

A neural signature for the subjective experience of threat anticipation under uncertainty

Received: 29 November 2023

Accepted: 22 January 2024

Published online: 20 February 2024

 Check for updates

Xiqin Liu ^{1,2,3}, Guojuan Jiao^{2,4}, Feng Zhou^{5,6}, Keith M. Kendrick ², Dezhong Yao ², Qiyong Gong ^{1,7}, Shitong Xiang ^{8,9}, Tianye Jia ^{8,9,10,11}, Xiao-Yong Zhang ^{8,9}, Jie Zhang^{8,9}, Jianfeng Feng ^{8,9,12,13} & Benjamin Becker ^{14,15} ✉

Uncertainty about potential future threats and the associated anxious anticipation represents a key feature of anxiety. However, the neural systems that underlie the subjective experience of threat anticipation under uncertainty remain unclear. Combining an uncertainty-variation threat anticipation paradigm that allows precise modulation of the level of momentary anxious arousal during functional magnetic resonance imaging (fMRI) with multivariate predictive modeling, we train a brain model that accurately predicts subjective anxious arousal intensity during anticipation and test it across 9 samples (total $n = 572$, both gender). Using publicly available datasets, we demonstrate that the whole-brain signature specifically predicts anxious anticipation and is not sensitive in predicting pain, general anticipation or unspecific emotional and autonomic arousal. The signature is also functionally and spatially distinguishable from representations of subjective fear or negative affect. We develop a sensitive, generalizable, and specific neuroimaging marker for the subjective experience of uncertain threat anticipation that can facilitate model development.

Uncertainty refers to the inability to predict the outcome of a situation, or the likelihood, valence, intensity, time, or type of future events¹. It represents an essential part of our daily lives ranging from uncertainty in the early stages of a romantic relationship to the likelihood of catching a COVID-19 infection or the magnitude of climate change.

Uncertainty about potential threats in the future and the associated anticipatory processes are central to the feeling of anxiety^{2,3}. While anxiety serves an important adaptive function to avoid or cope with potential danger⁴, excessive anticipatory anxiety in the face of uncertainty represents a key symptom of anxiety disorders^{2,5}. Contemporary

¹Huaxi MR Research Center (HMRR), Department of Radiology, West China Hospital of Sichuan University, Chengdu, Sichuan, China. ²MOE Key Laboratory for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China. ³The Center of Psychosomatic Medicine, Sichuan Provincial Center for Mental Health, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, Sichuan, China. ⁴The Second Xiangya Hospital of Central South University, Changsha, Hunan, China. ⁵Faculty of Psychology, Southwest University, Chongqing, China. ⁶MOE Key Laboratory of Cognition and Personality, Chongqing, China. ⁷Department of Radiology, West China Xiamen Hospital of Sichuan University, Xiamen, Fujian, China. ⁸Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. ⁹Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, (Fudan University), Ministry of Education, Shanghai, China. ¹⁰The Centre for Population Neuroscience and Stratified Medicine (PONS), ISTBI, Fudan University, Shanghai, China. ¹¹SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ¹²MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. ¹³Zhangjiang Fudan International Innovation Center, Shanghai, China. ¹⁴State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China. ¹⁵Department of Psychology, The University of Hong Kong, Hong Kong, China. ✉e-mail: ben_becker@gmx.de

neurobiological frameworks for mental disorders therefore consider ‘potential threat’ and ‘uncertainty intolerance’ as candidate mechanisms of anxiety (e.g., the Research Domain Criteria, RDoC)^{6,7}. However, the neural pathways that underlie the actual subjective experience of uncertainty-induced threat anticipation remain unclear.

Uncertain threat anticipation represents a prototypical paradigm to evoke experimental anxiety^{2,8,9}. Rodent models have identified brain systems that mediate the behavioral and physiological responses to uncertain situations (e.g., shock-probe burying)¹⁰. The bed nucleus of the stria terminalis (BNST), for instance, critically mediates defensive behaviors (e.g., freezing, flight and avoidance) during uncertain threat anticipation^{11–14} and has also been proposed as a key anxiety system in the RDoC framework^{6,7}. Other core regions include the medial prefrontal cortex (mPFC), ventral hippocampus (vHPC), amygdala, insula and thalamus^{10,13,15}. Recent animal research have reconciled the region-focused perspective into circuit-level frameworks demonstrating that anxiety-related responses are mediated by distributed circuits¹⁰. However, the contribution of these neural systems to the conscious experience of anxiety remains unknown because subjective feelings cannot be assessed in animal models^{16,17}. Circuits that underlie defensive behaviors are also distinct from those that generate subjective emotional experiences^{18–22}. Given current treatments based on behavioral and physiological indices are less effective than initially expected^{23,24}, and feelings of excessive anxiety are the primary reason for patients to seek treatment and reduction of subjective symptoms marks treatment success²², a mechanistic understanding of the neural representation that supports the subjective experience of uncertain threat anticipation can facilitate the development of anxiety models^{10,18,22,25,26}.

Human functional magnetic resonance imaging (fMRI) studies have reported BNST activation to uncertain threat anticipation^{27–32}, yet evidence that the BNST encodes subjective anticipatory experience remains controversial. Previous fMRI studies comparing uncertain threat versus safe anticipation conditions (overview see³³) have revealed increased activity in a broad range of brain regions such as BNST^{30,31,34,35}, amygdala^{36–38}, periaqueductal gray (PAG)^{29,34,39}, anterior insula (aINS)^{27–29,40,41}, anterior cingulate cortex (ACC)^{28,29,40,42,43} and lateral and medial frontal regions^{41,44–46}. However, the comparison does not allow for isolating the subjective feeling of uncertain threat anticipation given that the conditions may differ in several other mental processes (e.g., defensive responses or arousal), and the identified regions are involved in fundamental cognitive processes including salience or arousal¹³. Emotional experiences are highly subjective yet accessible via introspective self-report⁴⁷. Recent constructionist theories suggest that the subjective experience is supported by distributed interacting brain regions involved in emotional and non-emotional operations^{48,49}. However, there is limited research focusing on the brain mechanism of actual subjective experiences during uncertain threat anticipation and methodological limitations of the prevailing neuroimaging designs and analytic approaches make it difficult to isolate its specific neural substrates^{21,22}.

One fMRI study determined the neural basis of the subjective experience of anxious anticipation as a function of trial-by-trial self-reported anxiety levels during aversive and neutral trials and showed an association between anxious feelings and reactivity in amygdalo-insular systems⁵⁰. However, a conventional local mapping (mass-univariate) approach employed in this study can only identify isolated brain regions associated with subjective ratings and lacks functional specificity to provide a sufficient brain-level model of an emotional state⁵¹. A growing body of research suggests that many cognitive and emotional processes involve distributed neural coding across multiple brain regions or networks^{52–54}, and it has been proposed to examine population coding of neural activity in terms of multivariate activation patterns⁵⁵. Multivariate pattern analysis (MVPA), therefore, has emerged as a powerful method for capturing

emotion-specific brain states at a fine-grained level⁵⁶. Specifically, multivariate predictive modeling, a machine learning technique based on MVPA, is suggested to be more suitable for identifying biomarkers for specific subjective emotional states with high effect sizes by providing predictions (instead of local mapping) about the emotional experience from distributed neural activity patterns^{51,57}. Chang et al. showed, for instance, that a whole-brain multivariate model explained considerably more variance in predicting experiencing negative affect than local regions^{51,52}, and this predictive modeling approach has been successfully utilized to develop models that can sensitively and specifically predict emotional experiences such as fear²¹, pain⁵³, and pleasure⁵⁸.

Here, we developed an ‘uncertainty-variation threat anticipation (UVTA)’ fMRI paradigm that modulates different aspects of shock uncertainty to capture momentary variations in subjective reports of anxious arousal. Using multivariate predictive modeling, we aimed to determine (1) whether it is possible to develop a process-specific, robust, and generalizable neural representation (‘signature’) predictive of the intensity of subjective anxious experience during uncertain threat anticipation on the population level, (2) whether this signature can accurately track momentary trial-by-trial variations of anxious anticipation on the individual level, and (3) whether and which regions make consistent contributions to the whole-brain predictive models. Next, we systematically determined the extent to which the signature (4) depends on unspecific processes (e.g., negative emotional and autonomic arousal, anticipation per se), (5) differs from signatures of subjective fear or general negative experience, and (6) whether regions such as BNST or ‘salience network’ are sufficient to predict subjective anxious arousal.

To this end, we used 9 datasets (Studies 1–11, $n = 572$), including three fMRI datasets (Studies 1–3, $n = 124$) during which participants experienced varying levels of anticipatory anxious arousal induced by the UVTA paradigm (Fig. 1a). The UVTA paradigm was based on the established ‘threat of shock (TOS)’ paradigm frequently used to induce anxiety during the anticipation of uncertain electric shocks in experimental contexts (see ref. 9). We here modulated uncertainty along different dimensions (occurrence, timing and number of shocks) to induce within-individual variations in subjective ratings of anxious arousal (Fig. 1b). We trained a predictive model of ratings using support vector regression (SVR) on whole-brain activity patterns (Study 1, $n = 44$), termed the ‘shock uncertainty-induced threat anticipation signature’ (SUITAS), which was evaluated in a validation dataset (Study 2, $n = 30$, identical paradigm, Fig. 1c). Using a prospective independent dataset (Study 3, $n = 50$, modified paradigm) and two publicly available datasets (Study 4–5, $n = 127$), we tested the generalizability of SUITAS across cohorts, paradigms, MRI systems and scanning parameters. Next, four independent datasets were used to demonstrate the specificity of the SUITAS for subjective anxious arousal rather than pain experience (Study 6, $n = 33$), general anticipation (Study 7, $n = 100$), unspecific negative emotional (Study 8, $n = 48$, subsample of Study 3) and autonomic arousal (Study 9, $n = 65$, subsample of Studies 2 and 3; Fig. 1d). Finally, two additional datasets were used to determine whether the SUITAS is functionally and topographically distinguishable from established signatures for subjective fear (Study 10, $n = 67$)²¹ and subjective negative affect (Study 11, $n = 121$)⁵² (Fig. 1d). This design allowed a sufficient brain-level description of subjective experience of threat anticipation under uncertainty and we hypothesized that the subjective anxious experience would be decoded by a sensitive, generalizable and specific distributed neural representation that is (partly) separable from representations of fear exposure and negative affect.

Results

Validation of the UVTA paradigm

In Study 1–3, participants underwent the event-related UVTA paradigm, a modified version of the classical ‘TOS’ anxiety induction

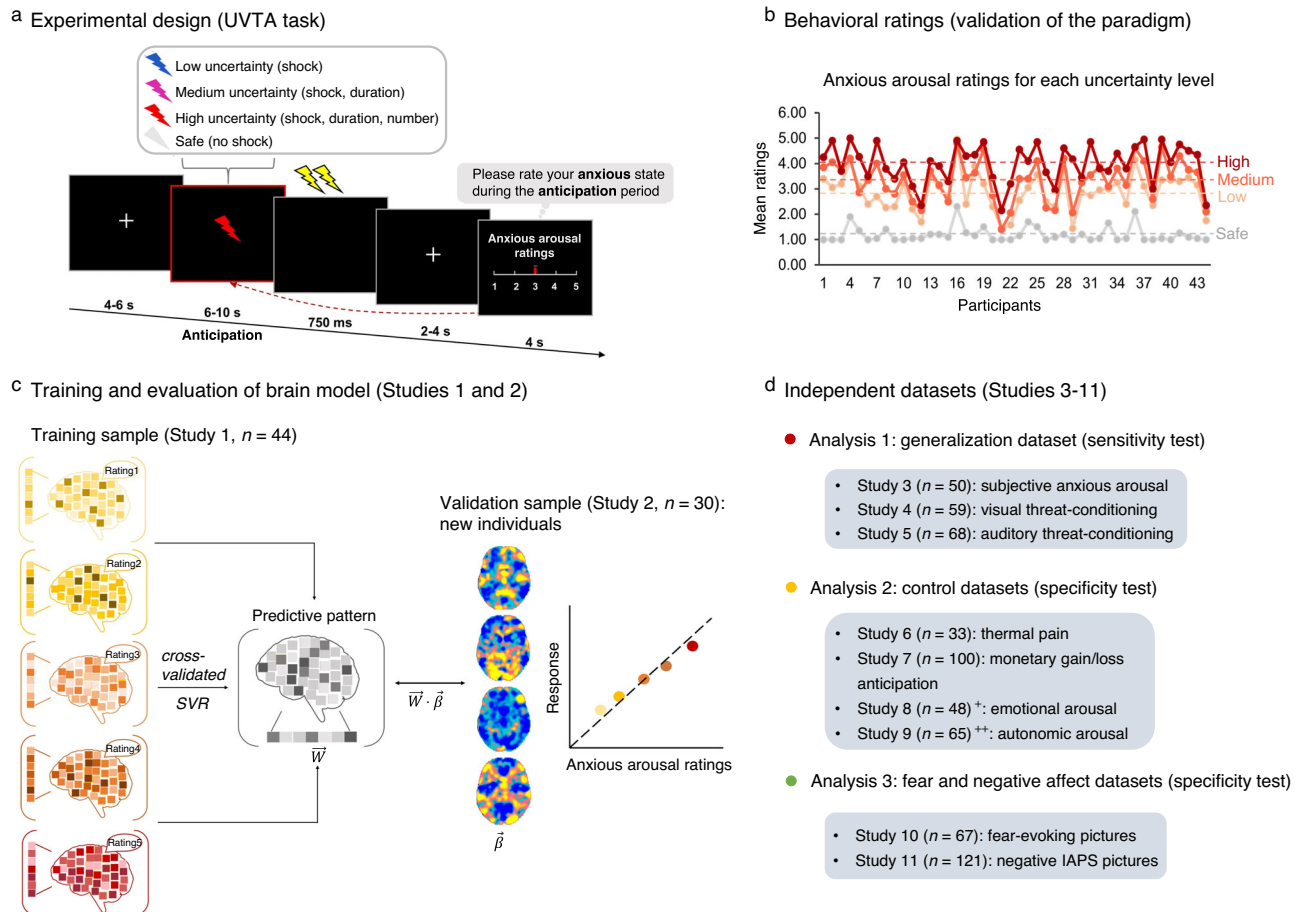


Fig. 1 | Experimental design and analytic strategy. **a** In the UVTA task, participants anticipated potential highly aversive shocks during the cue periods. Each trial began with a 4–6 s fixation cross followed by an anticipatory cue (blue, purple, red, and white, 6–10 s), after which an outcome (either 2–3 shocks or no shocks) was delivered to the participants' left wrists accompanied by a black screen (750 ms). Each cue type signaled a specific uncertainty level. At the end of each trial, participants reported their level of anxious arousal for the anticipatory phase (6–10 s) on a 1–5 Likert scale (4 s). **b** Mean anxious arousal ratings according to the four different uncertainty levels plotted for each participant in Study 1 ($n = 44$). The dash lines indicate the mean ratings for each condition. **c** We computed brain activation images (beta estimates) for the anticipation period (6–10 s) for each level of reported anxious arousal (1–5), and used a support vector regression (SVR) algorithm to predict anxious levels based on brain activity in the training sample (Study

1, $n = 44$, 10 × 10-fold cross-validation) and then applied the model to new individuals in the validation sample (Study 2, $n = 30$) by calculating the predicted level of anxious arousal as the dot product of the trained signature with individual's brain activation map. **d** Validation on independent datasets to test the signature's generalizability (Studies 3–5, $n = 50, 59, 68$), specificity with respect to neural activity during related processes including pain (Study 6, $n = 33$), anticipation of non-shock events (Study 7, $n = 100$), negative emotional arousal (Study 8, $n = 48$) and autonomic arousal (Study 9, $n = 65$) as well as predictive and topological specificity in comparison to subjective fear (Study 10, $n = 67$) and general negative affect (Study 11, $n = 121$). * The sample ($n = 48$) in Study 8 was a subsample of Study 3. ** The sample ($n = 65$) in Study 9 was a subsample of Study 2 and Study 3. UVTA, uncertainty-variation threat anticipation. Source data are provided as a Source Data file.

paradigm⁹. Participants were instructed to anticipate potential threats (highly aversive electric shocks) with four different uncertainty levels (i.e., certain safety, low, medium, and high uncertainty) varied in occurrence, timing and number of shocks indicated by four different colored cues (Fig. 1a). For the certain safety condition, no shocks would be administered after a white cue (6–10 s); for low uncertainty condition, two consecutive shocks might occur immediately after a blue cue (8 s); for medium uncertainty condition, two consecutive shocks might occur immediately after a purple cue (6–10 s); for high uncertainty condition, two or three shocks might occur immediately after a red cue (6–10 s) (details see Methods). At the end of each trial, participants retrospectively rated their subjective level of anxious arousal during the anticipation period (6–10 s) on a 1–5 Likert scale (for details, see Fig. 1a and Methods). The paradigm robustly induced the entire range of anxious arousal, such that 84%, 87%, and 98% of participants rated '1–4' and 59%, 50% and 80% of participants reported all 5 levels of anxious arousal in Study 1, 2 and 3, respectively.

To validate whether different uncertainty conditions resulted in varying subjective ratings of anxious anticipation, we estimated linear mixed-effects models (LMMs) with self-reported ratings as the dependent variable and uncertainty condition as the independent variable in Studies 1, 2, and 3, respectively. We observed the main effects of condition in all three studies (Study 1: $F_{(3,129)} = 367.00$, $P < 0.001$, $\eta^2_p = 0.90$ [0.87,1.00]; Study 2: $F_{(3,87)} = 164.90$, $P < 0.001$, $\eta^2_p = 0.85$ [0.80,1.00]; Study 3: $F_{(3,147)} = 448.54$, $P < 0.001$, $\eta^2_p = 0.90$ [0.88,1.00]), with subjective ratings increased with uncertainty levels (safety < low < medium < high, all post-hoc P s < 0.001, Fig. 1b and Supplementary Fig. 1a), suggesting that our paradigm robustly induced varied levels of anxious arousal.

Moreover, we asked whether participants' subjective ratings for different uncertainty conditions were affected by personality traits such as trait anxiety (TA, measured by the Spielberger State-Trait Anxiety Inventory, STAI⁵⁹) and intolerance of uncertainty (IOU, measured by the Intolerance of Uncertainty Scale-short form, IUS-12⁶⁰) via SurveyCoder 3.0 (<https://www.surveycoder.com/>). LMM analyses

Table 1 | Prediction performance (correlation) of the SUITAS on anxious arousal, fear and negative affect ratings

Prediction	Anxious arousal	Fear	Negative affect
Training dataset	0.59 [0.48, 0.69]; 0.74 ± 0.01 ^a	0.23 [0.12,0.33]; 0.47 ± 0.06	0.22 [0.15, 0.30]; 0.49 ± 0.04
Validation dataset	0.61 [0.50, 0.70]; 0.77 ± 0.03	0.25 [0.05,0.42]; 0.46 ± 0.08	0.26 [0.16, 0.37]; 0.55 ± 0.06
Generalization dataset	0.57 [0.48, 0.65]; 0.75 ± 0.04	0.34[0.19,0.47]; 0.38 ± 0.09	

We applied the SUITAS to subjective anxious arousal, fear, and negative affect datasets and calculated the overall (bootstrapped 95% CI) as well as within-individual (mean ± SE) prediction-outcome correlations between the signature responses and the actual ratings. SUITAS, shock uncertainty-induced threat anticipation signature.

^a indicates cross-validated.

revealed that higher IOU scores were linked to higher anxious experience in uncertain threat conditions (with safety as baseline), yet no effect of TA scores was observed (see Supplementary Results and Supplementary Fig. 2), confirming a specific involvement of intolerance of uncertainty in the anxiety-provoking processes of our paradigm.

SUITAS – a sensitive neural signature predictive of shock uncertainty-induced threat anticipation

We applied machine learning-based MVPA techniques using an SVR algorithm (i.e., predictive modeling) to develop a population-level whole-brain signature in the training dataset [Study 1, $n = 44$; 198 activation maps total, 3–5 maps per participant, gray matter mask, modeled for entire anticipation period (6–10 s), see Methods]. Then, we evaluated the performance of the SUITAS by conducting 10 × 10-fold cross-validation and applying the SUITAS to new individuals in the validation dataset (Study 2, $n = 30$) to calculate the SUITAS pattern expressions for each participant in Study 2 (entire anticipation period, one map per rating, Fig. 1c; see also Methods). The SUITAS accurately predicted subjective ratings in both the training and validation datasets and the overall prediction-outcome correlation coefficients were 0.59 (explained variance score (EVS) = 24%; bootstrapped 95% confidence interval (CI) = [0.48, 0.69], $P < 0.001$; within-participant $r = 0.74 \pm 0.01$, mean EVS = 45.6 ± 2.3%) and 0.61 (EVS = 28%; bootstrapped 95% CI = [0.50, 0.70], $P < 0.001$, one-sided permutation test; within-participant $r = 0.77 \pm 0.03$, mean EVS = 50.5 ± 5.9%; Table 1, Fig. 2a), respectively. Forced-choice tests indicated that the SUITAS accurately discriminated between high (average of rating 4 and 5) and low (average of rating 1 and 2) anxious arousal in the training and validation datasets (training dataset: accuracy = 100 ± 0%, $P < 0.001$, Cohen's $d = 10.44$; validation dataset: accuracy = 100 ± 0%, $P < 0.001$, Cohen's $d = 1.92$). Further control experiments and analyses were conducted to rule out the possibility that the results were confounded by color- and motor-related responses potentially involved in the paradigm (see Supplementary Methods, Results, and Supplementary Fig. 3–4).

To determine brain regions that reliably contribute to the predictive model and facilitate the interpretation, we performed a bootstrap test through random sampling of participants with replacement from the training dataset with 5000 iterations. Consistent model weights were thresholded to identify important voxels that reliably contributed to the prediction (uncorrected $P < 0.001$, two-sided; Fig. 2b for display purposes) which included positive weights in bilateral aINS, thalamus, ACC, superior parietal lobule (SPL), inferior parietal lobule (IPL), dorsolateral prefrontal cortex (dlPFC) and inferior occipital gyrus (IOG), as well as negative weights in ventromedial prefrontal cortex (vmPFC), supplementary motor area (SMA) and posterior insula (pINS).

Generalization of the SUITAS performance

To test whether the prediction performance of the population-level signature can be generalized to new datasets and paradigms, we applied the SUITAS to a prospective generalization dataset (Study 3, $n = 50$) with a different sample and slightly modified UVTA paradigm (e.g., different shock uncertainty baseline, details see Methods) by calculating the SUITAS pattern expressions for each participant (entire anticipation period, one map per rating). The SUITAS could significantly predict anxious arousal ratings and the overall prediction-outcome correlation coefficient was 0.57 (EVS = 19%; bootstrapped 95% CI = [0.48, 0.65], $P < 0.001$, one-sided permutation test; within-participant $r = 0.75 \pm 0.04$, mean EVS = 53.4 ± 4.6%; Fig. 2a). The forced-choice test indicated that the SUITAS accurately discriminated between high and low anxious arousal in Study 3 (accuracy = 94 ± 3%, $P < 0.001$, Cohen's $d = 4.41$).

To further determine whether the SUITAS could generalize to other paradigms that encompass an uncertain threat anticipation period, we capitalized on two publicly available datasets acquired during threat conditioning with different MRI systems and scanning parameters in which a visual cue (Study 4, $n = 59$, details see ref. 61) or auditory cue (Study 5, $n = 68$, details see ref. 62) (CS+) was paired with a shock on 43% or 33% of the trials, respectively, while a control cue (CS-) was unpaired. We tested whether the SUITAS generalized to distinguish CS+ versus CS-, which is equivalent to uncertain threat versus safe anticipation⁶³. Results showed that the SUITAS accurately classified CS+ versus CS- in both datasets (Study 4: accuracy = 79 ± 5%, $P < 0.001$, Cohen's $d = 1.09$; Study 5: accuracy = 69 ± 6%, $P < 0.005$, Cohen's $d = 0.49$). Overall, the generalizability tests demonstrated that the SUITAS could robustly generalize to new cohorts, paradigms, MRI systems and parameters.

The SUITAS performance in predicting within-individual anxious arousal

To test whether the population-level SUITAS can track moment-to-moment variations in subjective anxious experience on the individual level, the SUITAS was applied to single-trial activation maps of each participant in Studies 1–3. The SUITAS significantly predicted momentary trial-by-trial ratings within individuals (-44 trials for each participant in Study 1 and 2, and -64 trials for each participant in Study 3, see Methods) with mean prediction-outcome correlation between actual and predicted ratings of $r = 0.50$ ($P < 0.001$, bootstrap test) in Study 1, $r = 0.41$ ($P < 0.001$, bootstrap test) in Study 2, and $r = 0.37$ ($P < 0.001$, bootstrap test) in Study 3 (Fig. 2c–e), indicating that the SUITAS was also sensitive to predict within-individual momentary anxious arousal during uncertain threat anticipation.

A neurofunctional core system for the subjective experience of uncertain threat anticipation

Emotional experience is a highly subjective and individually constructed state with interindividual variations⁴⁹. The mental processes and brain patterns that underlie affective states as well as the construction of the affective state may vary greatly across participants²¹. We further developed within-individual predictive model for each participant using single-trial estimates of brain responses (Study 1, -44 trials, 10 × 10-fold cross-validated, Supplementary Fig. 5a) and transformed these within-individual patterns to 'activation patterns' using structure coefficients (Supplementary Fig. 5b) to determine the core system that reliably and consistently contributed to the prediction and encoded the model (Supplementary Fig. 5c, details see Methods).

The core regions included the aINS, thalamic regions (intralaminar and mediodorsal nuclei), dlPFC, IPL, and IOG observed on SUITAS together with anterior midcingulate cortex (aMCC) and parietal regions involved in self-referential processing, e.g., precuneus and

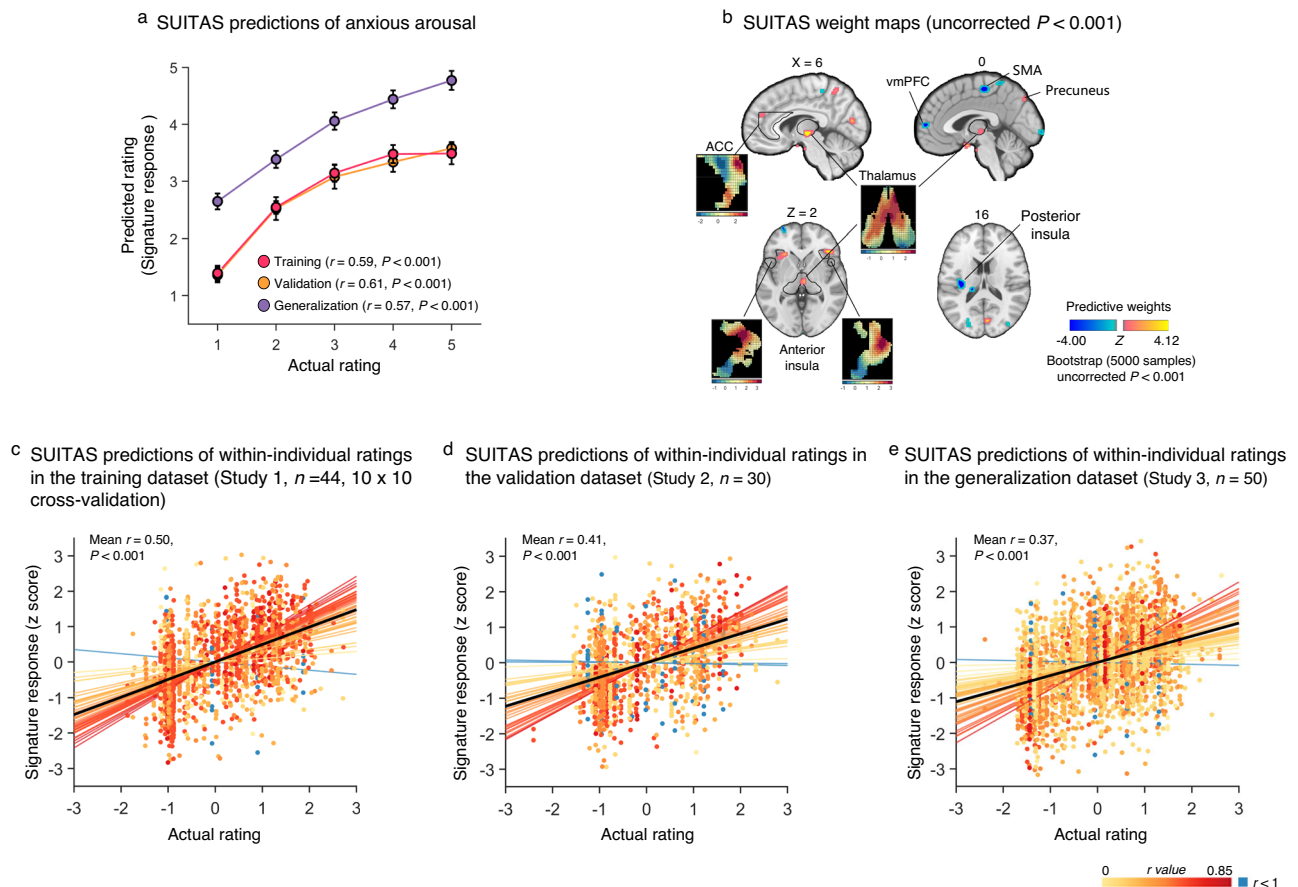


Fig. 2 | Shock uncertainty-induced threat anticipation signature (SUITAS) model evaluation and weight maps. **a** Predicted ratings (signature responses) modeled using the anticipatory brain activity (6–10 s) of $n = 44$ participants in the training dataset (Study 1) compared to actual ratings across participants in Study 1 (10×10 cross-validated, $P = 4.62 \times 10^{-20}$), Study 2 ($n = 30$ participants, validation dataset, one-sided permuted $P < 0.001$) and Study 3 ($n = 50$ participants, prospective generalization dataset, one-sided permuted $P < 0.001$), respectively. r indicates the Pearson correlation coefficient between predicted and actual ratings. The P values for validation and generalization datasets are derived non-parametrically through one-sided permutation tests (5000 random shuffles). Error bars reflect standard errors of the mean. **b** The multivariate pattern of fMRI activity predictive of subjective levels of anxious arousal during threat anticipation (SUITAS weight maps, 10×10 cross-validated, Study 1, $n = 44$) based on a 5000 samples bootstrap test (two-sided). The maps display voxel weights

thresholded at uncorrected $P < 0.001$ for display purposes. Inserts show the spatial topography of the unthresholded patterns in the regions previously proposed as core regions of the anxiety network³³, in particular the ACC, bilateral anterior insula and thalamus. **c–e** Z-scored actual ratings versus predicted ratings (signature response, z-scored) within participants in Study 1 (**c**, $n = 44$, cross-validated), Study 2 (**d**, $n = 30$), and Study 3 (**e**, $n = 50$). The signature response was calculated using the dot product of the SUITAS weight map with each participant's single-trial activation map. Mean r , mean within-participant Pearson correlation between predicted and actual ratings; P values are based on 5000 samples bootstrap tests of within-participant r values (one-sided). Each colored line represents a fitted line for each individual. The black line represents the fitted line across participants. ACC anterior cingulate cortex; SMA supplementary motor area; SUITAS shock uncertainty-induced threat anticipation signature; vmPFC ventromedial prefrontal cortex. Source data are provided as a Source Data file.

posterior parietal cortex (PCC), the putamen and precentral gyrus extending to SMA (FDR $q < 0.05$, one-sided, retaining positive values, Fig. 3; see also Supplementary Fig. 5c for both positive and negative values). To evaluate spatial distribution of the current results, we mapped the core system onto regions showing activations to uncertain threat versus safe conditions from a recent meta-analysis³³ (outlined in Fig. 3). Visualizing the results indicated a high overlap between the identified networks in the meta-analysis and our core systems (e.g., bilateral aINS, thalamus, dlPFC, SMA) but also emphasized a regional specificity of the core systems, probably reflecting a higher specificity for isolating the subjective experience sub-process. In line with the approach employed in the meta-analysis, we additionally examined the conventional univariate categorical effect (uncertain threat > certain safety anticipation) in our training dataset and observed an extensive and rather process-unspecific bilateral activity spanning the entire insular-cingulate network, lateral frontal and subcortical (e.g., BNST, amygdala, thalamus, and PAG), and occipital regions (FDR $q < 0.05$, one-sided, see Supplementary Fig. 6).

Testing the specificity of the SUITAS against pain, general anticipation and unspecific arousal

The multivariate model might capture processes not specific to anxious arousal but those inherently involved in the UVTA paradigm. We employed a series of analyses with independent datasets (details see Methods and Supplementary Table 1) to determine to which extent the SUITAS captures pain experience, general anticipation, and unspecific negative emotional and autonomic arousal. First, we examined whether the SUITAS measured antecedents of the pain response using a publicly available dataset from Wager et al.³³ (Study 6, $n = 33$). Applying the SUITAS to brain activation maps for thermal pain stimulation periods revealed that the SUITAS was less sensitive to predict pain experience ($r = 0.31$, $P = 0.05$, one-sided permutation test) than to predict anxious experience (Study 3: $r = 0.57$, $P < 0.001$, one-sided permutation test; difference in effect size: $\Delta r = 0.26$, $P < 0.005$, one-sided permutation test). Forced-choice tests indicated that SUITAS distinguished high versus low pain experience (accuracy = $85 \pm 3\%$, $P < 0.001$, Cohen's $d = 0.84$, Fig. 4a) less accurately than high versus

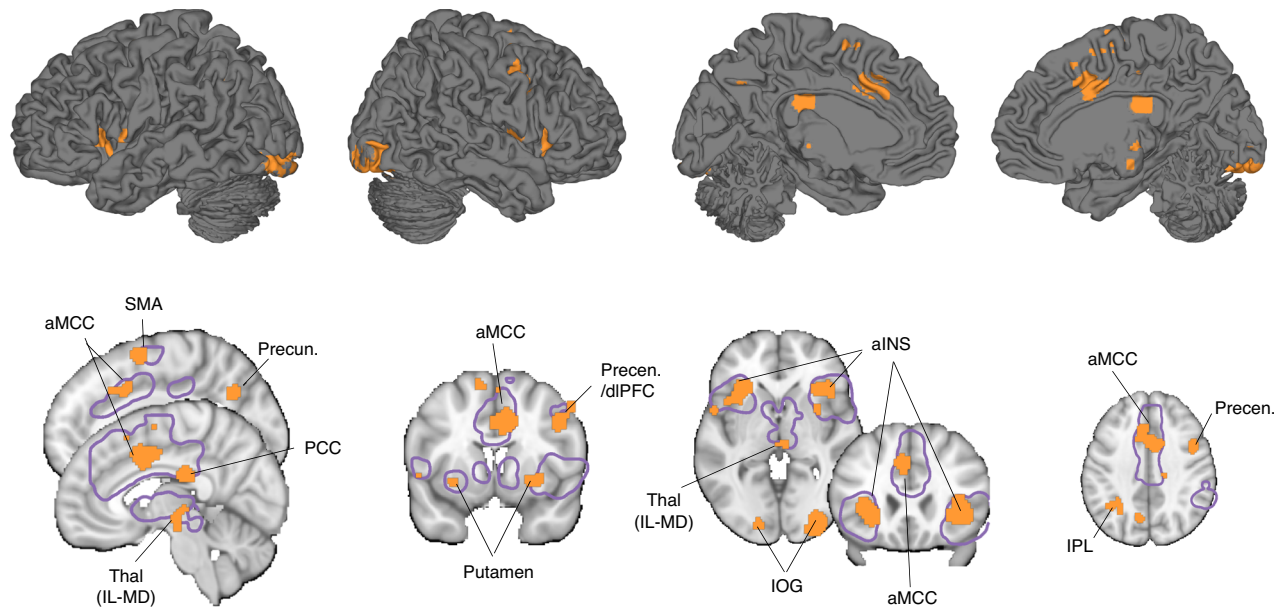


Fig. 3 | Core brain systems for the subjective experience of uncertain threat anticipation. Within-individual core system for subjective experience of threat anticipation under uncertainty using the conjunction (retaining positive values) of the within-individual model weight map (one-sample *t*-test, FDR-corrected $q < 0.05$, one-sided; based on SVR on each participant's single-trial estimates of brain responses) and the transformed within-individual model encoding map using structure coefficients (one-sample *t*-test, FDR-corrected $q < 0.05$, one-sided) of the

training dataset (Study 1, $n = 44$). The violet contour line delineates regions from a meta-analytic activity study comparing uncertain threat versus safe anticipation in healthy individuals³³. aINS anterior insula; aMCC anterior midcingulate cortex; dIPFC dorsolateral prefrontal cortex; IOG inferior occipital gyrus; IPL inferior parietal lobule; PCC posterior cingulate cortex; Precun precuneus; SMA supplementary motor area; Thal (IL-MD), intralaminar and mediodorsal nuclei of thalamus. Source data are provided as a Source Data file.

low anxious experience (accuracy = $94 \pm 3\%$, $P < 0.001$, Cohen's $d = 4.41$, Study 3). Second, we asked if the SUITAS could capture the anticipation of non-shock-related negative events (loss of money) or the anticipation irrespective of valence (anticipation of a positive event, i.e., gain of money). To this end, we tested the SUITAS on an independent dataset that used a monetary incentive delay (MID) task (<https://zib.fudan.edu.cn>)⁶⁴ by classifying negative (monetary loss) and positive (monetary gain) versus neutral (no gain or loss) anticipation (Study 7, $n = 100$). The SUITAS could significantly discriminate loss from neutral (accuracy = $65 \pm 5\%$, $P = 0.004$, Cohen's $d = 0.26$) but not gain from neutral anticipation (accuracy = $54 \pm 5\%$, $P = 0.48$, Cohen's $d = 0.28$, Fig. 4b), although the difference was not significant [(loss vs neutral) vs (gain vs neutral): accuracy = $53 \pm 5\%$, $P = 0.62$, Cohen's $d = 0.02$]. These results indicated that the SUITAS might be more effective at identifying negative anticipation than positive anticipation, yet further research is needed to confirm this.

Third, we accounted for unspecific negative emotional arousal which is inherently associated with uncertain threat anticipation^{30,65} by applying the SUITAS to classify the brain activations during picture viewing period of high-arousing negative (disgust) versus low-arousing neutral visual stimuli (Study 8, $n = 48$). The classification accuracy for high- versus low-arousing stimuli was at chance level (accuracy = $58 \pm 7\%$, $P = 0.31$, Cohen's $d = 0.24$, Fig. 4c), suggesting that unspecific negative emotional arousal did not explain the SUITAS. Since arousing pictures may generally rely stronger on visual processing than UVTA, we recomputed the classification accuracy after excluding the occipital lobe of both SUITAS and the test images for arousal. The classification accuracy remained insignificant (accuracy = $54 \pm 7\%$, $P = 0.67$, Cohen's $d = 0.12$).

Given that physiological responses are expected to (partly) covary with subjective emotional experiences albeit with dissociable neural bases⁶⁶, we finally explored whether the SUITAS predicts subjective anxious experience rather than its physiological correlates by applying the SUITAS to binned brain activation maps for

anticipation period of five skin conductance levels (SCLs) for each participant ($n = 65$). We found that the SUITAS predicted SCLs ($r = 0.22$, $P < 0.001$, one-sided permutation test) to a lesser degree than predicting subjective ratings ($r = 0.56$, $P < 0.001$, one-sided permutation test; difference in effect size: $\Delta r = 0.34$, $P < 0.001$, one-sided permutation test, Fig. 4d, see also Supplementary Results and Supplementary Fig. 7) in a subsample of the combined validation and prospective generalization datasets (Study 9, $n = 65$), demonstrating that the SUITAS captured autonomic arousal to some extent but with a smaller effect size, which is in line with previous studies on the dissociation between subjective fear experience and its physiological correlates⁶⁶.

Further evidence arguing against the effect of nonspecific negative arousal can be found below ('Comparing the SUITAS with the predictive models of fear exposure and nonspecific negative affect'). Together with the cross-signature evaluation in the following paragraph, a series of analyses confirm a comparably high specificity of the SUITAS for predicting anxious experiences during uncertain threat anticipation.

Comparing the SUITAS with the predictive models of fear exposure and nonspecific negative affect

We further determined the distinctiveness of our signature from that of fear exposure and general negative affect by comparing the functional and spatial similarities of the SUITAS with the established predictive models of the subjective experience of fear (visually induced fear signature, VIFS)²¹ and the subjective negative affect (Picture Induced Negative Emotion Signature, PINES)³² during picture viewing periods. The prediction performance of the SUITAS on subjective anxious arousal (Study 3: $r = 0.57$, $P < 0.001$, one-sided permutation test) was twice as high as on subjective fear ($r = 0.23$, $P = 0.015$; $\Delta r = 0.34$, $P < 0.001$, one-sided permutation test) and negative affect ($r = 0.22$, $P = 0.014$; $\Delta r = 0.35$, $P < 0.001$, one-sided permutation test; Table 1, Fig. 5a) while the VIFS and PINES more

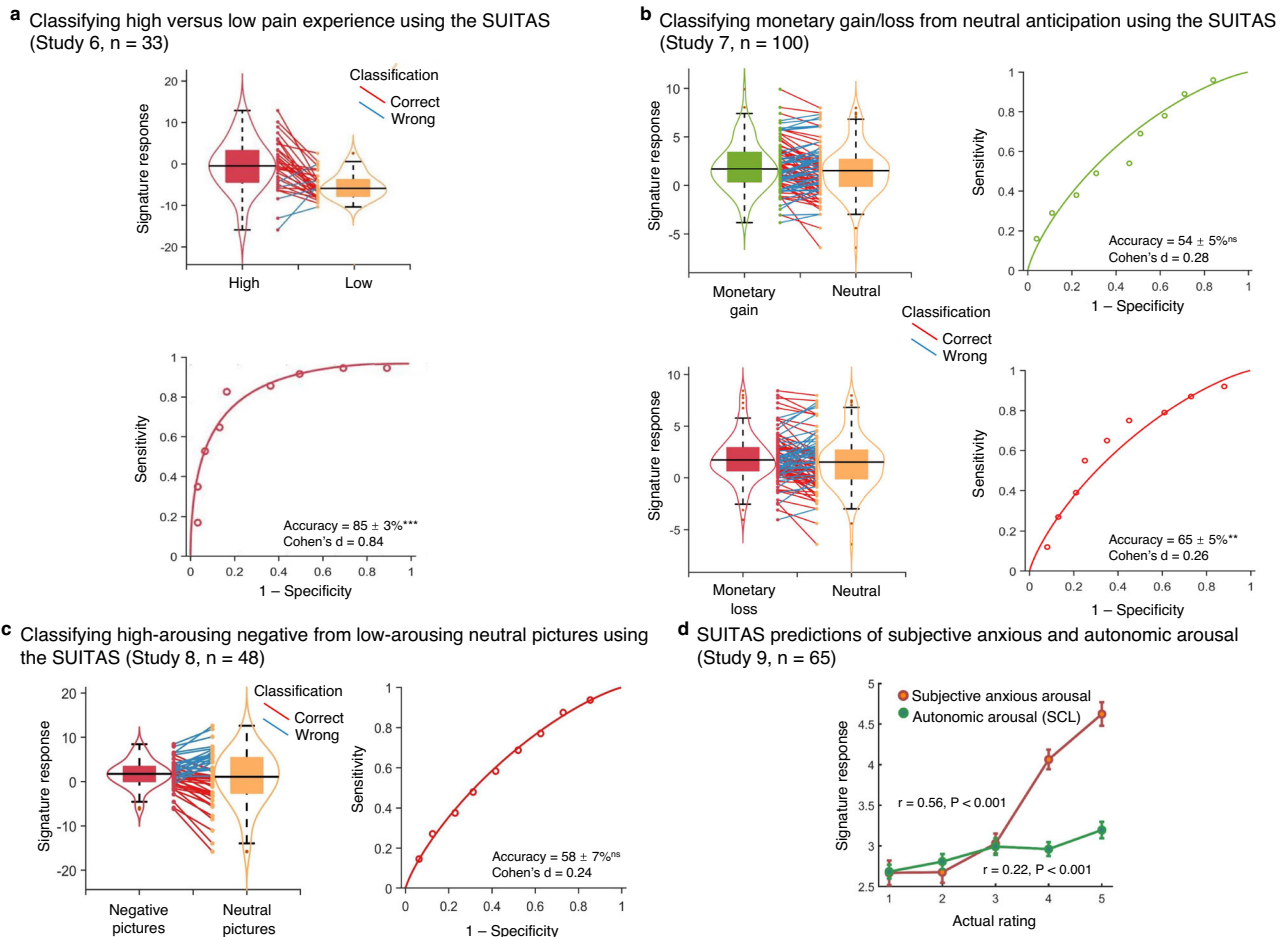


Fig. 4 | Testing the specificity of the SUITAS against pain, general anticipation and unspecific negative emotional and autonomic arousal. a–c We used SUITAS to classify between high versus low pain stimulation (Study 6, $n = 33$), monetary gain/loss versus neutral anticipation (Study 7, $n = 100$), and high arousing negative (disgust) versus low arousing (neutral) picture-viewing (Study 8, $n = 48$). The violin and box plots show the distributions of the signature response. The box was bounded by the first and third quartiles, and the whiskers stretched to the greatest and lowest values within the median ± 1.5 interquartile range. The data points outside of the whiskers were marked as outliers. Each colored line between dots represents each participant's paired data (red line, correct classification; blue line, incorrect classification). The receiver operating characteristic curves illustrate the

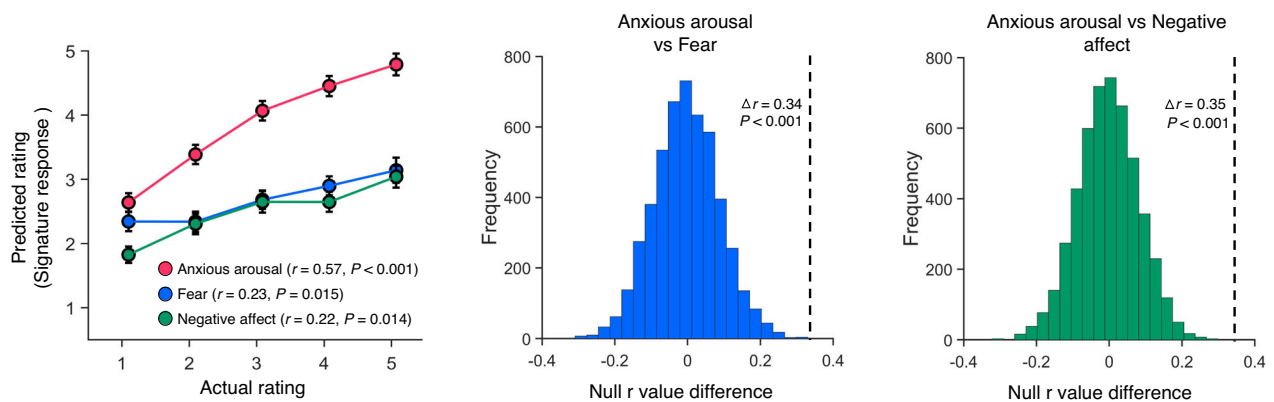
discrimination ability of the SUITAS using forced-choice tests for each of the three studies. Performance was shown as accuracy \pm SE and Cohen's d . *** $P < 0.001$, ** $P < 0.01$, ^{ns} $P > 0.05$, two-sided binomial tests. **d** Using the SUITAS to predict autonomic arousal (one-sided permutated $P < 0.001$) and subjective anxious experience (one-sided permutated $P < 0.001$) during the anticipation period in the UVTA task in a subsample of the combined validation and prospective generalization datasets (Study 9, $n = 65$). r , the Pearson correlation between actual and predicted ratings (signature response). The P values are derived nonparametrically through one-sided permutation tests (5000 random shuffles). SCL skin conductance level. Error bars reflect standard errors of the mean. Source data are provided as a Source Data file.

accurately predicted subjective ratings of fear and negative affect as compared to the SUITAS (Supplementary Table 2). Applying the SUITAS to classify high versus low levels of subjective anxious arousal, fear and negative affect further confirmed that the SUITAS distinguished high versus low subjective anxious arousal (accuracy = $94 \pm 3\%$, $P < 0.001$, Cohen's $d = 4.41$) more accurately than high versus low fear (accuracy = $79.10 \pm 5\%$, $P < 0.001$, Cohen's $d = 0.83$) and negative affect (accuracy = $77.14 \pm 4\%$, $P < 0.001$, Cohen's $d = 0.80$) in terms of classification accuracy and effect size (see also Supplementary Fig. 8). These results suggest that the SUITAS predicts subjective experience of anxious arousal with high specificity relative to that of fear, subjective negative affect or nonspecific negative arousal. Next, we compared the spatial topography of the SUITAS, VIFS and PINES (Fig. 5b). The pattern similarity among the SUITAS, VIFS and PINES weight maps restricted to the gray matter mask suggests that these models exhibited weak positive spatial correlations on the whole-brain level (SUITAS versus VIFS: $r = 0.07$; SUITAS versus PINES: $r = 0.05$; VIFS versus PINES: $r = 0.08$, all P s < 0.001 , one-sided permutation tests). The Pearson correlations decreased after

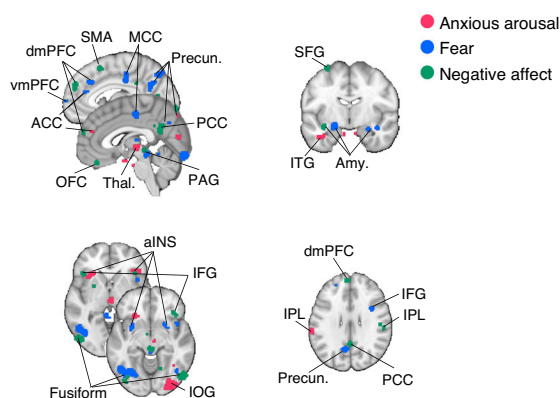
thresholding the model weights at both uncorrected $P < 0.01$ and uncorrected $P < 0.001$ (see Supplementary Results for details). To test to what extent the similarity or distinction of performance depended on the contribution of the visual cortex, we retrained these models excluding the occipital lobe and then compared the functional and spatial similarities and the results remained consistent (see Supplementary Results and Supplementary Fig. 9), suggesting little contribution of visual cortex in the differentiation among predictive models of threat anticipation, fear exposure and negative affect.

We further used a river plot to illustrate the spatial similarity between the predictive weights of these models (uncorrected $P < 0.001$) and a set of a priori regions of interest (ROIs, see Supplementary Table 3) previously linked to negative emotion processing including anxious anticipation, fear exposure and general negative affect^{33,67,68}. As shown in Fig. 5c, each ROI encoded at least one model with different contributions: thalamus, dlPFC, aINS, and vlPFC were contributed most by SUITAS, whereas MCC, amygdala, PAG, and vmPFC were contributed most by the VIFS and dmPFC was contributed

a SUITAS predictions of subjective anxious arousal, fear and negative affect



b Model weight maps for anxious arousal, and negative affect



c Spatial similarity with ROIs across anxious arousal, fear and negative affect

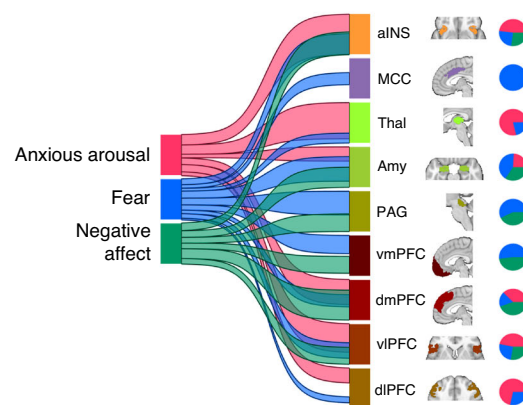


Fig. 5 | Comparing the SUITAS with the predictive models of fear exposure (VIFS) and negative affect (PINES). **a** Using the SUITAS to predict subjective reports of fear experience (VIFS, Study 10, $n = 67$) and negative affect experience (PINES, Study 11, $n = 121$) based on samples acquired in the previous studies^{21,52}. The Pearson correlation coefficient (r value) between actual and predicted ratings (signature response) of anxious arousal using the generalization dataset (Study 3, $n = 50$, one-sided permutated $P < 0.001$) was twice as high as that of fear (one-sided permutated $P = 0.015$) and negative affect (one-sided permutated $P = 0.014$) (left panel), and the differences of r values (Δr , dash lines in the middle and right panel) were significant based on permutation tests of r value differences with 5000 random shuffles (one-sided permutation test $P_s < 0.001$). Error bars reflect standard errors of the mean. **b** Spatial topography of the weight maps for anxious arousal, fear and negative affect, each thresholded at $P < 0.001$ (bootstrap test, two-sided, uncorrected) within gray matter (retaining positive values). **c** River plot depicting the spatial similarity (computed as cosine similarity) of the thresholded weight

maps (bootstrap test $P < 0.001$, uncorrected, two-sided, retaining positive voxels) in Fig. 5b with anatomical labels of predefined ROIs previously linked to negative emotion processing. Ribbons are normalized by the max cosine similarity across all ROIs. Ribbon locations in relation to the boxes are arbitrary. Pie charts show relative contributions of each model to each ROI (that is, the percentage of voxels with the highest cosine similarity for each predictive map). ACC anterior cingulate cortex; aINS anterior insula; Amy amygdala; dlPFC dorsolateral prefrontal cortex; dmPFC dorsomedial prefrontal cortex; IFG inferior frontal gyrus; IOG inferior occipital gyrus; IPL inferior parietal lobule; ITG inferior temporal gyrus; MCC midcingulate cortex; OFC orbitofrontal cortex; PAG periaqueductal gray; PCC posterior cingulate cortex; PINES Picture Induced Negative Emotion Signature; Precun precuneus; SFG superior frontal gyrus; SMA supplementary motor area; Thal thalamus; VIFS visually induced fear signature; vlPFC ventrolateral prefrontal gyrus; vmPFC ventromedial prefrontal gyrus. Source data are provided as a Source Data file.

most by the PINES. Summarizing, the double dissociation of predictive performance and the distinct spatial topography suggest that these signatures show distinct representations in predicting their corresponding subjective emotional states.

Single subsystems are not sufficient to predict subjective experience of threat anticipation under uncertainty

Previous 'structure-centric' theories proposed that emotions are localizable in single brain regions or networks^{18,69–71}. To determine whether the subjective emotional experience of threat anticipation under uncertainty can be reduced to specific systems, we re-trained the predictive models in (1) five prior regions considered as 'classical' anxiety systems^{13,33}; (2) salience network hubs associated with anxiety and anxiety disorders⁷²; (3) cortical network related to conscious emotional experience¹⁸; (4) seven large-scale functional networks⁷³,

and tested them on training, validation and generalization datasets. Isolated regions (i.e., aINS, ACC, thalamus, and BNST, but not PAG) and networks could predict subjective ratings, however, with considerably smaller effect size as compared to the whole-brain model (see Supplementary Results and Supplementary Fig. 10a). To control for the potential effect of the numbers of features in the prediction analyses (i.e., whole-brain model uses much more features/voxels), we randomly selected certain numbers of voxels from a uniform distribution spanning the whole brain or aforementioned networks. The asymptotic predictions with sampling from all brain systems were consistently higher than sampling the same number of voxels from the individual networks when exceeding 1000 voxels (Supplementary Fig. 10b), suggesting that the whole-brain model has a much larger effect size than those using the same number of features from single networks.

In line with accumulating evidence^{21,52}, the above results underscore that subjective experiences of uncertain threat anticipation require a distributed neurofunctional representation.

Discussion

Aberrant anticipatory responding towards uncertain threats is central to anxiety-related disorders²¹. We here developed an uncertainty-variation threat anticipation paradigm and utilized multivariate predictive modeling to systematically determine an accurate, robust, generalizable, and specific fMRI-based neuro-affective signature for the subjective experience of threat anticipation under uncertainty using several independent datasets in healthy adults. The developed SUITAS was predictive of the level of uncertainty-induced subjective anxious experience on both the population and individual levels. The signature showed a robust generalization across cohorts and paradigms and was less or not sensitive to processes inherently interwoven with the paradigm, including pain, positive anticipation (but negative anticipation), or unspecific negative emotional and autonomic arousal, suggesting a comparably high specificity for predicting anxious anticipation of aversive events. Comparison with established decoders for subjective fear and negative affect indicated a certain extent of cross-prediction, suggesting that the decoder may partly capture a common underlying process, e.g. conscious emotional experiences or interpretation⁷⁴, but the considerably higher effect size for predicting the target emotional state also suggests distinguishable neural representations of uncertain threat anticipation, fear exposure, and negative affect. Consistent model weights associated with subjective anxious arousal were observed in the aINS, thalamus, ACC, somatosensory cortices, and IOG, while the within-individual models additionally revealed consistent contributions of aMCC, precuneus, PCC, putamen, precentral gyrus and SMA to predict momentary trial-by-trial variations. No single brain region or network was sufficient for accurately predicting anxious experience, underscoring that conscious emotional experiences require a distributed neural representation^{10,21,33,52,66,75}. The neuro-affective signature provides a promising neuroimaging biomarker for subjective anxious experience during uncertain threat anticipation which may facilitate rapid and accurate evaluation of new interventions targeting subjective anxious arousal related to uncertainty (for similar approach see e.g. ref. 76).

Uncertainty about potential threats has long been considered as a key candidate mechanism underlying anxiety^{2,77,78}. However, how varying levels of uncertainty impact the subjective experience of threat anticipation remained unknown. Our UVTA paradigm based on the established TOS paradigm - but manipulating different aspects of anticipatory uncertainty - successfully induced sufficient and varying intensity levels of anxious anticipation. Moreover, individuals high in intolerance of uncertainty reported higher anxious arousal in the uncertain conditions of the UVTA task. These results together demonstrated uncertainty as a candidate mechanism underlying the experience of anxious arousal and allowed us to determine the underlying neuro-affective signature.

Employing a pattern recognition-based machine learning technique that has been previously successfully employed to other subjective emotional experience domains^{21,52,53,74,79,80} allowed us to identify a robust and process-specific predictive pattern for the subjective experience of uncertain threat anticipation and segregate it from other emotional states. Importantly, out-of-sample predictions in two independent samples (study 2 and 3) protected against overfitting and showed the robustness and generalizability of the brain-wide model⁸¹. The generally higher signature response across rating levels of the prospective generalization dataset might be attributed to the higher baseline of uncertainty related to including a higher proportion of uncertain trials of the UVTA task in Study 3 (see Fig. 2a).

The SUITAS provides evidence that the subjective experience of uncertain threat anticipation involves distributed neural systems⁷⁵,

which resembles current observations in animal models that the response to uncertain threats recruits a distributed array of interconnected neural ensembles¹⁰. Further analyses demonstrated that no single region or network was sufficient for predicting anxious experience, which aligns with predictive modeling results for other subjective emotional experiences^{21,52} and argues against the traditional structure-centric view^{70,71} but rather aligns with a constructionist perspective^{48,49}. The core system is partly consistent with recent meta-analyses on induced anxiety-associated brain activity³³ and structural alterations in anxiety-related disorders⁸². The aINS, thalamus, and aMCC are key regions of the cingulo-opercular network^{83,84}, which partly resembles the salience network⁸⁵. These regions in concert not only support the detection of threatening signals⁸⁴ but also interoceptive processes, emotional awareness of negative affect, and autonomic activity⁸⁶⁻⁹⁰. The dlPFC, IPL, and precuneus constitute core nodes of the fronto-parietal network which exerts top-down feedback control and is involved in constructing conscious experiences and self-related processes^{18,91}. These regions together promoted the subjective experience of anxious arousal on both the population and individual level, which may suggest that the conscious experience of uncertain threat anticipation is a constructed state that encompasses different functional modules^{48,49}.

The aINS has long been suggested to represent subjective feelings from the body and emotional awareness across emotional domains including anger, fear, sadness, happiness, disgust, and aversion⁸⁶. Moreover, the aINS is involved in uncertainty and risk processing besides its established role in the perception of interoceptive and subjective feeling states^{2,92}. For example, the aINS plays a critical role in the anticipation of uncertain threats relative to certain threats^{45,93,94}. Bilateral aINS activity has also been associated with risk prediction in reward anticipation under uncertainty⁹⁵. The aINS exhibits reciprocal connections with the aMCC^{96,97}, a region proposed to integrate information about punishment, interoceptive, and subjective emotional states to exert control in the face of uncertainty⁸⁹. Our results therefore support the critical roles of aINS and aMCC in detecting, interpreting, and reacting to salient internal and environmental changes in the face of uncertainty which is central to anxiety.

The thalamus consists of different nuclei that serve various functions including relaying sensory and motor signals, memory, and attention as well as the regulation of consciousness, alertness, and emotion^{98,99}. The present study robustly identified intralaminar and mediodorsal thalamic nuclei as critical modules predictive of subjective anxious arousal on both the population and individual level, which may be related to their function in the integration of information across multiple cortical circuits that influences the conscious experience of anxious arousal during uncertain anticipation of the future¹⁰⁰.

Importantly, although the aINS, thalamus and aMCC are key nodes of the salience network^{84,85}, our control analyses suggested that the SUITAS did not respond to salience or unspecific negative emotional and autonomic arousal alone (Study 8 and 9). Moreover, none of these regions or other networks was sufficiently predictive of subjective anxious arousal, suggesting that the subjective experience during uncertain threat anticipation requires a distributed and brain-wide engagement to support the idiosyncratic and complex emotional experience. Notably, although the BNST has been suggested to play a prominent role in animal and human models of anxiety^{14,30,101}, we did not find BNST in the signature predictive of anxious anticipation which may suggest that BNST representation did not code for subjective experience despite its role in automatic processing of uncertain threats³⁰⁻³².

An important contribution of the current study is the comparison of the neural representation of subjective experience of uncertain threat anticipation with that of subjective fear (Study 10) and negative affect (Study 11). The conceptual and neural differentiation of fear

(acute threat), anxiety (potential threat) and nonspecific negative affect has long been debated in neurobiological models of emotion^{17,21,30,33,82,102}, animal models^{11,15,103,104} and neuropsychiatric models such as the RDoC framework^{6,7}. Rodent models suggest that defensive and physiological responses related to fear or anxiety are mediated by distinct neural substrates, i.e., the central nucleus of the amygdala and BNST, respectively¹¹. In contrast, human fMRI findings on functional dissociations between fear responses and uncertain threat anticipation have been inconsistent^{29,30,33,36}. Recent evidence from a meta-analysis of human fMRI studies suggests that ‘fear’ (conditioned versus unconditioned stimulus) and ‘anxiety’ (uncertain threat versus safe anticipation) may engage a partly overlapping circuit including BNST, MCC, aINS, and PAG^{33,105}. However, the mass univariate and categorical contrast approach does not permit to clearly segregate the neural systems underlying different mental processes such that common neural circuits may alternatively reflect hard-wired defensive behaviors or physiological responses as well as the generally increased arousal that characterize both fear and anxiety. The present study demonstrates fine-grained distributed neural patterns can, to a certain extent, segregate subjective experiences of uncertain threat anticipation, fear exposure and negative affect functionally and spatially such that the classic ‘emotional’ regions showed different levels of involvement in representing different emotion domains (see Fig. 5c), which suggests that these regions may to a certain extent encode fear anxiety and negative emotions in distinguishable neural representations.

Several important caveats should be taken into account when interpreting the results of the present study. First, the developed model was based on uncertain anticipation of electric shocks, and the generalization datasets also used shocks as aversive stimuli. Including datasets with other types of uncertain threats would allow for a more robust test of the generalizability of our brain model. Second, the parameter settings (C and epsilon) of the SVR were based on previous studies with a similar purpose. However, a grid search may help to determine the optimal combination of hyperparameters that allows to yield the highest performance in future studies. Moreover, there might be differences in the task design (e.g., with rating or not) and experimental parameters (e.g., stimulus type, durations) between our original datasets and the external datasets for testing the specificity of the SUITAS. Although we employed control experiments and analyses to validate that the results were not explained by these potential confounding factors, datasets with comparable task design and stimulus type are needed in future studies for the specificity test. Lastly, the sample size for the training dataset was based on previous studies⁷⁹. While researchers recently proposed that testing the generalizability of a model in novel samples could provide evidence for robust biomarkers without large sample sizes¹⁰⁶, research on a priori power and sample size estimations for multivariate neuroimaging models in task-fMRI will critically advance the field.

Together, the current study provided a comprehensive brain-level description of the subjective experience of threat anticipation under uncertainty. The sensitive, generalizable and specific signature has the potential to be tested prospectively in future studies with different experimental settings and populations (e.g., patients).

Methods

Participants

A total of 124 healthy, right-handed, young Asian (Chinese) university students (63 females, based on self-reported gender identity; age range: 18–33; Supplementary Table 1) participated Studies 1–3 to develop (Study 1, $n = 44$, 22 females), validate (Study 2, $n = 30$, 14 females) and prospectively test the generalizability (Study 3, $n = 50$, 27 females) of a multivoxel-pattern based predictive model of subjective experience of uncertain threat anticipation. All studies were approved by the local ethics committee at the University of Electronic Science

and Technology of China and were in accordance with the latest revision of the Declaration of Helsinki. All participants provided written informed consent on which Ethnicity was assessed by writing prior to study procedures. Participants were remunerated 120 RMB for their participation. Data from Study 1 and Study 2 were acquired in June 2021 (T1) using the same experimental design and scanning parameters at the same study site, while Study 3 was conducted 10 months later (April 2022, T2) with a slightly different experimental design (e.g., more experimental trials, different threat probability and uncertainty baseline, details see Paradigm and procedures) and was a part of our ongoing project examining cue-based anticipation of multimodal affective input. We randomly selected 60% of the participants from T1 as the training dataset to develop the brain model (Study 1: $n = 44$, 23 females, mean \pm SD age = 22.07 ± 2.50 years). The remaining participants comprised the validation dataset (Study 2: $n = 30$, 14 females, mean \pm SD age = 22.47 ± 2.76 years). Data from T2 were used as the prospective generalization dataset (Study 3: $n = 50$, 27 females, mean \pm SD age = 20.08 ± 2.22 years) to test if the brain model identified in Study 1 can be prospectively applied to independent data and generalized to different shock probability and uncertainty environments. Participants were excluded if they reported a current or history of neurological, psychiatric, or major physical disease, psychotropic medication, substance abuse, MRI contraindications and any prior participation in experiments with electric stimulation. No gender-based analysis was performed because this work did not have a prior hypothesis related to gender differences.

Paradigm and procedures

In Study 1 and Study 2, participants completed an uncertainty-variation threat anticipation (UVTA) task while undergoing fMRI acquisition. Before the start of the actual experiment, participants were instructed that they would anticipate electric shocks with varying levels of uncertainty and provide ratings of the anxious levels they experienced during the anticipation period at the end of each trial. Prior to MRI data acquisition, a shock calibration was implemented during which a highly aversive but not unbearable shock level was determined for each participant. Participants rated subjective shock discomfort on a scale from 1 (not at all painful) to 5 (painful and difficult to tolerate) to reach a level of 4 (painful but not unbearable). The shock calibration procedure was repeated in the middle of the four experimental runs to avoid habituation. 75% of participants in Study 1 and 70% of participants in Study 2 adjusted the level of electric stimulation during the recalibration. Electric shocks were delivered to the underside of the left wrist using a Biopac STM100C (Biopac Systems Inc., Goleta, CA). The UVTA paradigm used an event-related design consisting of four conditions repeatedly presented over four runs in a pseudorandom order with no more than two consecutive trials of the same condition. The four conditions corresponded to four different uncertainty levels from certain safety, low, medium to high threat uncertainty modulated by different combinations of event (shock), temporal (duration), and number of shock uncertainty. Low, medium and high uncertainty conditions were cued by a colored lightning bolt (blue, purple and red, respectively) in the center of the screen and a border of the same color whereas the certain safety condition was indicated by a white isosceles triangle and a white border (Fig. 1a). Prior to the task, participants were told about the contingency between cues and outcomes, e.g. that the blue cue would last 8 s and two consecutive shocks might occur immediately after the cue presentation (low threat uncertainty: event-only uncertainty); the purple cue would disappear any time between 0 to 16 s and two consecutive shocks might occur immediately after the cue presentation (medium threat uncertainty: event and temporal uncertainty); the red cue would disappear any time between 0 to 16 s and two consecutive shocks might occur or three shocks might occur either in a row or separated into two consecutive shocks immediately after the cue presentation and one shock after the subsequent fixation

epoch (high threat uncertainty: event, temporal and number uncertainty); no shocks would be administered after the white cue lasting between 0 to 16 s (certain safety). The exact shock probability for low, medium, and high uncertainty conditions was not stated to the participants and they were only told that the likelihood of shock outcomes of these three conditions was equivalent. The shock probability was predetermined to 60% for each uncertain threat condition. The purple, red, and white cues in actuality ranged between 6 and 10 s (avg 8 s) and both medium and high uncertainty conditions contained 'fast' trials with 3 s of cue presentation and 'slow' trials with 13 s of cue presentation pseudo-randomly to increase the credibility of the instruction that the cues could disappear any time within 16 s after which shocks might occur. Each run contained 5 valid trials per condition and 4 dummy trials (2 fast trials and 2 slow trials, shock probability: 50% throughout all runs), resulting in 24 trials in total in each run. The dummy trials were not included in the behavioral and fMRI analysis. Participants were asked to retrospectively rate their subjective level of anxious arousal during the anticipation period (6–10 s) from 1 (no anxiety) to 5 (extreme anxiety). Stimuli were presented via E-Prime 2.0 (Psychology Software Tools, Sharpsburg, PA). Note that the varying levels of uncertainty conditions were deployed to induce different levels of subjective anxious feelings and the behavioral rating patterns confirmed that our paradigm successfully evoked varying levels of anxiety between conditions (Fig. 1b and Supplementary Fig. 1).

Study 3 implemented a modified UVTA task. The differences were that the total number of trials and the proportion of uncertain threat trials increased (each run contained 29 trials, 8 trials per low, medium, and high uncertainty conditions, 4 trials per certain safety condition, and 1 'fast' trial), which resulted in increased statistical power and a different shock uncertainty baseline. The shock probability was predetermined to 50% for each uncertain threat condition and the 'fast' trial for each run belonged to a different condition and was always presented in the first trial with shocks. The shock calibration procedure was the same and 70% of the participants adjusted the level of electric stimulation in the middle of the experimental runs. Participants were asked to retrospectively rate their subjective level of anxious arousal during the anticipation period (6–10 s) from 1 (not anxious at all) to 5 (extremely anxious). A chi-square test confirmed no habituation effects of shock pain on anxious experience over the course of the paradigm ($\chi^2(3) = 16.21, P = 0.18$; Study 1–3).

MRI data were collected using a GE Discovery MR750 3.0 T system (General Electric Medical System, Milwaukee, WI, USA). Functional images were acquired with an interleaved T2*-weighted gradient echo-planar imaging sequences (40 slices; repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; slice thickness = 3.8 mm; spacing = 0.6 mm; field of view (FOV) = 200 × 200 mm; flip angle = 90°; matrix size = 64 × 64; voxel size = 3.125 × 3.125 × 3.8 mm). High-resolution whole-brain T1-weighted images were additionally acquired to improve spatial normalization (3D spoiled gradient echo pulse sequence; 154 slices; TR = 6 ms; TE = 3 ms; slice thickness = 1 mm, FOV = 256 × 256 mm, acquisition matrix = 256 × 256, flip angle = 8°, voxel size = 1 × 1 × 1 mm).

Skin conductance data was continuously acquired during fMRI scanning using an MRI-compatible Biopac system (MP-150). Skin conductance (1000 Hz) was sampled using MRI-compatible disposable, radiotranslucent, pre-gelled electrodes (EL508) attached to the index and middle fingers of the non-dominant (left) hand.

fMRI preprocessing and analysis

The fMRI data were preprocessed and analyzed using Statistical Parametric Mapping (SPM12, <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>) and the procedures for preprocessing and general linear model (GLM) analyses were the same for Study 1, 2 and 3. The first five volumes of each run were discarded to allow for magnetic field equilibration. Prior to preprocessing, image intensity outliers were identified using CanlabCore tools (<https://github.com/canlab/>

CanlabCore, for details, see ref. 21). Each time point identified as outlier was included in the first-level model as a separate nuisance covariate. The remaining functional images were correct for slice timing differences, head movements, co-registered with the T1-weighted structural images, normalized to Montreal Neurological Institute (MNI) standard template (interpolated to 2 × 2 × 2 mm voxel size), and spatially smoothed with an 8-mm full-width at half maximum Gaussian kernel (For details, see ref. 21).

Preprocessed images were subjected to a first-level GLM using SPM12 for prediction analysis. The four runs were temporally concatenated beforehand to ensure that there were enough trials for each rating. The model included five separate boxcar regressors time-locked to the anticipation period (6–10 s) corresponding to each rating (i.e., 1–5), which allowed us to model brain activity in response to each anxious level separately. Note that we only included the anticipation periods after which no shocks occurred to obviate any potential interference of the shock pain^{107–109}. Accordingly, the anticipation periods followed by shocks, the 'fast' and 'slow' trials as well as missed trials, were treated as regressors of no interest. The outcome period was also modeled with a boxcar regressor to examine the effects of shock delivery. To model any effects related to motor activity, one boxcar regressor indicating the rating period was included. The two fixation-cross epochs served as implicit baseline. All regressors of interest were convolved with a double gamma canonical hemodynamic response function (HRF). Additionally, 24 head motion parameters (6 realignment parameters demeaned, their derivatives and the squares of these 12 regressors), indicator vectors specifying 'spikes' by framewise displacement (FD) that had deviations larger than 0.50 mm¹¹⁰ and outlier time points identified from image intensity (see above for details) were treated as nuisance regressors.

Developing the brain model

Using the training dataset (Study 1, $n = 44$), we developed a whole brain neural signature (gray matter masked) predictive of subjective anxious experience (i.e., SUITAS) by training a support vector regression (SVR) model. We first concatenated each participant's data across runs and then used individual first-level GLM images (one per rating of no-shock trials for each participant) across participants as features to predict participants' ratings of the anticipation periods (Fig. 1c). The SVR was performed using a linear kernel function ($C = 1$) with epsilon setting to 0.1 in the Spider toolbox (<http://people.kyb.tuebingen.mpg.de/spider>) with individual activation maps (one per rating for each participant) as features to predict participants' anxious arousal ratings of the anticipation period while undergoing fMRI.

Model evaluation

To evaluate the model performance and minimize overfitting, we used 10 × 10-fold cross-validation within the training dataset in Study 1 and applied the brain model (i.e., SUITAS) to new individuals in Study 2 (Fig. 1c). The cross-validation procedures were repeated ten times by producing different splits (10 subsamples) in each repetition. Activation maps from 9 subsamples were trained to predict the anxiety ratings, and then the performance was tested on the holdout participants (1 subsample). To obtain unbiased estimates of the model performance, we predicted subjective ratings in new individuals from the validation dataset (Study 2, $n = 30$) by calculating the signature responses using a dot product of the SUITAS weight map with each vectorized activation map (one per rating). To provide an interpretable effect size metric, we calculated Pearson correlations between the predicted ratings and the actual ratings across participants for both Study 1 and Study 2, and the statistical inference was determined using a permutation test with 5000 random shuffles only for Study 2 given that cross-validated permutation test is very time-consuming. Explained variance score (EVS) was computed to indicate the overall

prediction error (for similar approaches see ref. 21). We additionally assessed the classification accuracy between high (average of rating 4 and 5) versus low (average of rating 1 and 2) levels of subjective anxious ratings using two-sided forced-choice tests from receiver operating characteristic curves.

Generalization of the SUITAS

We tested whether SUITAS could predict subjective ratings in new individuals from Study 3 ($n = 50$) who were not included in the initial model training and validation and underwent a modified UVTA (Fig. 1d). The independent dataset in Study 3 could provide a good test for the model generalizability and also allow us to obtain unbiased estimates of the model sensitivity. Similarly, pattern responses were estimated for each test participant by computing the dot product of the SUITAS pattern with the participant's vectorized activation map. The calculations for Pearson correlation, permutation test, EVS, and classification accuracy were the same as in Study 2. The statistical inference of Pearson correlation was determined using a permutation test with 5000 random shuffles. Moreover, we used two publicly available datasets (Study 4, $n = 59$ from our previous study using a visual threat-conditioning paradigm⁶¹; Study 5, $n = 68$ from a previous study using an auditory threat-conditioning paradigm⁶²; details see Supplementary Methods and Supplementary Table 1) to further test the generalizability of the SUITAS in distinguishing uncertain threat versus safe anticipation across cohorts, paradigms, MRI systems, and scanning parameters.

Within-individual prediction

To further determine the sensitivity of the population-level model (i.e., SUITAS) in tracking within-individual variation in anxious ratings, we estimated single-trial responses using a GLM design matrix with separate regressors for each trial (no shock) in Study 1, 2 and 3, resulting in -44 trials for each participant in the training and validation datasets and -64 trials for each participant in the prospective generalization dataset. The SUITAS was next applied to the single-trial activation maps to obtain the signature responses, which were finally correlated with the true ratings for each participant separately. Note that we used 10×10 -fold cross-validation to obtain less biased test results for each participant in Study 1, whereas the within-individual prediction in Study 2 and Study 3 did not use cross-validation procedures because they were not included in the training of the model. The correlation coefficients were first r to z transformed and then averaged to obtain one mean correlation value. Bootstrap tests (5000 iterations) were used to test whether the distributions of within-individual prediction-outcome correlations were significantly higher than zero.

Identifying a core system consistently involved in subjective experience of anxiety

To identify the important brain regions for the subjective experience of uncertain threat anticipation, we located the consistent voxels across participants that (1) reliably contributed to the model prediction (i.e., model weights) and (2) were associated with subjective anxious arousal (i.e., model encoding). Previous studies have suggested that both model weight ('betas') and model encoding maps ('structure coefficients') are necessary to interpret the model^{74,111}. We therefore obtained both model weight maps and model encoding maps at the individual level and then calculated their conjunction (i.e., significant voxels in both maps) to determine the core brain system for subjective experience of threat anticipation under uncertainty.

To do this, we first ran a separate prediction analysis (linear SVR with $C = 1$) for each participant in the training dataset (Study 1, $n = 44$) using their single-trial activation maps (-44 trials, 10×10 -fold cross-validated). A one-sample t -test (FDR $q < 0.05$, one-sided) across within-individual patterns was performed to evaluate the consistency of each weight for every voxel in the brain. Next, model encoding ('structure

coefficient') maps were computed for each participant by transforming the within-individual predictive pattern using the following formula: $A = \text{cov}(X) \times W \times \text{cov}(W^T \times X)^{-1}$, where A is the reconstructed activation pattern, $\text{cov}(X)$ is the covariance matrix of individual data, W is the SUITAS predictive weights, and $\text{cov}(W^T \times X)$ is the covariance matrix of the latent factors. Structure coefficients map individual voxels to the overall multivariate model prediction from backward model to forward model and the significant brain regions were determined by one-sample t -test (FDR $q < 0.05$, one-sided) on the within-individual encoding maps where the voxel-wise activity correlates with the model prediction. The conjunction (FDR $q < 0.05$, one-sided, preserving positive values) of the thresholded second-level within-individual model weight map (one-sample t -test) and model encoding map (one-sample t -test) was considered as the core system map. The regions were moreover mapped onto the results from a recent meta-analysis on uncertainty-induced threat anticipation that included categorical comparisons of fMRI activation between unpredictable threat and safe anticipation in healthy individuals³³.

Testing the specificity of the SUITAS against pain, anticipation, and arousal

To examine whether the SUITAS specifically captures anxious feelings, we evaluated the extent to which the SUITAS was sensitive to pain experience, general positive/negative anticipation and unspecific negative emotional arousal involved in the paradigm by testing the SUITAS on four independent datasets (Study 6, $n = 33$ from previous studies using a thermal pain paradigm^{53,112}; available at https://figshare.com/articles/dataset/bmrk3_6levels_pain_dataset_mat/6933119; Study 7, $n = 100$ from an ongoing study using a monetary gain/loss anticipation paradigm - MID task¹¹³, to measure reward processing, <https://zib.fudan.edu.cn>; Study 8, $n = 48$ from an ongoing study using an emotional picture paradigm and was a subsample from Study 3, Fig. 1d; Study 9, $n = 65$ from Study 2 and 3 in which the skin conductance data was recorded; see also Supplementary Methods and Supplementary Table 1 for details), respectively. For Study 6, we used the SUITAS to predict the pain ratings by computing the signature responses using a dot product of the SUITAS weight map with each vectorized activation map, and the statistical inference was determined using a permutation test with 5000 random shuffles. We further used the SUITAS to classify the pain stimulation periods corresponding to high (5 and 6) versus low (1 and 2) pain ratings after each trial. For Study 7, we applied the SUITAS to distinguish anticipation of monetary gain versus neutral condition and monetary loss versus neutral condition where signature responses were compared for two conditions. For Study 8, we applied the SUITAS to classify exposure towards pre-selected high-arousing disgust versus low-arousing neutral pictures that were presented in an event-related design during fMRI acquisition (details see Supplementary Methods). Classification accuracy was calculated from receiver operating characteristic curves using forced-choice classification and P values were calculated using two-sided independent binomial tests for Study 6-8. Moreover, we tested whether the SUITAS could specifically predict subjective anxious feelings rather than its concomitant skin conductance levels using subsamples of Study 2 ($n = 22$) and Study 3 ($n = 43$) who had complete skin conductance data (Study 9, details see Supplementary Methods and Supplementary Table 1).

Comparing the SUITAS with predictive models of subjective fear and negative affect

To test whether or to which extent The SUITAS was distinct from neural representations of fear exposure and negative affect, we compared the functional and spatial similarity of the SUITAS with the subjective fear signature VIFS from our previous study (Study 10, $n = 67$; details see ref. 21) and the subjective negative affect signature PINES from Chang et al. (Study 11, $n = 121$; details see ref. 52) (Fig. 1d, see also Supplementary Methods for details). Specifically, we (1)

applied the SUITAS to the datasets of VIFS and PINES, and vice versa (i.e., applied the VIFS or PINES to the datasets of SUITAS) to examine the prediction performance, and (2) compared the spatial topography of the SUITAS, VIFS and PINES patterns by calculating the spatial correlation among these predicted weights (unthresholded and thresholded at uncorrected $P < 0.001$) and computed cosine similarity between these signatures and predefined ROIs documented in previous studies as regions showing preferential activation to anxiety³³, fear⁶⁷ and negative affect⁶⁸. To exclude the visual processing effect, we retrained these models excluding the occipital lobe and compared the performance of SUITAS with VIFS and PINES (see Supplementary Results and Supplementary Fig. 9).

Models using features from local regions or networks

To examine whether the subjective anxious experience was reducible to activation patterns in any single brain region or network, we retrained SVR models (10 × 10-fold cross-validated) restricted to local brain regions (e.g., ACC, aINS, thalamus, BNST, etc.) and networks (e.g., seven large-scale resting-state function networks⁷³, salience network⁷² and consciousness network⁴⁸) which have been previously suggested to be associated with anxious anticipation. Anatomical parcellations and the corresponding atlases used to select ROIs are detailed in Supplementary Table 3.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that has been used to develop and evaluate the SUITAS are available on the Open Science Framework (<https://osf.io/a8gcb/>). The meta-analytic map comparing uncertain threat versus safe anticipation in healthy individuals³³ is available at <https://neurovault.org/images/384665/>. Source data are provided with this paper.

Code availability

The custom code that supports the findings of this study is available at https://github.com/lucyliu666/Anxiety_decoder and <https://osf.io/a8gcb/>, and the CanlabCore Tools are available at <https://github.com/canlab/CanlabCore>. Zenodo, <https://zenodo.org/records/10409502>, (2023)¹¹⁴.

References

- Carleton, R. N. The intolerance of uncertainty construct in the context of anxiety disorders: theoretical and practical perspectives. *Expert Rev. Neurother.* **12**, 937–947 (2012).
- Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).
- Mobbs, D., Headley, D. B., Ding, W. & Dayan, P. Space, time, and fear: survival computations along defensive circuits. *Trends Cogn. Sci.* **24**, 228–241 (2020).
- Nitschke, J. B. et al. Startle potentiation in aversive anticipation: evidence for state but not trait effects. *Psychophysiol.* **39**, 254–258 (2002).
- Grillon, C. et al. Increased anxiety during anticipation of unpredictable but not predictable aversive stimuli as a psychophysiological marker of panic disorder. *Am. J. Psychiatry* **165**, 898–904 (2008).
- Insel, T. et al. Research domain criteria (rdoc): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
- Kozak, M. J. & Cuthbert, B. N. The NIMH research domain criteria initiative: background, issues, and pragmatics. *Psychophysiol.* **53**, 286–297 (2016).
- Grillon, C., Robinson, O. J., Cornwell, B. & Ernst, M. Modeling anxiety in healthy humans: a key intermediate bridge between basic and clinical sciences. *Neuropsychopharmacol.* **44**, 1999–2010 (2019).
- Robinson, O. J., Vytal, K., Cornwell, B. R. & Grillon, C. The impact of anxiety upon cognition: perspectives from human threat of shock studies. *Front. Hum. Neurosci.* **7**, 203 (2013).
- Calhoun, G. G. & Tye, K. M. Resolving the neural circuits of anxiety. *Nat. Neurosci.* **18**, 1394–1404 (2015).
- Davis, M., Walker, D. L., Miles, L. & Grillon, C. Phasic vs sustained fear in rats and humans: Role of the extended amygdala in fear vs anxiety. *Neuropsychopharmacol.* **35**, 105–135 (2010).
- Hammack, S. E., Todd, T. P. & Margaret Kocho-Schellenberg, M. E. B. Role of the bed nucleus of the stria terminalis in the acquisition of contextual fear at long or short context-shock intervals. *Behav. Neurosci.* **129**, 673–678 (2015).
- Robinson, O. J., Pike, A. C., Cornwell, B. & Grillon, C. The translational neural circuitry of anxiety supplements. *J. Neurol. Neurosurg. Psychiatry* **90**, 1353–1360 (2019).
- Shackman, A. J. & Fox, A. S. Contributions of the central extended amygdala to fear and anxiety. *J. Neurosci.* **36**, 8050–8063 (2016).
- Chen, Y. H. et al. Distinct projections from the infralimbic cortex exert opposing effects in modulating anxiety and fear. *J. Clin. Invest.* **131**, e145692 (2021).
- Anderson, D. J. & Adolphs, R. A framework for studying emotions across species. *Cell* **157**, 187–200 (2014).
- LeDoux, J. E. *Anxious: Using the Brain to Understand and Treat Fear and Anxiety*. (Viking, New York, 2015).
- LeDoux, J. E. & Pine, D. S. Using neuroscience to help understand fear and anxiety: A two-system framework. *Am. J. Psychiatry* **173**, 1083–1093 (2016).
- Ledoux, J. E. & Brown, R. A higher-order theory of emotional consciousness. *Proc. Natl Acad. Sci.* **114**, E2016–E2025 (2017).
- Barrett, L. F. et al. Of Mice and Men: Natural Kinds of Emotions in the Mammalian Brain? A Response to Panksepp and Izard. *Perspect Psychol Sci.* **2**, 297–312 (2016).
- Zhou, F. et al. A distributed fMRI-based signature for the subjective experience of fear. *Nat. Commun.* **12**, 1–16 (2021).
- Taschereau-Dumouchel, V., Michel, M., Lau, H., Hofmann, S. G. & LeDoux, J. E. Putting the “mental” back in “mental disorders”: a perspective from research on fear and anxiety. *Mol. Psychiatry* **27**, 1322–1330 (2022).
- Baldwin, D. S., Huusom, A. K. T. & Mæhlum, E. Escitalopram and paroxetine in the treatment of generalised anxiety disorder: Randomised, placebo-controlled, double-blind study. *Br. J. Psychiatry* **189**, 264–272 (2006).
- Griebel, G. & Holmes, A. 50 years of hurdles and hope in anxiolytic drug discovery. *Nat. Rev. Drug Discov.* **12**, 667–687 (2013).
- Yao, D., Qin, Y. & Zhang, Y. From psychosomatic medicine, brain-computer interface to brain-apparatus communication. *Brain-Appar. Commun. A J. Bacomics* **1**, 66–88 (2022).
- Afsaneh, E. & Zarei Ghobadi, M. The role of neuromodulation-related technologies in neurology for the next 10 years. *Brain-Apparatus Commun. A J. Bacomics* **2**, 2147405 (2023).
- Alvarez, R. P., Chen, G., Bodurka, J., Kaplan, R. & Grillon, C. Phasic and sustained fear in humans elicits distinct patterns of brain activity. *Neuroimage* **55**, 389–400 (2011).
- Alvarez, R. P. et al. Increased anterior insula activity in anxious individuals is linked to diminished perceived control. *Transl. Psychiatry* **5**, e591–e591 (2015).
- Grupe, D. W., Oathes, D. J. & Nitschke, J. B. Dissecting the anticipation of aversion reveals dissociable neural networks. *Cereb. Cortex* **23**, 1874–1883 (2013).
- Hur, J. et al. Anxiety and the neurobiology of temporally uncertain threat anticipation. *J. Neurosci.* **40**, 7949–7964 (2020).

31. Klumbers, F., Kroes, M. C. W., Baas, J. M. P. & Fernández, G. How human amygdala and bed nucleus of the stria terminalis may drive distinct defensive responses. *J. Neurosci.* **37**, 9645–9656 (2017).
32. Somerville, L. H. et al. Interactions between transient and sustained neural signals support the generation and regulation of anxious emotion. *Cereb. Cortex* **23**, 49–60 (2013).
33. Chavanne, A. V. & Robinson, O. J. The overlapping neurobiology of induced and pathological anxiety: a meta-analysis of functional neural activation. *Am. J. Psychiatry* **178**, 156–164 (2021).
34. Herrmann, M. J. et al. Phasic and sustained brain responses in the amygdala and the bed nucleus of the stria terminalis during threat anticipation. *Hum. Brain Mapp.* **37**, 1091–1102 (2016).
35. Pedersen, W. S., Muftuler, L. T. & Larson, C. L. A high-resolution fMRI investigation of BNST and centromedial amygdala activity as a function of affective stimulus predictability, anticipation, and duration. *Soc. Cogn. Affect. Neurosci.* **14**, 1167–1177 (2019).
36. Mobbs, D. et al. From threat to fear: the neural organization of defensive fear systems in humans. *J. Neurosci.* **29**, 12236–12243 (2009).
37. Nitschke, J. B., Sarinopoulos, I., MacKiewicz, K. L., Schaefer, H. S. & Davidson, R. J. Functional neuroanatomy of aversion and its anticipation. *Neuroimage* **29**, 106–116 (2006).
38. Xin, F. et al. Oxytocin differentially modulates amygdala responses during top-down and bottom-up aversive anticipation. *Adv. Sci.* **7**, 1–14 (2020).
39. Weis, C. N. et al. A 7-Tesla MRI study of the periaqueductal gray: Resting state and task activation under threat. *Soc. Cogn. Affect. Neurosci.* **17**, 187–197 (2022).
40. Drabant, E. M. et al. Experiential, autonomic, and neural responses during threat anticipation vary as a function of threat intensity and neuroticism. *Neuroimage* **55**, 401–410 (2011).
41. Holtz, K., Pané-Farré, C. A., Wendt, J., Lotze, M. & Hamm, A. O. Brain activation during anticipation of interoceptive threat. *Neuroimage* **61**, 857–865 (2012).
42. Reicherts, P. et al. Anxious anticipation and pain: the influence of instructed vs conditioned threat on pain. *Soc. Cogn. Affect. Neurosci.* **12**, 544–554 (2017).
43. Straube, T., Schmidt, S., Weiss, T., Mentzel, H. J. & Miltner, W. H. R. Dynamic activation of the anterior cingulate cortex during anticipatory anxiety. *Neuroimage* **44**, 975–981 (2009).
44. Balderston, N. L., Hsiung, A., Ernst, M. & Grillon, C. Effect of threat on right dlPFC activity during behavioral pattern separation. *J. Neurosci.* **37**, 9160–9171 (2017).
45. Geng, H. et al. Altered brain activation and connectivity during anticipation of uncertain threat in trait anxiety. *Hum. Brain Mapp.* **39**, 3898–3914 (2018).
46. Kirlic, N., Aupperle, R. L., Misaki, M., Kuplicki, R. & Alvarez, R. P. Recruitment of orbitofrontal cortex during unpredictable threat among adults at risk for affective disorders. *Brain Behav.* **7**, 1–11 (2017).
47. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**, 1437–1448 (2014).
48. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. The brain basis of emotion: A meta-analytic review. *Behav. Brain Sci.* **35**, 121–143 (2012).
49. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **12**, 1–23 (2017).
50. Carlson, J. M., Greenberg, T., Rubin, D. & Mujica-Parodi, L. R. Feeling anxious: anticipatory amygdalo-insular response predicts the feeling of anxious anticipation. *Soc. Cogn. Affect. Neurosci.* **6**, 74–81 (2011).
51. Woo, C. W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: Brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377 (2017).
52. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol.* **13**, 1–28 (2015).
53. Wager, T. D. et al. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
54. Haxby, J. V. Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage* **62**, 852–855 (2012).
55. Formisano, E. & Kriegeskorte, N. Seeing patterns through the hemodynamic veil - The future of pattern-information fMRI. *Neuroimage* **62**, 1249–1256 (2012).
56. Peelen, M. V. & Downing, P. E. Testing cognitive theories using multivariate pattern analysis of neuroimaging data. *Nat. Hum. Behav.* **7**, 1430–1441 (2023).
57. Kragel, P. A., Koban, L., Barrett, L. F. & Wager, T. D. Representation, pattern information, and brain signatures: from neurons to neuroimaging. *Neuron* **99**, 257–273 (2018).
58. Kragel, P. A., Treadway, M. T., Admon, R., Pizzagalli, D. A. & Hahn, E. C. A mesocorticolimbic signature of pleasure in the human brain. *Nat. Hum. Behav.* **7**, 1332–1343 (2023).
59. Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. *Manual for the state-trait anxiety inventory (Form Y)*. (Consulting Psychologists Press., 1983).
60. Carleton, R. N., Norton, M. A. P. J. & Asmundson, G. J. G. Fearing the unknown: a short version of the intolerance of uncertainty scale. *J. Anxiety Disord.* **21**, 105–117 (2007).
61. Zhou, F. et al. Human extinction learning is accelerated by an angiotensin antagonist via ventromedial prefrontal cortex and its connections with basolateral amygdala. *Biol. Psychiatry* **86**, 910–920 (2019).
62. Reddan, M. C., Wager, T. D. & Schiller, D. Attenuating neural threat expression with imagination. *Neuron* **100**, 994–1005 (2018).
63. Morriss, J., Gell, M. & van Reekum, C. M. The uncertain brain: a coordinate based meta-analysis of the neural signatures supporting uncertainty during different contexts. *Neurosci. Biobehav. Rev.* **96**, 241–249 (2019).
64. Casey, B. J. et al. The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
65. Critchley, H. D., Mathias, C. J. & Dolan, R. J. Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron* **29**, 537–545 (2001).
66. Taschereau-Dumouchel, V., Kawato, M. & Lau, H. Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Mol. Psychiatry* **25**, 2342–2354 (2020).
67. Fullana, M. A. et al. Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Mol. Psychiatry* **21**, 500–508 (2016).
68. Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J. & Barrett, L. F. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cereb. Cortex* **26**, 1910–1922 (2016).
69. Davis, M. The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* **15**, 353–375 (1992).
70. LeDoux, J. E. Fear and the brain: Where have we been, and where are we going? *Biol. Psychiatry* **44**, 1229–1238 (1998).
71. PD, I. M. Psychosomatic disease and the “visceral brain” recent developments bearing on the papez theory of emotion. *Psychosom. Med.* **11**, 338–353 (1949).
72. Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V. & Greicius, M. D. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb. Cortex* **22**, 158–165 (2012).

73. Thomas Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
74. Ćeko, M., Kragel, P. A., Woo, C., López-solà, M. & Wager, T. D. Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* **25**, 760–770 (2022).
75. Pessoa, L. How many brain regions are needed to elucidate the neural bases of fear and anxiety? *Neurosci. Biobehav. Rev.* 105039 (2023).
76. Zhang, R. et al. Angiotensin II regulates the neural expression of subjective fear in humans - precision pharmaco-neuroimaging approach. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **8**, 262–270 (2023).
77. Foa, E. B., Zinbarg, R. & Rothbaum, B. O. Uncontrollability and unpredictability in post-traumatic stress disorder: an animal model. *Psychol. Bull.* **112**, 218 (1992).
78. Grillon, C., Baas, J. P., Lissek, S., Smith, K. & Milstein, J. Anxious responses to predictable and unpredictable aversive events. *Behav. Neurosci.* **118**, 916–924 (2004).
79. Lee, J. J. et al. A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* **27**, 174–182 (2021).
80. Woo, C. W. et al. Quantifying cerebral contributions to pain beyond nociception. *Nat. Commun.* **8**, 14211 (2017).
81. Scheinost, D. et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage*. **193**, 35–45 (2020).
82. Liu, X., Klugah-brown, B., Zhang, R. & Chen, H. Pathological fear, anxiety and negative affect exhibit distinct neurostructural signatures: evidence from psychiatric neuroimaging meta-analysis. *Transl. Psychiatry* **12**, 1–19 (2022).
83. Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L. & Petersen, S. E. A dual-networks architecture of top-down control. *Trends Cogn. Sci.* **12**, 99–105 (2008).
84. Sylvester, C. M. et al. Functional network dysfunction in anxiety and anxiety disorders. *Trends Neurosci.* **35**, 527–535 (2012).
85. Seeley, W. W. et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* **27**, 2349–2356 (2007).
86. Craig, A. D. How do you feel — now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* **10**, 59–70 (2009).
87. Yao, S. et al. Oxytocin facilitates approach behavior to positive social stimuli via decreasing anterior insula activity. *Int. J. Neuropsychopharmacol.* **21**, 918–925 (2018).
88. Ferraro, S. et al. The central autonomic system revisited – convergent evidence for a regulatory role of the insular and mid-cingulate cortex from neuroimaging meta-analyses. *Neurosci. Biobehav. Rev.* **142**, 104915 (2022).
89. Shackman, A. J. et al. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* **12**, 154–167 (2011).
90. Wager, T. D. & Barrett, L. F. From affect to control: Functional specialization of the insula in motivation and regulation. Preprint at *bioRxiv* 102368. <https://doi.org/10.1101/102368> (2017).
91. Wang, L. et al. Fear of missing out (FOMO) associates with reduced cortical thickness in core regions of the posterior default mode network and higher levels of problematic smartphone and social media use. *Addict. Behav.* **143**, 107709 (2023).
92. Singer, T., Critchley, H. D. & Preuschoff, K. A common role of insula in feelings, empathy and uncertainty. *Trends Cogn. Sci.* **13**, 334–340 (2009).
93. Gorka, S. M., Kreutzer, K. A., Petrey, K. M., Radoman, M. & Phan, K. L. Behavioral and neural sensitivity to uncertain threat in individuals with alcohol use disorder: Associations with drinking behaviors and motives. *Addict. Biol.* **25**, e12774 (2020).
94. Shankman, S. A. et al. Anterior insula responds to temporally unpredictable aversiveness: An fMRI study. *Neuroreport* **25**, 596–600 (2014).
95. Preuschoff, K., Quartz, S. R. & Bossaerts, P. Human insula activation reflects risk prediction errors as well as risk. *J. Neurosci.* **28**, 2745–2752 (2008).
96. Cauda, F. et al. Functional connectivity of the insula in the resting brain. *Neuroimage* **55**, 8–23 (2011).
97. Moisset, X. et al. Anatomical connections between brain areas activated during rectal distension in healthy volunteers: A visceral pain network. *Eur. J. Pain.* **14**, 142–148 (2010).
98. Roy, D. S., Zhang, Y., Halassa, M. M. & Feng, G. Thalamic subnetworks as units of function. *Nat. Neurosci.* **25**, 140–153 (2022).
99. Torrico, T. J. & Munakomi, S. Neuroanatomy, Thalamus. In *StatPearls*. (2022).
100. Saalman, Y. B. Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition. *Front. Syst. Neurosci.* **8**, 1–8 (2014).
101. Avery, S. N., Clauss, J. A. & Blackford, J. U. The human BNST: functional role in anxiety and addiction. *Neuropsychopharmacol.* **41**, 126–141 (2015).
102. Mobbs, D. et al. Viewpoints: approaches to defining and investigating fear. *Nat. Neurosci.* **22**, 1205–1216 (2019).
103. Duvarci, S., Bauer, E. P. & Paré, D. The bed nucleus of the stria terminalis mediates inter-individual variations in anxiety and fear. *J. Neurosci.* **29**, 10357–10361 (2009).
104. Heinz, D. E. et al. Exploratory drive, fear, and anxiety are dissociable and independent components in foraging mice. *Transl. Psychiatry* **11**, 318 (2021).
105. Shackman, A. J. & Fox, A. S. Two decades of anxiety neuroimaging research: New insights and a look to the future. *Am. J. Psychiatry* **178**, 106–109 (2021).
106. Rosenberg, M. D. & Finn, E. S. How to establish robust brain-behavior relationships without thousands of individuals. *Nat. Neurosci.* **25**, 835–837 (2022).
107. Fung, B. J., Qi, S., Hassabis, D., Daw, N. & Mobbs, D. Slow escape decisions are swayed by trait anxiety. *Nat. Hum. Behav.* **3**, 702–708 (2019).
108. Qi, S., Footer, O., Camerer, C. F. & Mobbs, D. A collaborator’s reputation can bias decisions and anxiety under uncertainty. *J. Neurosci.* **38**, 2262–2269 (2018).
109. Qi, S. et al. How cognitive and reactive fear circuits optimize escape decisions in humans. *Proc. Natl Acad. Sci. USA* **115**, 3186–3191 (2018).
110. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
111. Courville, T. & Thompson, B. Use of structure coefficients in published multiple regression articles: β is not enough. *Educ. Psychol. Meas.* **61**, 229–248 (2001).
112. Woo, C. W., Roy, M., Buhle, J. T. & Wager, T. D. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biol.* **13**, e1002036 (2015).
113. Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L. & Hommer, D. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* **12**, 3683–3687 (2001).
114. Liu, X. A neural signature for the subjective experience of threat anticipation under uncertainty. *Zenodo*, <https://zenodo.org/records/10409502>. (2023).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC, Grant No. 82271583, 32250610208), the Ministry of Science and Technology (MOST2030, Grant No. 2022ZD0208500) and the

National Key Research and Development Program of China (Grant No. 2018YFA0701400) to B.B. The funders had no role in study design, data collection, and analysis, decision to publish or preparation of the manuscript. Any opinions, findings, conclusions or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission.

Author contributions

X.L. and B.B. conceived and designed the study. X.L. and G.J. collected the data. X.L. analyzed the data. X.L. and B.B. drafted the manuscript. F.Z., K.K. D.Y., Q.G., S.X., T.J., X.Z., J.Z. and J.F. contributed data and important intellectual input to the manuscript. All of the authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-45433-6>.

Correspondence and requests for materials should be addressed to Benjamin Becker.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024