *et al.*, 2011). A view promoted by Sander, Brosch and colleagues (Sander *et al.*, 2003; Brosch *et al.*, 2010) argues that the amygdala, rather than being simply a 'fear module', is involved in broader, subjective appraisal of the motivational relevance of a stimulus. In particular, stimuli which are evaluated as relevant to the organism's current motivational state robustly engage amygdala activity (Cunningham and Brosch, 2012). The amygdala is thought to accomplish this appraisal via a dual route approach: (i) a fast subcortical route that automatically alerts the organism to potentially threatening stimuli that need to be dealt with immediately and (ii) a slower cortical route that provides more 'top-down' processed information placing the stimulus in relevant context. Our results seem to show evidence of both types of processing: a sharp, early peak in the amygdala (and PAG) response to Direct Threat (and to a slightly lesser extent, Indirect Threat) stimuli; and late-developing peak to Merely Negative stimuli, which also activate regions involved in resolving the contextual association of the stimulus (PHC, RSC, mPFC; Kveraga *et al.*, 2011) to a greater extent than the Threat stimuli. Moreover, we have previously shown the amygdala involvement in processing both clear and ambiguous compound threat cues (fear faces combined with averted or direct eye gaze) depending on processing speed (Adams *et al.*, 2012), presumably engaging reflexive *vs* reflective processing, respectively (Lieberman *et al.*, 2002a,b; Lieberman, 2003). Larson *et al.* found that pictures of spiders evoked early and brief BOLD responses in the amygdala, whereas responses to non-phobic stimuli were weaker in amplitude and spread in time (Larson *et al.*,

2006). We observed sharp, early peaks for the threat images (Direct and Indirect), and a late-peaking rise for the Merely Negative images that was more diffuse in time. The activation for the Merely Negative images may be indicative of re-entrant processing suppressing fear in 'safe' contexts. The PAG likewise showed an earlier, sharper peak in the threat conditions, and a delayed peak for the Merely Negative scenes. The PAG is activated most highly when a predator is near (Mobbs *et al.*, 2007) and in this study the PAG amplitude and peak latency scaled approximately with the imminence and clarity of threat (Figure 3A).

In contrast, the Merely Negative images, which portrayed sad or unpleasant situations in which the threat had already passed, evoked greater BOLD activity in the PHC, RSC and mPFC, regions implicated in contextual association processing (Bar and Aminoff, 2003; Davachi, 2006; Aminoff *et al.*, 2007, 2013; Kveraga *et al.*, 2011), representing the mental state of others, autobiographical memory processing and self-reflection (Buckner *et al.*, 2008). This pattern of BOLD activity is consistent with examining the details of a scene to learn what happens to others in such situations (e.g. an accident, a fire, being the victim of an assault), incorporating self-relevant details of a scene into personal experiences, and perhaps simulating the circumstances and outcomes of such events, in case they were to involve the observer in the future. The prefrontal regions (mPFC, OFC and vPFC) showed increased responses for all the negative conditions, which comports with their high-level functions in threat reappraisal and iterative re-processing (Cunningham and Zelazo, 2007; Wager *et al.*, 2008). However, the peaks evoked by the Merely Negative conditions were delayed in comparison with the images in threat conditions (Figure 3A). Finally, Indirect Threat images, which depict impending harm to others (e.g. a fight that does not presently involve the observer) and produced an intermediate BOLD response (between that evoked by Direct Threat and Merely Negative images, Figure 3) in our a priori ROIs, may evoke something akin to 'mixed feelings' about threat: a combination of curiosity with apprehension about suffering potential harm in real-life situations of this type.

The partly overlapping, large-network activations we have observed in response to Direct Threat, Indirect Threat and Merely Negative images (Figures 2 and 3) lend support to the 'constructivist' view of emotion (e.g. Lindquist and Barrett, 2012; Cunningham *et al.*, 2013; Barrett and Russell, 2014). This view holds that the patterns of interoceptive signals that we categorize as discrete emotions ('anger', 'fear', 'sadness', and so forth) are not tied to specific brain structures that 'specialize' in producing those emotions (in contrast to 'basic emotion' views which posit that emotions can be localized to specific, dedicated neurons (cf. Tracy and Randles, 2011)).<sup>1</sup> Rather, gradients of activity in large and interacting networks shift depending on the perceived meaning of the 'objective' qualities of a stimulus. For example, although the amygdala and the PAG have been implicated in automatic detection of threat and initiation of defensive responses to it (LeDoux, 1996; Saptute and Lieberman, 2006) and were originally labeled as the loci of fear in the brain, they participate in a range of responses not necessarily tied to one particular type of 'discrete' emotion (Lindquist *et al.*, 2012; Oosterwijk *et al.*, 2012). These dynamic shifts produce the internal reactions that lead to overt categorization of our stimuli as images of personal threats, threats to another, negative situations without an impending threat, or situations without negative and/or threatening cues. Indeed, there is no particular reason why a scene in one of our conditions would produce a specific discrete emotion. A scene which portrays an impending personal threat may evoke

<sup>1</sup> what kinds of evidence do the contributors believe should be sought to determine whether a particular state qualifies as a basic emotion? . . . another agreed-upon gold standard is the presence of neurons dedicated to the emotion's activation' (p. 398).

feelings not only of fear but also of anger, resentment, or perhaps even excitement, depending on the observer's willingness to engage in aggression (LeDoux, 2012). A merely negative scene, in which there appears to be no current threat, could cause one to feel sadness, empathy, disgust, curiosity, anger, relief, fear, or a mixture of them.

### Conclusions

Potential threats to self should be dealt with most urgently and evoke an avoidance response, whereas threats to another or threats that are no longer viable instead may evoke an approach response, perhaps to render assistance or examine biologically significant details that could help one avoid a similar fate in the future. By beginning to unravel the interaction between threat cues and their context, we have shown that human observers are exquisitely sensitive to different types of negative stimuli, with their evaluation of the stimuli differing immensely depending on the spatial and temporal direction of the threat.

### SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

### REFERENCES

- Adams, R.B., Jr, Franklin, R.G., Jr, Kveraga, K., et al. (2012). Amygdala responses to averted vs direct gaze fear vary as a function of presentation speed. *Social Cognitive and Affective Neuroscience*, 7(5), 568–77.
- Aguirre, G.K., Detre, J.A., Alsup, D.C., D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, 6(6), 823–9.
- Aminoff, E., Gronau, N., Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, 27, 1493–503.
- Aminoff, E.M., Kveraga, K., Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, 17(8), 379–90.
- Bandler, R., Carrive, P. (1988). Integrated defence reaction elicited by excitatory amino acid microinjection in the midbrain periaqueductal grey region of the unrestrained cat. *Brain Research*, 439(1–2), 95–106.
- Bandler, R., Shipley, M.T. (1994). Columnar organization in the midbrain periaqueductal gray: modules for emotional expression? *Trends in Neurosciences*, 17(9), 379–89.
- Bar, M., Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, 38, 347–58.
- Bar, M., Aminoff, E., Schacter, D.L. (2008). Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience*, 28, 8539–44.
- Barbas, H. (1995). Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 19(3), 499–510.
- Barbas, H., De Olmos, J. (1990). Projections from the amygdala to basoventral and medio-dorsal prefrontal regions in the rhesus monkey. *The Journal of Comparative Neurology*, 300(4), 549–71.
- Barrett, L.F. (2006). Valence as a basic building block of emotional life. *Journal of Research in Personality*, 40, 35–55.
- Barrett, L.F., Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1521), 1325–34.
- Barrett, L.F., Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology*, 41, 167–218.
- Barrett, L.F., Russell, J.A. (1999). Structure of current affect. *Current Directions in Psychological Science*, 8, 10–4.
- Barrett, L.F., Russell, J.A., editors. (2014). *The Psychological Construction of Emotion*. New York: Guilford (in press).
- Bliss-Moreau, E., Toscano, J.E., Bauman, M.D., Mason, W.A., Amaral, D.G. (2011). Neonatal amygdala lesions alter responsiveness to objects in juvenile macaques. *Neuroscience*, 178, 123–32.
- Brosch, T., Pourtois, G., Sander, D. (2010). The perception and categorisation of emotional stimuli: a review. *Cognition and Emotion*, 24(3), 377–400.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1–38.
- Buckner, R.L., Carroll, D.C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57.
- Carmichael, S.T., Price, J.L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, 363(4), 615–41.
- Cunningham, W.A., Brosch, T. (2012). Motivational salience: amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science*, 21(1), 54–9.
- Cunningham, W.A., Dunfield, K.A., Stillman, P.E. (2013). Emotional states from affective dynamics. *Emotion Review*, 5, 344–55.
- Cunningham, W.A., Zelazo, P.D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, 11, 97–104.
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16(6), 693–700.
- Duckworth, K.L., Bargh, J.A., Garcia, M., Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, 13(6), 513–9.
- Duff, E., Xiong, J., Wang, B., Cunnington, R., Fox, P., Egan, G. (2007). Complex spatio-temporal dynamics of fMRI BOLD: a study of motor learning. *Neuroimage*, 34(1), 156–68.
- Ellsworth, P.C., Scherer, K.R. (2003). Appraisal processes in emotion. In: D.R., S.K.R., G.H.H., editors. *Handbook of Affective Sciences*. New York, NY: Oxford University Press, pp. 572–95.
- Epstein, R., Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598–601.
- Feinstein, J.S., Adolphs, R., Damasio, A., Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current Biology*, 21(1), 34–8.
- Fischl, B., Van Der Kouwe, A., Destrieux, C., et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1), 11–22.
- Fuhrmann Alpert, G., Sun, F.T., Handwerker, D., D'Esposito, M., Knight, R.T. (2007). Spatio-temporal information analysis of event-related BOLD responses. *Neuroimage*, 34(4), 1545–61.
- Ghashghaei, H., Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*, 115, 1261–79.
- Kondo, H., Saleem, K.S., Price, J.L. (2003). Differential connections of the temporal pole with the orbital and medial prefrontal networks in macaque monkeys. *The Journal of Comparative Neurology*, 465(4), 499–523.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, 12(5), 535–40.
- Kringelbach, M.L., Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, 72(5), 341–72.
- Kveraga, K., Boshyan, J., Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27(48), 13232–40.
- Kveraga, K., Ghuman, A.S., Kassam, K.S., et al. (2011). Early onset of neural synchronization in the contextual associations network. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), 3389–94.
- Larsen, R.J., Diener, E. (1992). Promises and problems with the circumplex model of emotion. In: Clark, M.S., editor. *Review of Personality and Social Psychology: Emotion*, Vol. 13. Newbury Park, CA: Sage, pp. 25–59.
- Larson, C.L., Schaefer, H.S., Siegle, G.J., Jackson, C.A., Anderle, M.J., Davidson, R.J. (2006). Fear is fast in phobic individuals: amygdala activation in response to fear-relevant stimuli. *Biological Psychiatry*, 60(4), 410–7.
- LeDoux, J. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155–84.
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73(4), 653–76.
- LeDoux, J.E. (1996). *The Emotional Brain*. New York: Simon & Schuster.
- Lieberman, M., Cunningham, W.A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, 4, 423–8.
- Lieberman, M., Gaunt, R., Gilbert, D., Trope, Y. (2002a). Reflexion and reflection: a social cognition neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34, 199–249.
- Lieberman, M.D. (2003). Reflective and reflexive judgment processes: a social cognitive neuroscience approach. In: Forgas, J.P., Williams, K.R., von Hippel, W., editors. *Social Judgments: Implicit and Explicit Processes*. New York, NY: Cambridge University Press, pp. 44–67.
- Lieberman, M.D., Gaunt, R., Gilbert, D.T., Trope, Y. (2002b). Reflection and reflexion: a social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34, 199–249.
- Lindquist, K.A., Barrett, L.F. (2012). A functional architecture of the human brain: emerging insights from the science of emotion. *Trends in Cognitive Sciences*, 16(11), 533–40.
- Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F. (2012). The brain basis of emotion: a meta-analytic review. *The Behavioral and Brain Sciences*, 35(3), 121–43.
- Mobbs, D., Petrovic, P., Marchant, J.L., et al. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, 317(5841), 1079–83.
- Moriguchi, Y., Negreira, A., Weierich, M., et al. (2011). Differential hemodynamic response in affective circuitry with aging: an fMRI study of novelty, valence, and arousal. *Journal of Cognitive Neuroscience*, 23(5), 1027–41.
- Ochsner, K.N., Gross, J.J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–9.
- Oosterwijk, S., Lindquist, K.A., Anderson, E., Dautoff, R., Moriguchi, Y., Barrett, L.F. (2012). States of mind: emotions, body feelings, and thoughts share distributed neural networks. *Neuroimage*, 62(3), 2110–28.
- Peli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4), 437–42.

- Pessoa, L., Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, 11, 773–83.
- Peters, J., Daum, I., Gizewski, E., Forsting, M., Suchan, B. (2009). Associations evoked during memory encoding recruit the context-network. *Hippocampus*, 19(2), 141–51.
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–78.
- Sabatinelli, D., Lang, P.J., Bradley, M.M., Costa, V.D., Keil, A. (2009). The timing of emotional discrimination in human amygdala and ventral visual cortex. *Journal of Neuroscience*, 29(47), 14864–8.
- Sander, D., Grafman, J., Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in Neurosciences*, 14(4), 303–16.
- Sapute, A.B., Lieberman, A.B. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, 1079(1), 86–97.
- Touroutoglou, A., Hollenbeck, M., Dickerson, B.C., Barrett, L.F. (2012). Dissociable large-scale networks anchored in the right anterior insula subserve affective experience and attention. *Neuroimage*, 60(4), 1947–58.
- Tracy, J.L., Randles, D. (2011). Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, 3, 397–405.
- Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., Ochsner, K.N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, 59, 1037–50.
- Watson, D., Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219–23.
- Weierich, M.R., Wright, C.I., Negreira, A., Dickerson, B.C., Barrett, L.F. (2010). Novelty as a dimension in the affective brain. *Neuroimage*, 49(3), 2871–8.
- Whalen, P.J. (2007). The uncertainty of it all. *Trends in Cognitive Sciences*, 11(12), 499–500.
- Wilson-Mendenhall, C.D., Barrett, L.F., Barsalou, L.W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science*, 24(1), 947–56.
- Zhang, S.P., Bandler, R., Carrive, P. (1990). Flight and immobility evoked by excitatory amino acid microinjection within distinct parts of the subventricular midbrain periaqueductal gray of the cat. *Brain Research*, 520(1–2), 73–82.