Rotten to the core – a neurofunctional signature of subjective core disgust generalizes to socio-moral contexts

Xianyang Gan[1,2,8], Feng Zhou[3,8], Ting Xu[1,2], Xiaobo Liu[4], Ran Zhang[1,2], Zihao Zheng[1,2], Xi Yang[5], Xinqi Zhou[6], Fangwen Yu[1,2], Jialin Li[7], Ruifang Cui[1,2], Lan Wang[1,2], Jiajin Yuan[6], Dezhong Yao[1,2], and Benjamin Becker[1,2,*]

[1]The Center of Psychosomatic Medicine, Sichuan Provincial Center for Mental Health, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 610054, China

[2]Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China

[3]Faculty of Psychology, Southwest University, Chongqing 400715, China

[4]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal H2X0A8, Canada

[5]Faculty of Health, Medicine and Life Sciences, Maastricht University, 6226NB, Netherlands

[6]Institute of Brain and Psychological Sciences, Sichuan Normal University, Chengdu 610066, China

[7]Max Planck School of Cognition, Leipzig 04103, Germany

[8]These authors contributed equally

*Correspondence: ben_becker@gmx.de (B.B.)

Abstract

While disgust originates in the hard-wired mammalian distaste response, the conscious experience of disgust in humans strongly depends on subjective appraisal and may even extend to sociomoral contexts. In a series of studies, we combined functional magnetic resonance imaging (fMRI) with machine-learning based predictive modeling to establish a comprehensive neurobiological model of subjective disgust. The neurofunctional signature accurately predicted momentary self-reported subjective disgust across discovery (*n*=78) and pre-registered validation (*n*=30) cohorts and generalized across core disgust (*n*=26) and sociomoral (unfair offers; *n*=43) contexts. Disgust experience was encoded in distributed cortical and subcortical systems, and exhibited distinct and shared neural representations with subjective fear or negative affect in interoceptive-emotional awareness and conscious appraisal systems while the signatures most accurately predicted the respective target experience. We provide an accurate fMRI-signature for disgust with a high potential to resolve ongoing evolutionary and clinical debates.

Keywords

Disgust; revulsion; unfairness; moral; neural decoding; fMRI; Multivariate pattern analysis; emotion; subjective; neural signature

INTRODUCTION

Across cultures, disgust is commonly triggered by rotting food, maggots, and scavenging animals but also by highly culture- and individual-specific triggers such as specific foods or behaviors that are appraised as morally reprehensible. Disgust has been conceptualized as a multifaceted defensive-avoidance response that encompasses hard-wired physiological and motor responses such as retching or a distinct facial expression and a strongly aversive feeling of revulsion that together facilitate the avoidance of potentially poisonous, contagious, or morally offensive situations and individuals. From an evolutionary perspective, disgust originates in the mammalian bitter taste (distaste) rejection system aiming to protect against the ingestion of potentially toxic food[1-3]. In response to bitter food, the distaste mechanism directly triggers defensive reactions such as withdrawal and retching and distinct facial expressions across (some) mammalian species via brainstem circuits[4] or the posterior insular cortex[5], respectively (disgust output system[1], see also Figure 1A). In humans, the distaste response was preadapted during evolution into core disgust which is essentially characterized by the highly aversive feeling of revulsion in response to contaminated, infectious, or poisonous stimuli[6-9]. Core disgust is distinct from the distaste reaction given that the experience of revulsion can be evoked in the absence of actual sensory stimuli like bitter taste and is strongly shaped by conscious appraisals (e.g., food being rejected based on where it might have been)[9]. Subsequently, core disgust may have been preadapted into the social context with revulsion being elicited by sociomoral transgressions[1,2,10] which may facilitate avoidance of potentially harmful interactions (sociomoral disgust)[11].

While several neurobiological models on the exact brain mechanisms of the distaste reaction have been established in animal models[4,12,13], recent conceptual and computational advances indicate that the neural mechanisms that underlie conscious emotional experiences are distinguishable from those that mediate the hard-wired defensive and social-communicative facets of the response[7,14-18]. While these studies primarily focused on fear, the neural mechanisms underlying the conscious experience of disgust, whether these mechanisms differ from the experience of evolutionary similar defensive functions (e.g., fear), and whether the same mechanisms mediate the aversive experience in sociomoral transgressions remain fiercely debated.

A plethora of human studies employed functional Magnetic Resonance Imaging (fMRI) to determine the neural basis of disgust, yet findings with respect to the neural systems that mediate conscious disgust experience remained highly inconsistent. For example, the insula has been traditionally considered as a key neural substrate of disgust[6,7,19,20], however several studies did not observe insula engagement in response to disgust stimuli[21-25] and recent evidence suggests a broader role of the insula in interoceptive[26,27] and central autonomic processes[28]. Subcortical regions like the amygdala and basal ganglia have also been considered as key disgust regions[7,19,29], probably due to their roles in early threat detection and defensive motor responses[7,14,15]. However, many neuroimaging studies failed to observe an involvement of amygdala[23-25,30-32] while a recent meta-analysis did not confirm robust engagement of the basal ganglia[7] during disgust processing. These inconsistencies may arise via a combination of (1) experimental designs that do not account for individual differences in disgust appraisal and experience[33] and the employed categorical analytic approach of comparing highly disgusting versus neutral stimuli which inherently also captures neural activity associated with the defensive-avoidance response, salience or arousal[34], and, (2) methodological limitations of the conventional mass-univariate fMRI approach with respect to establishing accurate, comprehensive and process-specific neural models for mental

processes[35,36].

Recent methodological progress combined fMRI with machine-learning based multivariate pattern analysis (MVPA) to develop more comprehensive and precise neurofunctional models of internal mental processes, including sensitive and generalizable neuroaffective signatures for subjective emotional experiences (e.g., negative affect or fear)[15,34,37,38]. Identifying multivariate brain signatures predictive of the momentary emotional experience may allow to overcome critical limitations of the mass-univariate fMRI approach while accounting for contextual and individual variations on the appraisal and experience level[39]. Briefly, MVPA examines neural responses across multiple brain regions and voxels[35,40-42] thus allowing to determine neural representations of mental processes with higher resolution and considerably larger effect sizes[35,43] than the conventional mass-univariate approach. The resultant fMRI-based neurofunctional signatures enabled for the first time to separate the subjective emotional experience from hard-wired defensive and physiological responses[15,17] and to explore common and separable neural representations of aversive experiences across domains[38,44]. In line with recent conceptual frameworks, these findings indicate that the concerted interplay between subcortical and cortical systems mediates the appraisal and construction of conscious emotional experiences[45,46]. Despite long-standing discussions on the neurobiological and cultural evolution of disgust and increasing interest in disgust as a transdiagnostic domain involved in the onset and maintenance of prevalent clinical disorders (e.g., obsessive-compulsive disorder, eating disorders, and Huntington's disease)[6,8,29], and initial studies demonstrated the promise of MVPA for regional or categorical disgust decoding[47,48], a precise and generalizable whole-brain signature predictive of momentary self-reported subjective disgust experience is currently lacking.

Here, we capitalized on recent methodological advances in MVPA-based neural decoding techniques to determine whether (1) it is possible to develop a neural signature that predicts momentary subjective disgust experience elicited by core disgust stimuli on the population-level with high sensitivity and robustness across a discovery (*n*=78 healthy participants) and a pre-registered validation (*n*=30 participants) cohort and that generalizes to data from an independent research team using a different disgust induction paradigm and MRI system (*n*=26 participants)[49] (Figures 1B and 1C); (2) this neural signature can sensitively track variations in momentary disgust experience on the individual-level in terms of predicting trial-wise disgust experience in discovery and validation cohorts (Figure 1D); (3) specific brain regions or systems contribute to the encoding of disgust utilizing forward (association) and backward (prediction) encoding models[50] (Figure 1E); and (4) isolated regions or networks traditionally involved in disgust such as the insula, amygdala or default mode network (DMN)[51] are sufficient to predict subjective disgust experience (Figure 1E). The developed visually-induced disgust signature (VIDS) was next utilized to (5) determine the functional specificity of the subjective disgust signature by conducting predictive and topographical comparisons with established decoders for subjective fear[15] and negative affect[34], respectively (Figure 1F), and (6) determine whether core disgust also extends to social experiences by applying the VIDS to data from two experiments conducted in social contexts, including neural reactivity in response to unfair offers in an Ultimatum Game (sociomoral disgust, *n*=43 individuals) and in response to the suffering of other individuals in a pain empathy paradigm[44] (*n*=238 individuals) (Figure 1G). Finally, control analyses were employed in independent datasets that examined neural reactivity in response to positive and negative visual emotional stimuli[52,53] (*n*=150, *n*=30 individuals, respectively) to explore the extent to which the VIDS may capture unspecific

emotional arousal. Together, this systematic examination enables us to establish a comprehensive brain-based model for subjective disgust experience as well as to determine the extent of neurofunctional similarity with the experience of other negative emotions (fear and negative affect) and the experience of unfairness in social situations (commonly referred to as sociomoral disgust).

RESULTS

Visual disgust stimuli robustly evoked different levels of subjective disgust experience

During fMRI, disgust was induced by means of validated disgust-specific affective pictures[54] (DIRTI disgust images database, 'METHODS'). Participants were instructed to react naturally to the stimuli and report their momentary level of disgust during each trial on a 5-point Likert scale ranging from 1 (neutral/slightest disgust) to 5 (very strong disgust). 80 stimuli were presented and robustly induced the entire intensity range of subjective disgust balanced across the most common disgust-related contexts (animal, human, and scene; Figure S1). In the discovery cohort (*n*=78) utilized to develop the disgust signature, the stimuli evoked sufficient levels of subjective disgust experience across contexts such that over 12% of stimuli in each context induced very strong disgust (rating 5), with 72 out of 78 participants reporting all 5 levels of subjective disgust (6 reported levels 1-4).

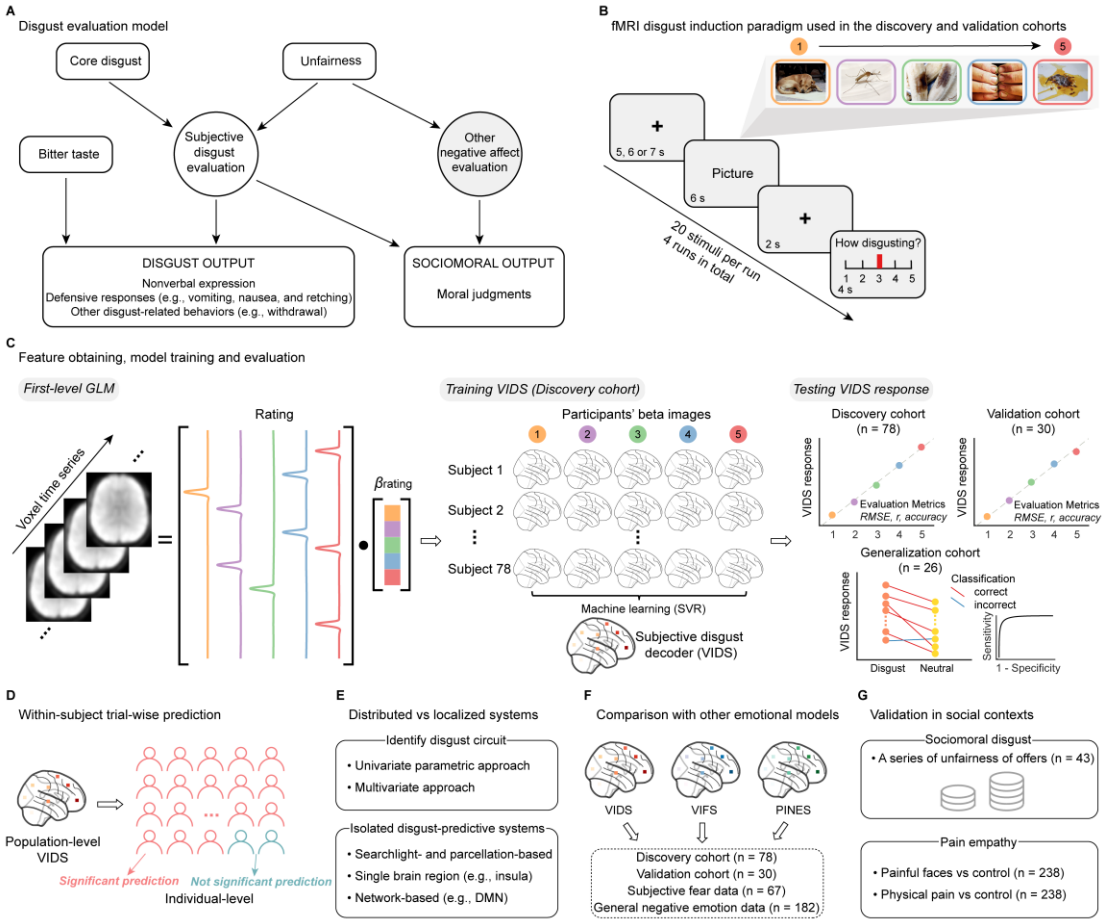

Figure 1. Disgust evaluation model, task design, and analytic workflow

(A) Disgust evaluation model proposed based on previous studies[1,11].

(B) Discovery and validation cohorts underwent the identical disgust induction fMRI paradigm. Example stimuli from the DIRTI disgust images database[54].

(C) Feature determination, model training, and evaluation. Voxel-level brain maps (beta images) were used as features in the prediction analysis. A whole-brain multivariate pattern predictive of the level of subjective disgust (VIDS) was trained on the discovery sample (*n*=78) using SVR and further evaluated in discovery (cross-validated), validation (*n*=30), and generalization (*n*=26) cohorts.

(D) Within-subject trial-wise prediction by applying the population-level VIDS to individual-level trial-

wise disgust experience in both discovery and validation cohorts.

(E) Systematic tests of distributed versus localized disgust systems hypotheses. Univariate and multivariate approaches were employed to determine the contribution of specific brain systems to predict subjective disgust. Multiple prediction analyses were next conducted to test the performance of isolated brain regions or systems in predicting subjective disgust experience.

(F) Testing the specificity of the VIDS in terms of distinguishable representations compared to subjective fear or general negative affect.

(G) Validation of disgust in social contexts. The VIDS was applied to two social contexts, i.e., sociomoral disgust (unfairness), and pain empathy induced by physical (limb) and affective (face) pain-displaying stimuli.

DMN=default mode network; GLM=general linear model; RMSE=root mean squared error; SVR=support vector regression; VIDS=visually-induced disgust signature developed in this study; VIFS=visually-induced fear signature from Zhou et al.[15]; PINES=picture-induced negative emotion signature developed from Chang et al.[34] See 'METHODS' and 'Table S1' for further details.

A brain signature sensitive to predict subjective disgust experience

Consistent with previous studies developing neuro-affective decoders[15,17,34], we employed SVR to identify a whole-brain signature of fMRI activation predictive of subjective disgust experience (Figures 2A and S2). The decoder was developed in the discovery cohort, next performance of the VIDS was evaluated using cross-validation in the discovery cohort (10×10-fold cross-validated, 'METHODS') and by determining reactivity of the VIDS to disgust experience in an independent and pre-registered validation cohort (*n*=30) that underwent the identical paradigm. The developed VIDS accurately predicted the level of subjective disgust in the discovery cohort (averaged within-subject correlation between predicted and true disgust ratings, 5 or 4 pairs of scalar values per subject: *r*=0.88±0.02 (mean±standard error, SE); averaged RMSE=1.43±0.08; overall, i.e., between- and within-subject prediction-outcome correlation for 384 pairs, *r*=0.56, averaged across ten repetitions, bootstrapped 95% confidence interval (CI)=[0.49, 0.62], Figure 2B). To independently determine the robustness of the VIDS, we pre-registered an independent validation cohort (https://osf.io/y4s78). Applying the predictive model - with no further model fitting – to the validation cohort revealed comparable high prediction-outcome correlations (within-subject *r*=0.89±0.03, average RMSE=1.21±0.14; overall prediction-outcome *r*=0.62, 95% CI=[0.53, 0.70]; Figure 2C), indicating a sensitive and robust neurofunctional subjective disgust signature. To further determine the sensitivity of the model for subjective disgust, we compared pairs of activation maps within each subject by means of a two-alternative forced-choice test choosing the activation map with higher VIDS response as the one with higher disgust experience. The VIDS response accurately classified high (average of rating 4 and 5) versus moderate (rating 3) and moderate versus low (average of rating 1 and 2) disgust in both cohorts with 86%-96% accuracy (Cohen's *d*=1.22–1.96), and high versus low with 99%-100% accuracy in both cohorts (Cohen's *d*=2.40–2.92). Moreover, the VIDS response could distinguish each successive pair of disgust rating levels (e.g., rating 1 versus rating 2) with at least 80% accuracy (significantly higher than 50% chance-level; *P*<0.001; except rating 5 versus 4 in the validation cohort, with accuracy being 76% and *P*<0.01) (Figures 2B and 2C).

Given that previous studies emphasized the contribution of the visual cortex in decoding emotional content[7,55], we retrained the decoder excluding the entire occipital lobe and yielded

similar performance, indicating that the prediction is not driven by emotional schemas encoded in visual regions (Supplementary Results, Figure S3).

To further determine the specificity of the VIDS to the experience of disgust (rather than anticipation or cognitive evaluation), we explored the VIDS reactivity over the stimulus presentation interval covering the pre-stimulus interval and onset of the stimulus in discovery (10×10-fold cross-validated) and validation cohorts. In line with the standard hemodynamic model, the VIDS response started approximately 4s following the picture onset and increased during 6-12s (Figures 2D and 2E), suggesting that the VIDS captures disgust experience rather than expectation (pre-stimulus) or cognitive evaluation (response-reporting) of disgust.

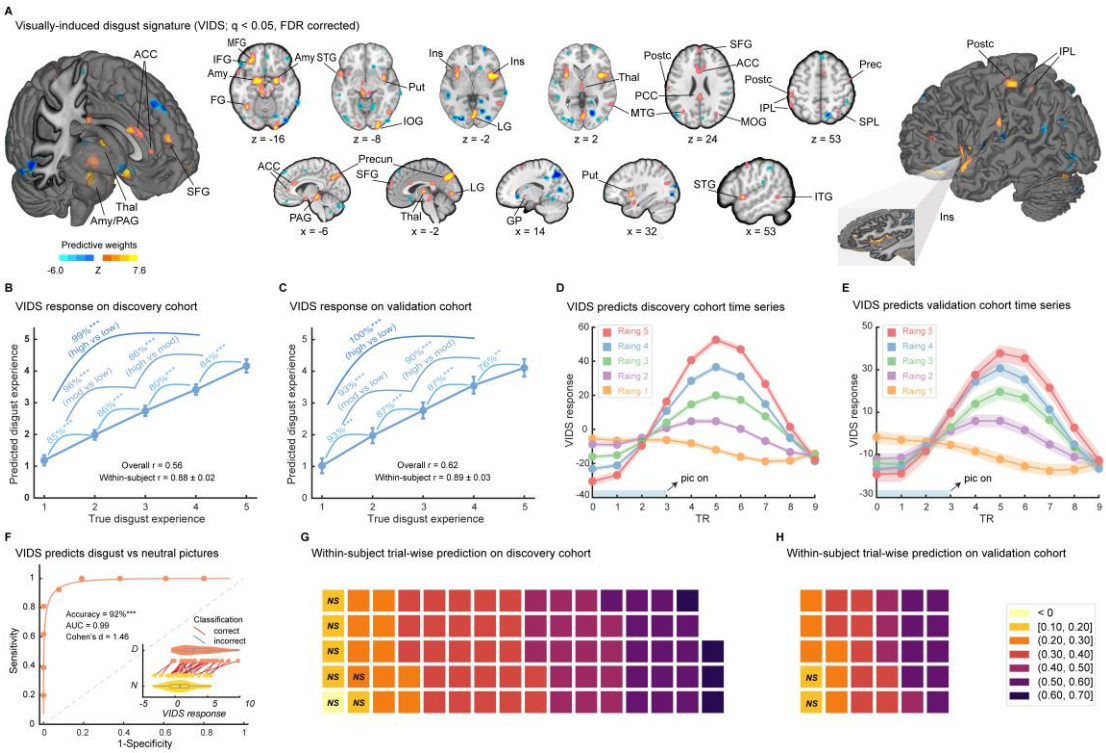

Figure 2. Visually-induced disgust signature (VIDS)

(A) The VIDS pattern thresholded at q<0.05 (FDR corrected, 10,000-sample bootstrap). Figure S2 shows the topography of the unthresholded pattern in anatomical regions of interest previously associated with disgust and emotional experience.

(B and C) Predicted disgust experience (subjective ratings; mean±SE) compared to actual disgust ratings in the cross-validated discovery cohort (n=78) and the independent validation cohort (n=30), respectively. Accuracy provided for forced-choice comparisons. P values based on two-sided independent binomial tests. r indicates Pearson correlation coefficient between predicted and true ratings.

(D and E) Averaged peristimulus plot (mean±SE) of the VIDS response in the cross-validated discovery cohort and the independent validation cohort, respectively, at every repetition time (TR; 2s) for each disgust intensity rating separately.

(F) Generalization of the VIDS in an independent dataset (using two-alternative forced-choice to classify disgust vs neutral stimuli). The figure shows the receiver-operating characteristics (ROC), and the inlet shows the VIDS response for different conditions (D=disgust; N=neutral). Each line connecting dots

represents paired data from the same participant (red=correct classification; blue=incorrect classification). *P* value was based on a binomial test, two-sided.
(G and H) Range of the within-subject trial-wise prediction-outcome correlation coefficients for each participant in discovery and validation cohorts, respectively.
** *P*<0.01, *** *P*<0.001, NS not significant. Error bars and shaded regions indicate SEs. ACC=anterior cingulate cortex; Amy=amygdala; AUC=area under the curve; FDR=false discovery rate, FG=fusiform gyrus; GP=globus pallidus; IFG=inferior frontal gyrus; Ins=insula; IOG=inferior occipital gyrus; IPL=inferior parietal lobule; ITG=inferior temporal gyrus; LG=lingual gyrus; MFG=middle frontal gyrus; MOG=middle occipital gyrus; MTG=middle temporal gyrus; PAG=periaqueductal gray; PCC=posterior cingulate cortex; Postc=postcentral gyrus; Prec=precentral gyrus; Precun=precuneus; Put=putamen; SFG=superior frontal gyrus; SPL=superior parietal lobule; STG=superior temporal gyrus; Thal=thalamus.

Generalization of the VIDS prediction

Accurate and population-level neurofunctional decoders require robust predictions across different paradigms, laboratories, and MRI systems. We therefore tested whether the VIDS could discriminate disgust versus neutral stimuli in a previously published independent study[49] (study 3, Table S1) that examined emotional reactivity to disgusting stimuli. The VIDS classified disgust versus neutral with high accuracy (92%, *P*<0.001, AUC=0.99, and Cohen's *d*=1.46, Figure 2F), indicating robust generalization to predict disgust experience across populations, experiments, and MRI systems.

Within-subject trial-wise prediction

The conscious experience of disgust is a highly subjective and dynamic process[33,46]. A key question is thus to what extent the population-level VIDS can predict momentary variations in disgust on the individual-level (on a trial-by-trial basis). To this end, single-trial models were employed to generate trial- and individual-specific activation maps (~80 beta maps per subject in both cohorts). Next, we calculated the VIDS pattern expressions of these single-trial beta maps, which were finally correlated with the true ratings for each subject separately with statistical significance evaluated by prediction−outcome correlation (Pearson) for each subject. The results showed that the VIDS could significantly predict trial-by-trial disgust ratings in over 90% of the individuals in both, discovery (71 from 78, cross-validated) and validation (28 from 30) cohorts (Figures 2G and 2H). Mean prediction−outcome correlations were 0.39±0.02 and 0.41±0.02 for discovery and validation cohorts, respectively. Our findings thus suggest that although disgust experience differs between individuals[33,46], the VIDS could accurately and robustly predict the intensity of momentary disgust experience on the individual-level across two independent populations.

Subjective disgust is associated with and predicted by distributed neural systems

We next systematically determined which brain regions and systems contribute to the whole-brain disgust prediction. We first examined regions that made reproducible (reliable) contributions to the disgust prediction within the VIDS itself by applying a bootstrap test to identify regions with significant, consistent model weights (*q*<0.05, FDR corrected)[56]. Given that some brain features could contribute to controlling for noise in the data[50] rather than the emotional process per se, we next transformed the population-level VIDS into 'activation pattern' ('structure coefficient', details see 'METHODS'). Results showed that a set of distributed subcortical and cortical brain systems

exhibited significant model weights (Figure 3A) and structure coefficients (Figure 3B), including the bilateral amygdala, anterior and mid- and posterior-insular cortex, hippocampus, dorsal ACC, PAG, thalamus and putamen as well as the right mid-cingulate cortex, left posterior cingulate cortex, left inferior frontal gyrus, left superior temporal gyrus, right precentral gyrus, occipital and supplementary motor areas (Figure 3C). Brain regions associated with and predictive of disgust experience on the individual-level determined with convergent univariate and multivariate approaches identified a similar set of broadly distributed regions ('Supplementary Methods' and 'Supplementary Results'). Overall, the analyses indicated that the conscious experience of disgust is represented in distributed subcortical and cortical systems involved in early detection and defensive responses towards potential threats (amygdala, PAG, thalamus, putamen)[7,15], regions involved in interoceptive awareness, self-referential processing, and explicit emotional processing such as the anterior insula, dorsal ACC and frontal regions[26,27,39,57], as well as regions that have been identified as core systems mediating the bitter taste reaction in animal models (posterior insula[5], brainstem[4]). Negative associations with disgust were most consistently identified in superior frontal and middle temporal regions.

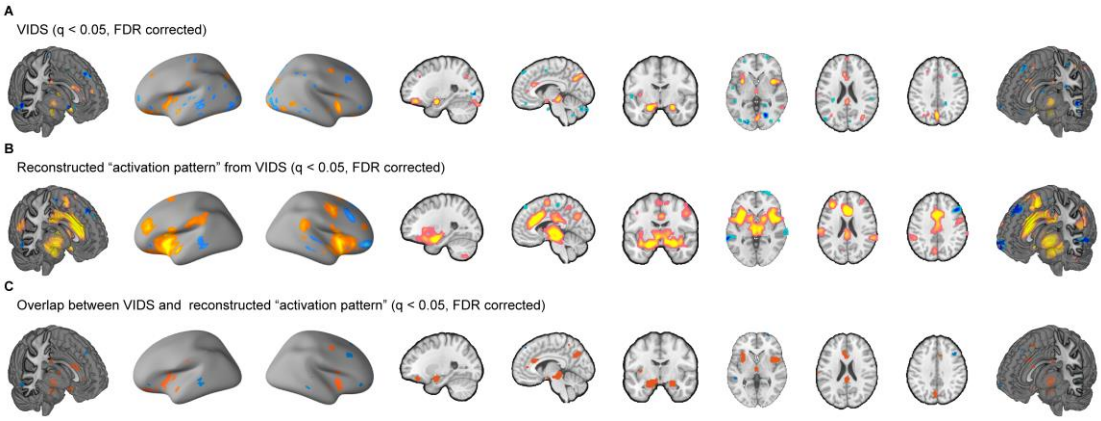

Figure 3. Subjective disgust experience is associated with and predicted by distributed brain regions

(A) Thresholded VIDS.

(B) Thresholded transformed VIDS 'activation pattern'.

(C) Conjunction between VIDS and transformed 'activation pattern'.

Images thresholded at $q<0.05$, FDR corrected. Hot color indicates weights (A) or positive associations (B) whereas cold color indicates weights (A) or negative associations (B).

Spatial overlap of the neurofunctional disgust signature with the Brainnetome Atlas further confirmed a biological-plausible disgust model

To further evaluate the neurobiological plausibility and validity of the developed model, we calculated the spatial overlap percentage between the thresholded VIDS with the modified 279-region version of the Brainnnetome Atlas[58,59] ('METHODS').The top 15 atlas regions visualized in Figure 4A included four amygdala sub-regions, five insula sub-regions, orbital gyrus, thalamus, rostral hippocampus, rostroventral cingulate gyrus as well as brainstem and midbrain regions involved in defensive responses and conscious emotional experience (i.e., PAG and substantia nigra[60]). Meta-analytic behavioral decoding utilizing the BrainMap database confirmed that the

identified brain systems were primarily associated with negative emotional processes, in particular disgust showed the largest activation likelihood ratio (Figures 4B and 4C).

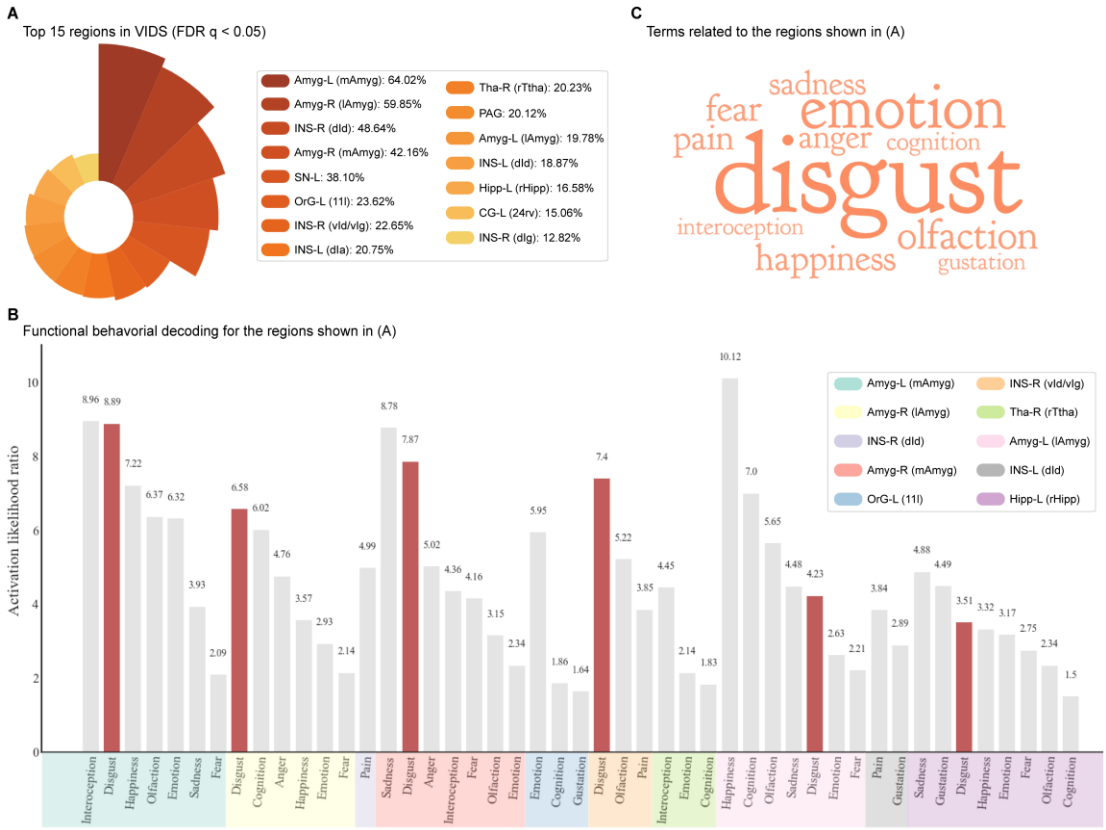

Figure 4. Neurobiological validity of the thresholded VIDS

(A) The top 15 atlas regions overlapping with the thresholded VIDS (retaining positive values).

(B) Functional behavioral decoding for 10 exemplary regions for the thresholded VIDS (provided by the Brainnetome website).

(C) Meta-analytic behavioral decoding of the identified regions. Font size in the word cloud is proportional to the activation likelihood ratio related to each term provided by BrainMap, with larger activation likelihood ratio value depicted by larger font size.

Alternative models to determine the contribution of isolated disgust-predictive systems: local searchlights, pre-defined regions, and networks show lower predictive accuracy compared to the VIDS

Given the continuing debate on the contribution of specific brain regions, such as insula[7,20,61,62], amygdala[7,63], and the DMN[51,64], to subjective disgust experience, (1) both searchlight- and parcellation-based analyses were employed to determine local brain regions that were predictive of subjective disgust experience, and (2) models were trained on single brain region and network to examine to what extent these models could predict subjective disgust experience compare to the whole-brain VIDS. As shown in Figures 5A and 5B, subjective disgust experience could be significantly predicted by activations in widely distributed regions (averaged across 10×10-fold cross-validation procedure). However, none of the local models predicted subjective disgust to the extent the VIDS did (see Figures S5A and S5B for predictions of models trained on discovery cohort

on validation cohort).

We next re-trained predictive SVR models restricted to activations in (1) the bilateral insula; (2) the bilateral amygdala; (3) a consciousness network; (4) a subcortical network; (5) each of seven large-scale resting-state functional networks ('METHODS'). Results showed that the insula (prediction-outcome correlation r=0.23 and 0.38 for discovery (cross-validation) and validation cohorts, respectively), the amygdala (prediction-outcome correlation r=0.27 and 0.39 for discovery (cross-validation) and validation cohorts, respectively) as well as other brain networks (Figures 5C-5E and S5C-S5E, Table S2) could, to some extent, predict subjective disgust experience. Nonetheless, although statistically significant (Ps<0.001), the effect sizes in terms of prediction-outcome correlations (including searchlight- and parcellation-based predictions) were substantially smaller than those obtained from the VIDS, which used features spanning multiple brain systems. See also 'Supplementary Methods' and 'Supplementary Results' for separating a classically insula-associated process (pain empathy) from the representation of disgust experience.

To control for potential effects of the number of features/voxels in prediction analyses (i.e., the whole-brain model contains much more features), we randomly selected voxels (repeated 1,000 times) from a uniform distribution spanning the entire brain (black), consciousness network (light orange), subcortical (brown) or individual resting-state networks (averaged over 1,000 iterations)[15,35]. The asymptotic prediction when sampling from all brain systems as we did with the VIDS (black line in Figures 5E and S5E) was substantially higher than the asymptotic prediction within individual networks (colored lines in Figures 5E and S5E; see also Table S2 for details). This analysis thus demonstrated that whole-brain models have much larger effect sizes than those using features from a single network. Moreover, model performance was optimized (i.e., reaching asymptote) when approximately 10,000 voxels were randomly sampled across the whole-brain, as long as voxels were drawn from multiple brain systems, further confirming that information about subjective disgust experience is contained in patterns of activity that span multiple systems. In addition, similar results were observed when applying the models trained on discovery cohort to the validation cohort (Figure S5E), indicating that models trained on approximately 10,000 randomly sampled voxels were robust enough (for similar findings, see ref.[15]). Together, converging lines of evidence from the above systematic analyses point to the fact that subjective disgust experience is encoded in distributed neural patterns that span multiple systems, adding to increasing evidence that emotions are represented in distributed brain systems rather than single brain regions or networks.

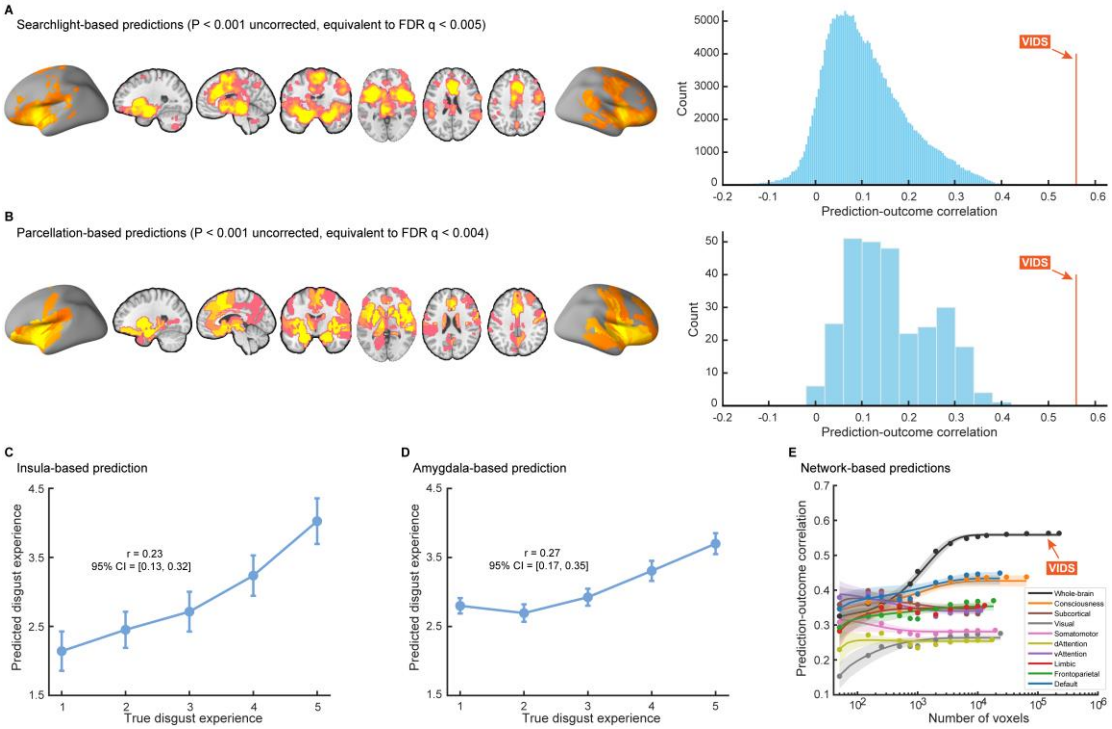

Figure 5. Local brain region and network predictions in the discovery cohort

(A and B) Brain regions that significantly predict subjective disgust experience revealed by searchlight and parcellation-based analyses, respectively. Uncorrected *P* values equivalent to *q*<0.05 might be considered lenient, therefore brain regions that survived *P*<0.001 uncorrected (corresponding to *q*<0.005 and 0.004, FDR corrected, for searchlight- and parcellation-based predictions, respectively) are displayed. Histograms: cross-validated predictions (correlations) from local searchlights or parcellations. Orange lines indicate the prediction-outcome correlation from VIDS.

(C and D) Cross-validated predictions (mean±SE) from insula- and amygdala-based prediction analyses, respectively. Error bar indicates standard error of the mean; *r* indicates overall (between- and within-subjects; i.e., *n*=384 pairs) prediction-outcome Pearson correlation coefficient.

(E) Subjective experience of disgust is distributed across multiple systems. Model performance was evaluated as increasing numbers of voxels/features (x-axis) were selected to predict subjective disgust in different regions of interest including the entire brain (black), consciousness network (light orange), subcortical network (brown), or individual resting-state networks (other colored lines). The y-axis denotes the cross-validated prediction-outcome correlation. Colored dots indicate the mean correlation coefficients, solid lines indicate the mean parametric fit, and shaded regions indicate standard deviation.

The experience of disgust, fear, and non-specific negative affect have shared yet distinguishable neurofunctional representations

Disgust is inherently inter-related with general negative affect and functionally similar avoidance responses like fear[7]. To determine shared and separable neurofunctional representations, we conducted a series of analyses. First, we investigated spatial similarities between stable decoding maps and a set of ROIs as well as networks. Amygdala, anterior insula, basal ganglia, lPFC, and MCC contained stable predictive voxels in all three models, but VIDS showed larger contributions to the first four ROIs than VIFS and PINES while VIFS exhibited the largest contributions to MCC; furthermore, the remaining three ROIs (PAG, ACC, and thalamus) exhibited a degree of specificity

for VIDS and VIFS (Figure 6A). Although all networks showed stable predictive voxels across the three models, the relative contributions of each network to each model varied, such that the ventral attention, limbic, and consciousness networks contributed stronger to VIDS than to VIFS and PINES whereas the somatomotor network strongly contributed to PINES (Figures 6B and S7A). Second, we examined functional similarities between the VIDS and PINES[34] (general negative emotion experience) and VIFS[15] (fear), respectively. The results showed that VIDS was more specific to predict high versus low disgust as compared with PINES or VIFS, as reflected by effect sizes 1.56–2.76 and 2.09–2.61 times larger than those for PINES or VIFS in disgust discovery and validation cohorts, respectively (Figure 6C). PINES was more sensitive to predict high versus low negative emotion with effect sizes 2.18–3.74 higher than those for VIDS or VIFS, and VIFS more accurately predicted high versus low fear with effect sizes 1.95–2.10 higher than those for PINES or VIDS (Figure 6C). Particularly, additional analyses revealed that the three decoders were rather specific for their targeted emotion during discrimination at moderate versus low intensity (e.g., VIDS had effect sizes 1.84–27 and 2.31–11.53 times higher than those for PINES or VIFS in disgust discovery and validation cohorts, respectively (Figure 6D). The above findings were further substantiated by the comparisons of the overall and within-subject prediction-outcome correlations of the three decoders across four datasets (Table S3).

Finally, we employed multilevel mediation models to determine the neurofunctional relationship between the representations of subjective emotional experiences encoded in VIDS, PINES, and VIFS. While PINES could to some degree track disgust (discovery cohort: r=0.24; validation cohort: r=0.30), the VIDS response fully mediated the effect of PINES response on subjective disgust ratings in the discovery cohort, and in the validation cohort the VIDS response partially mediated the effect of PINES response on disgust ratings (Figure 6E). In contrast, the PINES response failed to mediate the effect of VIDS response on disgust ratings (Figure 6F). VIFS could predict subjective disgust (discovery cohort: r=0.35; validation cohort: r=0.32), and corresponding multilevel mediation models revealed that VIDS partially mediates the effect of VIFS response on disgust ratings and vice versa (Figures S7B and S7C); however, the first model (Figure S7B) had 3 times larger effect sizes than the second one (Figure S7C) in the discovery cohort and a similar trend was observed in the validation cohort, suggesting a stronger influence of VIDS in partially mediating the effects of VIFS response on disgust ratings. See Supplementary Results for further details.

Together, these findings underscore that neural representations of subjective experiences of negative emotions engage shared yet distinct representations which may shape distinguishable subjective experiences.

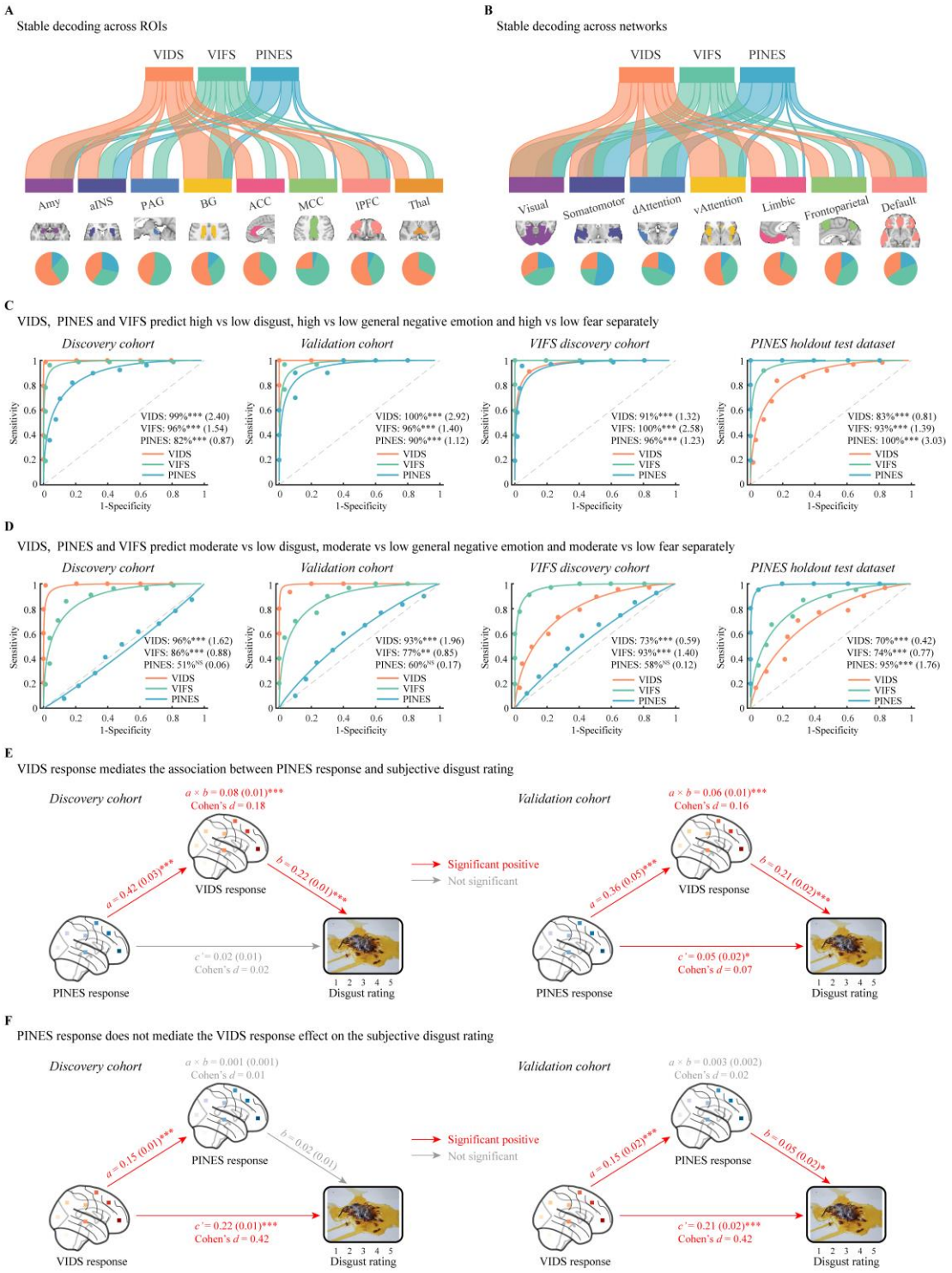

Figure 6. Comparing neurofunctional decoders for disgust, fear, and negative affect (VIDS, VIFS, PINES)

(A and B) River plots showing spatial similarity (cosine similarity) between stable decoding maps and selected brain systems. Ribbons are normalized by the max cosine similarity across all ROIs and networks. Stable decoding models were thresholded at FDR $q < 0.05$ and positive voxels were retained only for similarity calculation and explanation. Ribbon locations in relation to the boxes are arbitrary. Pie charts show relative contributions of each model to each ROI or network (i.e., percentage of voxels with highest cosine similarity for each map).

(C and D) VIDS, PINES, and VIFS more accurately (shown as forced-choice classification accuracy and Cohen's *d*) predict the targeted emotion during discrimination at high versus low and moderate versus low intensities, respectively.
(E) VIDS response fully mediates the association between PINES response and subjective disgust ratings in the discovery cohort, and VIDS response partially mediates the PINES response - disgust rating association in the validation cohort.
(F) PINES response does not mediate the effect of VIDS response on disgust ratings, neither in the discovery cohort nor the validation cohort.
(C and D) ** $P<0.01$, *** $P<0.001$, NS not significant.
(E and F) * $P<0.05$, *** $P<0.001$ (bootstrap tests with 10,000 samples; two-sided).
aINS=anterior insula; BG=basal ganglia; MCC=midcingulate cortex; lPFC=lateral prefrontal cortex.

The neural signature of subjective disgust experience captures sociomoral disgust in terms of neural reaction to unfairness

In everyday life, humans frequently refer to disgust experience in social (i.e., sociomoral) contexts and this has also entered pop-cultural contexts (e.g., '… I was ***disgusted*** by all the injustice, all the injustice…' from Scream, Michael Jackson). However, debates in social and affective neurosciences continue on whether the evolutionary 'core disgust' that has evolved to avoid pathogens has been adopted towards social situations[7], with some evidence showing that disgust experience evoked by moral transgressions (e.g., unfair offers) may be similar to the disgust experience elicited by core disgust stimuli (e.g., uncleanliness, pathogens)[3,65], while other evidence suggested that unfair offers may evoke a general negative emotional response rather than disgust experience[48]. Here, we thus test whether the VIDS - which captures subjective disgust experience elicited by core disgust stimuli - could track sociomoral disgust. To this end, we performed exploratory predictions by applying the VIDS pattern along with PINES and VIFS to an independent fMRI dataset acquired while participants received a series of unfair offers in a social exchange (Ultimatum Game) task (study 6, n=43; 'Supplementary Methods' and Table S1). As shown in Figures 7A and 7B, VIDS could predict high versus low unfairness with high accuracy (accuracy=74%, P=0.0019, Cohen's d=0.73), and could also predict high versus moderate unfairness (accuracy=70%, P=0.0137, Cohen's d=0.37). PINES could predict high versus low unfairness (accuracy=77%, P=0.0006, Cohen's d=0.54) yet failed to predict high versus moderate unfairness (accuracy=56%, Cohen's d=0.20). VIFS did not predict either high versus low unfairness or high versus moderate unfairness. These results indicated that the neurofunctional reaction to unfair offers is highly similar to the emotional reaction towards core disgust stimuli.

To further test whether VIDS simply captures negative emotional processing in high arousing social contexts, we applied the VIDS to another independent fMRI dataset (study 7, n=238; Table S1) that acquired neural response during physical pain and painful facial stimuli as well as respective non-painful control stimuli. Results showed that the VIDS could not discriminate high arousing painful faces from respective control faces (with accuracy near chance-level, Figure 7C); although VIDS successfully distinguished physical pain versus corresponding control pictures (Figure 7D), the effect size was substantially lower than that using VIDS to predict high versus low unfairness (Figure 7A). Combined, these findings indicated that VIDS is not sensitive to separate high arousing pain empathy from control stimuli, yet it is relatively specific for disgust experience in social contexts. For further analyses confirming that VIDS tracks unspecific arousal across

negative and positive stimuli to a comparably low extent see also 'Supplementary Methods' and 'Supplementary Results'.

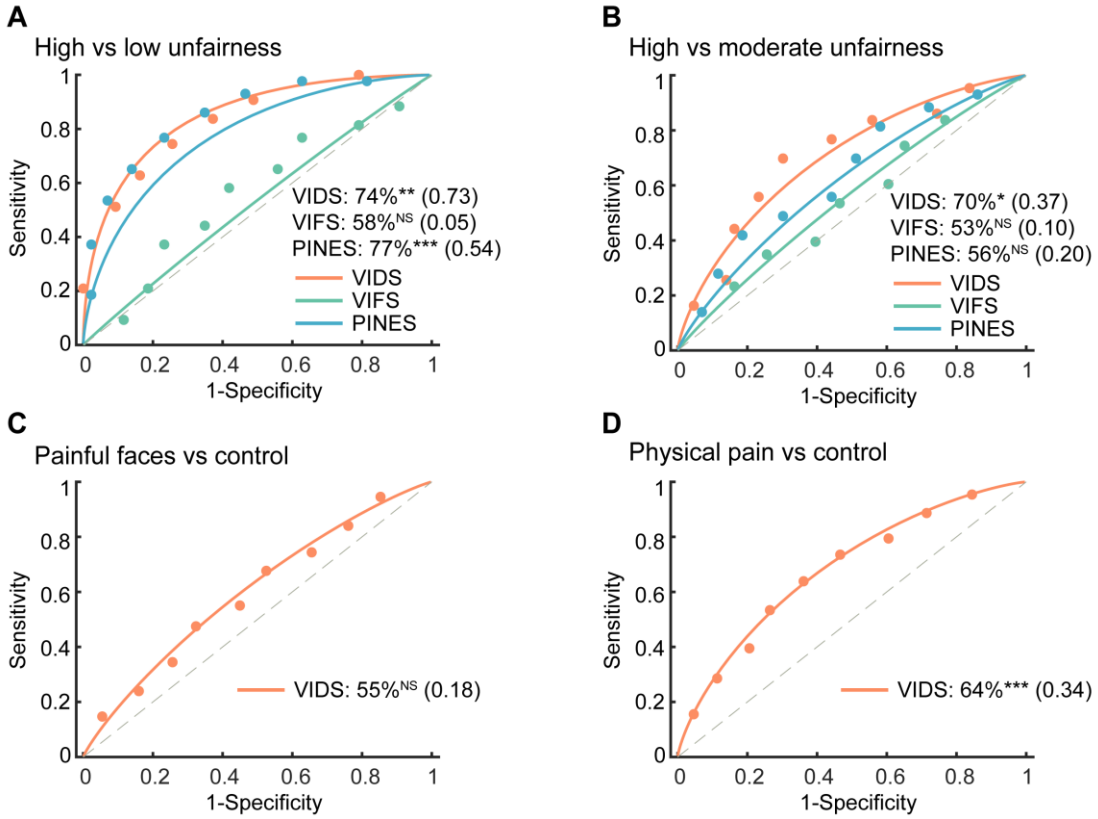

Figure 7. Additional validation tests of VIDS in social contexts

(A) VIDS and PINES could predict high versus low unfairness (shown as forced-choice classification accuracy and Cohen's d), while VIFS failed to predict high versus low unfairness.

(B) Neither PINES nor VIFS could predict high versus moderate unfairness, however, VIDS predicted high versus moderate unfairness.

(C) VIDS failed to predict painful faces versus respective control.

(D) VIDS could predict physical pain versus corresponding control.

* $P<0.05$, ** $P<0.01$, *** $P<0.001$, NS not significant.

DISCUSSION

Recent affective and clinical neuroscience perspectives propose a paradigm shift towards subjective and conscious emotional experiences[15,39,45,66,67], yet neurobiological models that accurately describe the respective neural representations are scarce. Disgust originates in the hard-wired mammalian distaste reflex, but in humans its conscious and subjective emotional experience is considerably shaped by subjective appraisal and may extend towards sociomoral contexts and underlie some psychopathological dysregulations (e.g., obsessive-compulsive disorder). Utilizing machine-learning based neural decoding, we established and validated a comprehensive neurobiological model for subjective disgust experience (VIDS) that generalizes across samples and study contexts with regard to accurately predicting disgust in response to core disgust stimuli as well as unfair offers, suggesting that the experience of disgust may extend into sociomoral domains. Subjective disgust was encoded in distributed subcortical and cortical brain systems while no isolated region or network reached the predictive accuracy of the whole-brain model. In a series of analyses, we further demonstrated that subjective disgust experience showed shared yet separable neural representations with established predictive models for the subjective experience of non-specific negative affect or fear, such that all models engaged subcortical regions and regions implicated in emotional awareness and appraisal, whereas the neural signature of disgust robustly mediated the response of the other two models (PINES, VIFS) on subjective disgust ratings but not vice versa. The VIDS moreover outperformed the other signatures in predicting disgust experience in response to both core and sociomoral disgust (unfair offers); while it tracked unspecific arousal in response to painful or positive pictures, the extent was comparably low. Together, the current study provides the first comprehensive neurofunctional model of subjective disgust, and the resulting VIDS can predict subjective disgust experience with high robustness and generalization and thus may have a high clinical utility as fMRI-based marker for subjective disgust and read-out for the evaluation of novel interventions.

Previous studies emphasized the role of single brain systems or networks as core modules for disgust, including the insula[6,7,19,20] or the ventral attention network which may facilitate detection and attentional processing of core disgust stimuli[68]. Contrary to structure- and network-centric views, our findings indicate that subjective disgust experience requires concerted engagement of brain-wide distributed representations with comparably strong contributions of subcortical regions involved in rapid threat detection and avoidance responses (amygdala, PAG, thalamus, putamen)[7,15] and cortical systems engaged in interoceptive awareness and emotional appraisal such as the anterior insula, dorsal ACC and lateral frontal regions[26,27,39,57], as well as regions that have been identified as mediators of the bitter taste reaction in animal models (posterior insula[5], brainstem[4]). However, neural activity in these isolated regions including systems such as insula or amygdala which have been typically implicated in disgust was only weakly associated with the subjective experience.

While no single network (e.g., visual network and ventral attention network) was critical for predicting subjective disgust, the consciousness network and DMN exhibited comparably strong contributions. According to two-system and higher-order consciousness models of emotion[57], the consciousness network is vital for the generation and appraisal of conscious emotional states. In the context of constructionist theories of emotion, the DMN has been proposed to shape current affective experiences by drawing on autobiographical experiences and knowledge[51] and to facilitate contextually sensitive escape decisions[69]. Consistent with previous predictive models for

subjective emotional experiences[15], our analyses also revealed that approximately 10,000 voxels that were randomly sampled across the whole-brain could lead to high predictive performance for subjective disgust. Together, our results dovetail with a growing body of research from both univariate[70] and multivariate perspectives[15,34,37,38,47] which demonstrated that capturing affective experience requires integration across multiple distributed neural systems. The distributed representation perspective aligns with appraisal[71] and constructionist[46,70,72] theories of emotions which propose that shared but also distinct distributed functional assemblies integrate to facilitate subjective emotional experiences.

In terms of the adaptive function, pathological dysregulations as well as subjective triggers and conscious experience level, disgust is - to a certain extent – distinguishable from related emotional domains like fear or negative affect (see e.g., also Research Domain Criteria, RDoC[73]). For robust fMRI-neuromarkers, a precise differentiation between subjective states is critical for model development and translation[39,41,43]. The VIDS exhibited partly overlapping yet also distinct neurofunctional representations compared to the PINES (non-specific negative affect)[34] or the VIFS (fear experience)[15]. While prediction of all neuroaffective signatures was driven by contributions of distributed subcortical and cortical systems on the distributed signature level, the VIDS was more sensitive to predict subjective disgust than general negative affect or subjective fear. Moreover, the VIDS response mediated the association between PINES response and disgust ratings but not vice versa. Additionally, the correlation between VIFS response and disgust ratings was robustly mediated by the VIDS response. Together, these findings may underscore that the VIDS is predictive of subjective disgust experience rather than predicting unspecific salience or arousal or non-specific aversive emotional states and that shared yet also separable neurofunctional representations shape subjective emotional experiences.

We next utilized the disgust predictive model to determine whether disgust experience extends into the sociomoral domain[3,48,65], in particular whether exposure to unfairness activates the VIDS. Evolutionary and disgust appraisal models have long hypothesized an association[1,11], but the lack of a precise neuromarker for subjective disgust precluded a direct test of the hypothesis. Chapman et al. thus emphasized that the MVPA approach is suitable to examine the extent of core disgust appraisal being involved in sociomoral disgust as it provides evidence for the activation of shared appraisal while it is concomitantly critical to decode other aversive states associated with withdrawal (e.g., fear)[11]. Our results revealed that VIDS could predict different levels of unfairness with high accuracy (high, moderate, low unfairness) while PINES could differentiate high unfairness from low unfairness yet failed to predict high versus moderate unfairness, and VIFS did not react to unfairness. These findings largely converged with the observations from Chapman et al.[3], where the authors showed that participants judged their experience as most similar to disgust when they received unfair offers, while other emotions like fear did not change with unfair offers. More importantly, however, the authors noted that the emotional response to unfairness was not characterized by disgust alone; rather, it is a complex blend of multiple emotions – anger and sadness also play a role, but to a less degree than disgust[3]. Indeed, through item analysis, Chang et al. found that the PINES response was strongly associated with normative ratings of multiple discrete emotions (e.g., disgust, r=0.94; anger, r=0.94; sadness, r=0.92)[34], which explains why the PINES also captured unfairness to a certain extent. Our findings extend the initially proposed disgust appraisal model depicted in Figure 1A, which combined several possibilities raised in previous disgust appraisal models[1,11]. Rozin et al. once mentioned that '…only if evidence is found

for a route from unfairness to the disgust evaluation system can it be concluded that disgust at unfairness is "the same" as disgust that is elicited through the core route (such as in response to cockroaches) ...'[1]. We here demonstrate neurobiological evidence for the route between unfairness and the subjective disgust evaluation system. From the evolutionary perspective, these findings support the notion that disgust may have been preadapted to avoid harm and regulate behavior in social contexts.

From a biomarker perspective, it is imperative that a neuroaffective signature captures the respective mental process across variations of experimental contexts[35]. The VIDS generalized across cohorts, core and sociomoral paradigms, MRI systems and predicted disgust with considerably higher effect sizes than reactivity to high arousing positive or high arousing negative stimuli or strongly aversive vicarious pain stimuli (pain infliction stimuli). See 'Supplementary Results' for further discussions on the arousal issue. Since the insula plays an important role in encoding pain empathy and disgust[44,48,74], we further tested a physical pain empathy pattern[44] on discovery and validation cohorts, yet the performance was substantially inferior to VIDS (see Figures S6A and S6B). Together with the in-depth comparison of the predictive and topographical features of the fear, non-specific negative affect, and disgust signatures, we demonstrate that the VIDS is a comparably robust, sensitive, and specific brain marker to track and differentiate subjective disgust experience across contexts.

The present study has the following limitations. While we observed partly overlapping systems between the subjective experience representations of disgust and fear, the extent to which these common representations represent hard-wired physiological defensive responses, autonomous reactivity, or general emotional arousal remains unclear. Furthermore, this study performed explanatory predictions by testing the developed VIDS, together with VIFS and PINES, in the sociomoral domain; however, we could not provide a direct prediction of the degree of other negative emotions like anger engaged throughout this process (for the role of anger in sociomoral disgust, see ref.[11,75]).

In conclusion, the present study developed a whole-brain neural signature for subjective disgust experience. This visually-induced disgust pattern was validated and generalized across participants, paradigms, study contexts, and MRI systems. We showed that the neural basis of subjective disgust is encoded in multiple distributed regions. The robustness and specificity of the resulting VIDS were further tested with general negative emotion experience, subjective fear experience, and social contexts (e.g., pain empathy and sociomoral disgust). This study provides objective brain measures that can serve to understand the neurobiological underpinnings of subjective disgust experience which is centered on many disorders, track the progression of disorders over time, and serve as intermediate markers for treatments and interventions.

METHODS

Participants

Discovery Cohort (Study 1) Discovery cohort (labeled as study 1) included eighty healthy participants recruited from the University of Electronic Science and Technology of China (UESTC). All participants were right-handed, reported having normal or corrected to normal vision, had no history of psychiatric or physical disorders, no MRI contraindications, and were free of current or regular substance or medication use. Data from two participants (one female) were excluded due to excessive head movement (>3mm) during fMRI scanning, therefore, leading to a final sample of *n*=78 participants (44 females; mean±SD age=22.10±2.68 years). Informed consent was obtained before the experiment, the experimental protocol was approved by the UESTC Ethics Board and in line with the latest revision of the Declaration of Helsinki. Participants were reimbursed 110 RMB.

Validation Cohort (Study 2) In order to validate the performance of the disgust decoder (VIDS) developed using discovery cohort (i.e., study 1), we next collected an independent dataset (validation cohort, study 2) using the identical paradigm as in the discovery cohort. The data acquisition plan and analyses were pre-registered before the start of the validation experiment (https://osf.io/y4s78). Study 2 included thirty-one healthy participants recruited from UESTC. Enrollment criteria were Identical to study 1. Data from one male participant were excluded due to excessive head movement (>3mm) during fMRI scanning, leading to a final sample of *n*=30 participants (16 females; mean±SD age=21.13±2.18 years). Informed consent was obtained before the experiment, the experimental protocol was approved by the UESTC Ethics Board and in line with the latest revision of the Declaration of Helsinki. Participants were reimbursed 110 RMB.

Experimental procedures and paradigms

Discovery Cohort (Study 1)

Stimuli and Experimental Paradigm

The disgust rating task incorporated 80 pictorial stimuli from the Disgust-related Images (DIRTI) database[54] (67 pictures), Nencki Affective Picture System (NAPS)[76] (5 pictures), and International Affective Picture System (IAPS)[77] (8 pictures), distributing over 4 runs (with 20 stimuli per run). Of note, to avoid specificity driven by different categories of disgust, we drew a balanced sample from animals, humans, and scenes. Stimuli were presented using the E-Prime stimulus presentation software (Version 2.0; Psychology Software Tools, Sharpsburg, PA). Participants were instructed to pay attention to the pictures and naturally experience the induced emotion. Each trial consisted of a 6s presentation of the picture followed by a 2s fixation-cross separating the stimuli from the rating period. Participants then had 4s to report the level of disgust they experienced for the stimuli using a 5-point Likert scale where 1 indicated neutral/slightest disgust and 5 indicated most strongly disgust. Finally, there was a jittered fixation-cross epoch (5, 6, or 7s) before the presentation of the next picture (Figure 1B). All participants reported '1-4' in their responses while 6 out of 78 participants did not use rating '5'.

Validation Cohort (Study 2)

Stimuli and Experimental Paradigm

Stimuli and paradigm used in the validation cohort were identical to the discovery cohort. All participants reported '1-4' in their responses while 1 out of 30 participants did not use rating '5'.

MRI data acquisition for discovery and validation cohorts

MRI data were acquired on a 3T system (GE MR750, General Electric Medical System, Milwaukee, WI, USA). Structural images were collected using a high-resolution T1 spoiled gradient recall (SPGR) sequence (repetition time=6ms, echo time=2ms, flip angle=9°, field of view=256×256mm, 1×1×1mm voxels, 256×256 acquisition matrix, 156 slices, 1mm slice thickness) and were used for improving spatial normalization and excluding participants with apparent brain pathologies. Functional images were acquired with a T2*-weighted echo planar imaging (EPI) sequence (repetition time=2000ms, echo time=30ms, flip angle=90°, field of view=240×240mm, 3.75×3.75×4mm voxels, 64×64 resolution, 39 slices, 3mm slice thickness, 1mm gap).

fMRI data preprocessing in discovery and validation cohorts

In line with our previous studies[15,44], all fMRI data were preprocessed and analyzed using SPM12 (Statistical Parametric Mapping, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The first five volumes of the fMRI data in each run were removed to allow for T1 equilibration. Briefly, the preprocessing steps included: correction for differences in slice timing, realignment of head motion, unwarp to correct for magnetic field inhomogeneity, followed by the tissue segmentation of high-resolution anatomical image (into gray matter, white matter, cerebrospinal fluid, bone, fat and air), co-registration of the skull-stripped and bias-corrected structural image to the functional images, normalization of functional images to the standard Montreal Neurological Institute (MNI) space (interpolated to $2\times2\times2mm^3$ voxels) and spatial smoothing with an 8-mm full width at half maximum (FWHM) Gaussian kernel. Before first-level analysis, global motion outliers in each run were identified according to any of the following criteria: (a) signal intensity exceeding three standard deviations from the global mean or (b) signal intensity and Mahalanobis distances exceeding ten mean absolute deviations (https://github.com/canlab/CanlabCore). Each time point identified as outliers was later included as a separate nuisance covariate in the first-level model.

First-level fMRI analysis in discovery and validation cohorts

We designed two separate participant-level univariate general linear models (GLM). Specifically, the first GLM model used to create beta images for the following prediction analyses included five separate regressors time-logged to the presentations of pictures for each rating (i.e., 1-5), enabling us to model brain activity in response to each disgust level separately. The model also included one boxcar regressor indicating the 4s rating period aiming to model effects correlated with motor activity. The fixation cross epoch was used as an implicit baseline. The second GLM model employed a parametric modelling approach using subjective disgust ratings encompassing two regressors of interest, with one modeling the picture presentation period and the other modeling the subjective disgust rating period. The self-reported disgust ratings (i.e., 1-5) for each picture were included in the statistical model as a parametric modulator for the presentation period. The model thus allowed to determine brain regions where neural activity changed as a function of disgust ratings.

Task regressors were convolved with the canonical hemodynamic response function and a high-pass filter of 128s was applied. The four runs of the fMRI task were concatenated for each

participant (via SPM spm_fmri_concatenate.m function). Nuisance variables encompassed: (a) 'dummy' regressors representing each run (intercept for each run); (b) the six estimated head movement parameters (*X*, *Y*, *Z*, roll, yaw, and pitch), their squares, their derivatives and squared derivatives for each run (24 columns in total); (c) vectors indicating motion outlier time points.

Multivariate pattern analysis

We used whole-brain multivariate machine-learning pattern analysis[15,34,38,56] to determine a pattern of weighted brain activity that best predicted the self-reported disgust ratings. Here, we used the support vector regression (SVR) algorithm (linear kernel with C=1) implemented in the Spider toolbox (http://people.kyb.tuebingen.mpg.de/spider). Whole-brain data masked with a gray matter mask from the discovery cohort was used. Features (i.e., predictors) constituted 384 beta maps (one per rating for each participant, i.e., 72×5 + 6×4), aggregated into an images × voxels matrix (stacked across participants), and the outcome variable included subjective disgust ratings across participants. To test the performance of the disgust-predictive pattern, we used a rigorous 10×10-fold cross-validation procedure on the discovery cohort during which all participants were randomly assigned to 10 subsamples of 7 or 8 participants using MATLAB's cvpartition function. The optimal hyperplane was computed based on the multivariate pattern of 70 or 71 participants (training set) and evaluated by the excluded 8 or 7 participants (test set). This procedure was repeated ten times with each subsample being the testing set once. To avoid a potential bias of training-test splits, the cross-validation procedures throughout the study were repeated ten times by producing different splits in each repetition and the resultant prediction performance were averaged to produce a convergent estimation[15,44,78].

To further evaluate the performance of the disgust-predictive pattern (trained on the whole discovery cohort), we assessed overall (between- and within-subjects; 384 pairs in total) and within-subject (5 or 4 pairs per subject) Pearson correlations (r) between the cross-validated predictions and the actual ratings to indicate the effect sizes and the RMSE to illustrate overall prediction error (for similar model evaluation metrics, see ref.[38]). Furthermore, we computed the classification accuracies of the VIDS between each successive pair of disgust rating levels (rating 2 versus 1, rating 3 versus 2, rating 4 versus 3, and rating 5 versus 4) as well as between low, moderate and high disgust rating levels (i.e., high versus low, high versus moderate, and moderate versus low) from receiver operating characteristic curves using forced-choice classification, where signature responses were compared for two conditions tested within the same participant (the higher was chosen as more disgust) and is therefore 'threshold free'. Two-sided binomial tests were used to test whether the classification accuracies were higher than chance-level (50%).

Additionally, to test how well the VIDS predicted disgust on datasets other than the discovery cohort, we applied the VIDS to the validation and generalization cohorts (See 'Generalization cohort' section below) to obtain a pattern response for each map (that is, the dot-product of each vectorized univariate GLM-derived fMRI activation map with the VIDS weight map).

Generalization cohort

A previous study measured disgust experience and regulation which included disgust and neutral stimuli (study 3[49]). In line with the aim of the current study, we focused on the disgust experience condition. Briefly, 26 participants (10 females; mean±SD age=21.73±1.69 years) performed a disgust experience task where they were required to pay attention to the pictures shown on the

screen. There were 10 blocks (5 disgust and 5 neutral blocks), and each block encompassed 3 consecutive pictures (2s each). After each block, the participants were required to rate how negative they felt. Given that the ratings were averaged across several pictures, we focused on the classification of disgust versus neutral to increase statistical power. See 'Supplementary Methods' for the MRI acquisition details.

Within-subject trial-wise prediction

Here we tested whether the VIDS could predict individual trial-by-trial subjective disgust. To this end, we performed a single-trial analysis[15,79,80], which was achieved by constructing a GLM design matrix with separate regressors for each stimulus. Each task regressor was convolved with the canonical hemodynamic response function. Nuisance regressors and high-pass filter were identical to the above GLM analyses. When using a single-trial analysis, it is worth noting that the regressor (i.e., beta estimates) for a given trial could be strongly contaminated by artifacts (e.g., participant's head movement, scanner pulse artifacts, etc.) that co-occur during that trial. Given this, for each subject, variance inflation factors (a measure of design-induced uncertainty due to collinearity with nuisance regressors) were calculated for each regressor, based on whether it can be predicted by a combination of the others. And any trials with variance inflation factors over 3 were excluded from subsequent analyses (overall ~3.7% of trials in the discovery cohort as well as ~3.4% of trials in the validation cohort were excluded). Next, these single-trial beta maps were used in VIDS pattern expression analysis (i.e., the dot-product of vectorized activation images with the VIDS weights). For subjects in the discovery cohort, a 10×10-fold cross-validation procedure was employed to obtain the VIDS response of each single-trial beta map for each subject.

Determining brain regions associated with and predictive of subjective disgust

A series of analyses were employed to identify neural substrates underlying subjective experience of disgust. Firstly, we conducted bootstrap tests to determine brain areas that made stable contributions to the prediction across participants, where we constructed 10,000 samples (with replacement) from the discovery cohort, repeated the prediction process with each bootstrap sample, and evaluated *z*-scores and two-tailed uncorrected *P* values at each voxel based on the mean and standard deviation of the bootstrap distribution, on the population-level disgust-predictive pattern (i.e., the VIDS). After correction for multiple comparisons (FDR $q<0.05$), those regions with significant and consistent model weights were obtained[56]. Secondly, as in previous studies that computed model encoding maps[15,38], we transformed the population-level pattern (i.e., VIDS, also backward model) into 'activation pattern' (forward model), which maps each voxel to the response (fitted values) in the multivariate model[50]. The reconstructed 'activation pattern' is known as 'structure coefficient' in the statistical literature[81,82]. And the one-sample *t* test, thresholded at FDR $q<0.05$ corrected for multiple comparisons, identified reliable disgust predictive and associative brain regions of the VIDS. Core regions that were viewed as the most consistently correlated with and predictive of the target outcome (i.e., subjective disgust) were defined as having voxels that were consistently observed across backward (that is, model predictive weights) and forward (that is, model encoding voxels in which the prediction correlates with fMRI activation) models (for the methodology detail of determining regions involved in the subjective experience of disgust on the individual-level, see 'Supplementary Methods').

Next, to independently evaluate the neurobiological validity of the developed pattern, we

used the thresholded VIDS (FDR q<0.05, retaining positive values) from the bootstrap test and examined its overlap percentages with the modified 279-region version of the Brainnetome Atlas included additional brainstem, midbrain, and cerebellum regions[58,59]. We then displayed the top 15 regions. Moreover, to further decode the functions correlated with these regions in the thresholded VIDS, functional characterizations of ten exemplary regions as provided by the Brainnetome website (https://atlas.brainnetome.org/bnatlas.html) were reported through probabilistic maps indicating the behavioral domain based on meta data labels of the BrainMap database (http://www.brainmap.org/taxonomy/)[58].

Furthermore, we tested whether disgust processing could be reducible to activations in a single brain region (e.g., insula, amygdala) or network (e.g., the DMN). To examine this hypothesis, we performed whole-brain searchlight (three-voxel radius spheres) – and parcellation (279 cortical, subcortical, and cerebellum regions[58,59]) – based analyses to identify local regions predictive of disgust and compared model performances of local regions with the whole-brain model (i.e., the VIDS). Additionally, we compared the prediction performances of insula and amygdala (based on the 279-region version of the Brainnetome Atlas[58,59]) as well as large-scale networks to the whole-brain model. The networks of interest comprised of seven large-scale cerebral networks[83], a subcortical network (including the striatum, thalamus, hippocampus, and amygdala), and a cortical 'consciousness network'[57]. To reduce potential biases arising from different atlases, we continued to use the modified 279-region version of the Brainnnetome Atlas (which also combined Yeo's seven networks) to extract the nine networks. For these analyses, we trained and tested a model for each searchlight sphere, parcellation, brain region, or network separately using the discovery data (10×10-fold cross-validated).

Comparing the performance of VIDS with the PINES and the VIFS

Previous studies have developed and evaluated whole-brain emotional decoders for general negative emotion experience (picture-induced negative emotion signature, PINES[34]) and subjective fear experience (visually-induced fear signature, VIFS[15]). To compare the performance of VIDS with the PINES and the VIFS, we applied the three decoders to the discovery, validation, PINES holdout test[34] (Study 4, n=182, see Table S1 for details) and VIFS discovery[15] (study 5, n=67, Table S1) cohorts and assessed the overall as well as within-subject prediction−outcome correlations between the pattern expressions and the true ratings. Two-alternative forced-choice classification accuracies between the separate disgust intensity levels (and general negative emotion as well as fear) based on the pattern expressions were further calculated.

Spatial similarity between stable decoding maps and a priori regions of interest as well as networks of interest

River plots were created to illustrate spatial similarity between stable decoding maps derived from bootstrap tests and a prior regions of interest previously documented as regions linking to disgust, fear, and negative affect processes[7,15,34,38]. We further depicted spatial similarity between stable decoding maps and seven resting-state functional networks, the subcortical and consciousness networks. In line with a recent study[38], spatial similarity was computed as cosine similarity between the ROI or network and the thresholded VIDS, VIFS, and PINES (FDR q<0.05, retaining positive values) from bootstrap tests.

Multilevel two-path mediation analysis

To explore the relationship between VIDS response, disgust rating, and PINES response, multilevel two-path mediation analyses were performed using the Mediation Toolbox, available via https://github.com/canlab/MediationToolbox[15,84]. Briefly, the mediation analysis examines whether the observed covariance between the independent/predictor variable (*X*) and the dependent/outcome variable (*Y*) can be explained by the third variable (*M*, also mediator). The predictor-mediator relation, mediator-outcome relation, and predictor-outcome relation before and after controlling for the mediator are characterized by paths *a*, *b*, *c,* and *c'*, respectively. Specifically, the total effect of the predictor on the outcome (path *c*) is the sum of direct/non-mediation effect (path *c'*) and indirect/mediation effect (the product of the path coefficients of path *a* and path *b*, i.e., *a*×*b*). Significant mediation effect is obtained when *a*, *b,* and *a*×*b* are all significant. Furthermore, when *c'* is significant, *M* (i.e., the mediator) is considered to have a partial mediation effect; otherwise, *M* plays a full mediation role. In this study, we constructed two multilevel mediation analyses: (1) the trial-by-trial VIDS responses were entered as predictors (*X*), disgust ratings were entered as outcomes (*Y*), and the trial-by-trial PINES responses were entered as mediators (*M*); (2) the trial-by-trial PINES responses were entered as predictors (*X*), disgust ratings were entered as outcomes (*Y*), and the trial-by-trial VIDS responses were entered as mediators (*M*). To do this, the VIDS and PINES responses were calculated by dot-product of the single-trial beta maps (trials with variance inflation factors over 3 were excluded) with the VIDS and PINES patterns, respectively, for each participant. Bootstrap tests with 10,000 iterations were used to assess the statistical significance of mediation effects. If the bootstrapped 95% CI does not include zero, the effect will be considered to be significant ($P<0.05$). Of note, we also tested whether (1) VIFS response mediated the relationship between VIDS response and disgust ratings; and (2) VIDS response mediated the relationship between VIFS response and disgust ratings.

Validation in social contexts

First, to examine whether the VIDS – developed during the exposure to concrete and physical disgust stimuli (also termed core or physical disgust in the corresponding literature; see[6,7,11]) – could be extended into the sociomoral disgust domain (e.g., unfairness[3,65]), we applied the VIDS pattern to another independent fMRI dataset during which participants were confronted with a series of unfair offers in an Ultimatum Game task (study 6, *n*=43; 'Supplementary Methods' and Table S1). Specifically, we calculated forced-choice classification accuracies between high (average of unfairness level 4 and 5), moderate (unfairness level 3) and low (average of unfairness level 1 and 2) unfairness based on the pattern responses. Noteworthy, disgust is not the only emotion evoked in response to unfairness, other negative emotions might also be involved,[3] we thus compared the performance of PINES and VIFS with VIDS in the sociomoral context. Next, to further examine how specific the VIDS is for unfairness as compared to other social contexts (e.g., pain empathy), we tested the performance of VIDS on another independent dataset (study 7, *n*=238, fMRI brain responses to physical pain and respective non-painful control images as well as to painful faces and respective non-painful control images[44]; Table S1). Pattern response was estimated as explained above.

ACKNOWLEDGMENTS

We thank David Coynel and Dominique J.-F. de Quervain as well as Shengdong Chen for sharing their data. We also thank the CANLab for providing the PINES signature and the PINES holdout dataset. This work was supported by the National Natural Science Foundation of China (Grants No. 32250610208, 82271583), National Key Research and Development Program of China (Grant No. 2018YFA0701400), and the China MOST2030 Brain Project (Grant No. 2022ZD0208500).

AUTHOR CONTRIBUTIONS

X.Y.G., F.Z., and B.B. conceived and designed the experiment. X.Y.G., F.Z., and B.B. analyzed the data and were responsible for interpretation of data. T.X., X.B.L., R.Z., Z.H.Z., X.Y., X.Q.Z., F.W.Y., J.L.L. and R.F.C. provided important suggestions during formal analysis. X.Y.G., R.Z., T.X., and L.W. conducted the experiment. X.Y.G. and R.F.C. were responsible for visualization. X.Y.G. and B.B. drafted the manuscript, F.Z., J.J.Y., and D.Z.Y. provided feedback and revised the manuscript. B.B. supervised the project and acquired the funding. All authors meet the four ICMJE authorship criteria and were responsible for revising the manuscript, approving the final version for publication and for accuracy and integrity of the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Data and code availability

fMRI data are available at https://figshare.com/articles/dataset/Discovery_dataset_disgust/22827974 and https://figshare.com/articles/dataset/validation_dataset_disgust/22841117. Data were analyzed using CANLab neuroimaging analysis tools available at https://github.com/canlab and https://github.com/ganxianyang/fMRI-studies/tree/main/Subjective_disgust_experience_signature.

REFERENCES

1. Rozin, P., Haidt, J. & Fincher, K. From oral to moral. *Science* **323**, 1179-1180 (2009).
2. Rozin, P. & Fallon, A. E. A perspective on disgust. *Psychol. Rev.* **94**, 23-41 (1987).
3. Chapman, H. A., Kim, D. A., Susskind, J. M. & Anderson, A. K. In bad taste: evidence for the oral origins of moral disgust. *Science* **323**, 1222-1226 (2009).
4. Xie, Z., et al. The gut-to-brain axis for toxin-induced defensive responses. *Cell* **185**, 4298-4316.e4221 (2022).
5. Dolensek, N., Gehrlach, D. A., Klein, A. S. & Gogolla, N. Facial expressions of emotion states and their neuronal correlates in mice. *Science* **368**, 89-94 (2020).
6. Vicario, C. M., Rafal, R. D., Martino, D. & Avenanti, A. Core, social and moral disgust are bounded: a review on behavioral and neural bases of repugnance in clinical disorders. *Neurosci. Biobehav. Rev.* **80**, 185-200 (2017).
7. Gan, X., et al. Common and distinct neurofunctional representations of core and social disgust in the brain: coordinate-based and network meta-analyses. *Neurosci. Biobehav. Rev.* **135**, 104553 (2022).
8. Amoroso, C. R., Hanna, E. K., LaBar, K. S., Schaich Borg, J., Sinnott-Armstrong, W. & Zucker, N. L. Disgust theory through the lens of psychiatric medicine. *Clin. Psychol. Sci.* **8**, 3-24 (2019).
9. Jones, D. The depths of disgust. *Nature* **447**, 768-771 (2007).
10. Haidt, J., McCauley, C. & Rozin, P. Individual differences in sensitivity to disgust: a scale sampling seven domains of disgust elicitors. *Pers. Individ. Differ.* **16**, 701-713 (1994).
11. Chapman, H. A. & Anderson, A. K. Things rank and gross in nature: a review and synthesis of moral disgust. *Psychol. Bull.* **139**, 300-327 (2013).
12. Zhang, C., Vincelette, L. K., Reimann, F. & Liberles, S. D. A brainstem circuit for nausea suppression. *Cell Rep.* **39**, (2022).
13. Zhang, C., Kaye, J. A., Cai, Z., Wang, Y., Prescott, S. L. & Liberles, S. D. Area postrema cell types that mediate nausea-associated behaviors. *Neuron* **109**, 461-472.e465 (2021).
14. Becker, B., et al. Fear processing and social networking in the absence of a functional amygdala. *Biol. Psychiatry* **72**, 70-77 (2012).
15. Zhou, F., et al. A distributed fMRI-based signature for the subjective experience of fear. *Nat. Commun.* **12**, 6643 (2021).
16. 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).
17. 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).
18. Reddan, M. C., Wager, T. D. & Schiller, D. Attenuating neural threat expression with imagination. *Neuron* **100**, 994-1005.e1004 (2018).
19. Chapman, H. A. & Anderson, A. K. Understanding disgust. *Ann. N. Y. Acad. Sci.* **1251**, 62-76 (2012).
20. Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V. & Rizzolatti, G. Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* **40**, 655-664 (2003).
21. Stark, R., et al. Erotic and disgust-inducing pictures - differences in the hemodynamic responses of the brain. *Biol. Psychol.* **70**, 19-29 (2005).

22. Schienle, A., Übel, S., Schöngaßner, F., Ille, R. & Scharmüller, W. Disgust regulation via placebo: an fMRI study. *Soc. Cogn. Affect. Neurosci.* **9**, 985-990 (2014).
23. Baumann, O. & Mattingley, J. B. Functional topography of primary emotion processing in the human cerebellum. *Neuroimage* **61**, 805-811 (2012).
24. Schienle, A., Höfler, C., Keck, T. & Wabnegger, A. Neural underpinnings of perception and experience of disgust in individuals with a reduced sense of smell: an fMRI study. *Neuropsychologia* **141**, 107411 (2020).
25. Viol, K., et al. Erroneously disgusted: fMRI study supports disgust-related neural reuse in obsessive-compulsive disorder (OCD). *Front. Behav. Neurosci.* **13**, 81 (2019).
26. Li, J., et al. Common and dissociable contributions of alexithymia and autism to domain-specific interoceptive dysregulations: a dimensional neuroimaging approach. *Psychother. Psychosom.* **88**, 187-189 (2019).
27. Yao, S., et al. Oxytocin modulates attention switching between interoceptive signals and external social cues. *Neuropsychopharmacology* **43**, 294-301 (2018).
28. Ferraro, S., et al. The central autonomic system revisited – Convergent evidence for a regulatory role of the insular and midcingulate cortex from neuroimaging meta-analyses. *Neurosci. Biobehav. Rev.* **142**, 104915 (2022).
29. Bhikram, T., Abi-Jaoude, E. & Sandor, P. OCD: obsessive–compulsive … disgust? The role of disgust in obsessive–compulsive disorder. *J. Psychiatry Neurosci.* **42**, 300 (2017).
30. Shapira, N. A., et al. Brain activation by disgust-inducing pictures in obsessive-compulsive disorder. *Biol. Psychiatry* **54**, 751-756 (2003).
31. Schienle, A., Schäfer, A., Hermann, A. & Vaitl, D. Binge-eating disorder: reward sensitivity and brain activation to images of food. *Biol. Psychiatry* **65**, 654-661 (2009).
32. Benuzzi, F., Lui, F., Duzzi, D., Nichelli, P. & Porro, C. Does it look painful or disgusting? Ask your parietal and cingulate cortex. *J. Neurosci.* **28**, 923-931 (2008).
33. Stark, R., et al. Hemodynamic brain correlates of disgust and fear ratings. *Neuroimage* **37**, 663-673 (2007).
34. 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**, e1002180 (2015).
35. 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).
36. Haxby, J. V. Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage* **62**, 852-855 (2012).
37. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**, 1437-1448 (2015).
38. Čeko, M., Kragel, P. A., Woo, C.-W., López-Solà, M. & Wager, T. D. Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* **25**, 760-770 (2022).
39. Wager, T. D., Krishnan, A. & Hitchcock, E. How are emotions organized in the brain? In The nature of emotion. Fundamental questions (eds Fox, A. S., Lapate, R. C., Shackman, A. J., Davidson, R. J.) (pp. 112-118). Oxford University Press (2018).
40. Weaverdyck, M., Lieberman, M. & Parkinson, C. Multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. *Soc. Cogn. Affect. Neurosci.* **15**, (2020).
41. Kragel, P. A. & LaBar, K. S. Decoding the nature of emotion in the brain. *Trends Cogn. Sci.* **20**, 444-455 (2016).

42. Haynes, J.-D. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* **87**, 257-270 (2015).
43. 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).
44. Zhou, F., et al. Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. *eLife* **9**, e56929 (2020).
45. LeDoux, J. E. & Brown, R. A higher-order theory of emotional consciousness. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2016-E2025 (2017).
46. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **12**, 1-23 (2017).
47. Saarimäki, H., et al. Discrete neural signatures of basic emotions. *Cereb. Cortex* **26**, 2563-2573 (2015).
48. Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P. & Singer, T. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nat. Commun.* **7**, 10904 (2016).
49. Chen, S., et al. Functional decoupling of emotion coping network subsides automatic emotion regulation by implementation intention. *Neural Plast.* **2021**, 6639739 (2021).
50. Haufe, S., et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* **87**, 96-110 (2014).
51. Satpute, A. B. & Lindquist, K. A. The default mode network's role in discrete emotion. *Trends Cogn. Sci.* **23**, 851-864 (2019).
52. Xu, T., et al. Angiotensin Antagonist Inhibits Preferential Negative Memory Encoding via Decreasing Hippocampus Activation and Its Coupling With the Amygdala. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **7**, 970-978 (2022).
53. Fastenrath, M., et al. Human cerebellum and corticocerebellar connections involved in emotional memory enhancement. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2204900119 (2022).
54. Haberkamp, A., Glombiewski, J. A., Schmidt, F. & Barke, A. The DIsgust-RelaTed-Images (DIRTI) database: validation of a novel standardized set of disgust pictures. *Behav. Res. Ther.* **89**, 86-94 (2017).
55. Kragel, P. A., Reddan, M. C., LaBar, K. S. & Wager, T. D. Emotion schemas are embedded in the human visual system. *Sci. Adv.* **5**, eaaw4358 (2019).
56. Kohoutová, L., et al. Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat. Protoc.* **15**, 1399-1435 (2020).
57. 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).
58. Fan, L., et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508-3526 (2016).
59. Lee, J.-J., et al. A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* **27**, 174-182 (2021).
60. Damasio, A. & Carvalho, G. B. The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* **14**, 143-152 (2013).
61. Kirby, L. A. J. & Robinson, J. L. Affective mapping: an activation likelihood estimation (ALE) meta-analysis. *Brain Cogn.* **118**, 137-148 (2017).
62. Vytal, K. & Hamann, S. Neuroimaging support for discrete neural correlates of basic emotions:

a voxel-based meta-analysis. *J. Cogn. Neurosci.* **22**, 2864-2885 (2010).

63. Costafreda, S. G., Brammer, M. J., David, A. S. & Fu, C. H. Y. Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res. Rev.* **58**, 57-70 (2008).
64. Touroutoglou, A., Lindquist, K. A., Dickerson, B. C. & Barrett, L. F. Intrinsic connectivity in the human brain does not reveal networks for 'basic' emotions. *Soc. Cogn. Affect. Neurosci.* **10**, 1257-1265 (2015).
65. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755-1758 (2003).
66. 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).
67. Kyzar, E. J. & Denfield, G. H. Taking subjectivity seriously: towards a unification of phenomenology, psychiatry, and neuroscience. *Mol. Psychiatry* **28**, 10-16 (2023).
68. Pujol, J., et al. Mapping the sequence of brain events in response to disgusting food. *Hum. Brain Mapp.* **39**, 369-380 (2018).
69. Qi, S., Hassabis, D., Sun, J., Guo, F., Daw, N. & Mobbs, D. How cognitive and reactive fear circuits optimize escape decisions in humans. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3186-3191 (2018).
70. 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).
71. Clore, G. L. & Ortony, A. Appraisal theories: how cognition shapes affect into emotion. In Handbook of emotions, 3rd ed. (eds Lewis, M., Haviland-Jones, J. M., Barrett, L. F.) (pp. 628-642). The Guilford Press (2008).
72. Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B. & Barrett, L. F. A Bayesian model of category-specific emotional brain responses. *PLoS Comput. Biol.* **11**, e1004066 (2015).
73. 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).
74. Lamm, C., Decety, J. & Singer, T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54**, 2492-2502 (2011).
75. Russell, P. S. & Giner-Sorolla, R. Bodily moral disgust: what it is, how it is different from anger, and why it is an unreasoned emotion. *Psychol. Bull.* **139**, 328-351 (2013).
76. Marchewka, A., Żurawski, Ł., Jednoróg, K. & Grabowska, A. The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Beh. Res. Meth.* **46**, 596-610 (2014).
77. Lang, P., Bradley, M. & Cuthbert, B. *International Affective Picture System (IAPS): affective ratings of pictures and instruction manual*. Technical Report A-8. University of Florida, Gainesville (2008).
78. Zhou, F., et al. Shifted balance of dorsal versus ventral striatal communication with frontal reward and regulatory regions in cannabis-dependent males. *Hum. Brain Mapp.* **39**, 5062-5073 (2018).
79. Atlas, L. Y., Lindquist, M. A., Bolger, N. & Wager, T. D. Brain mediators of the effects of noxious heat on pain. *Pain* **155**, 1632-1648 (2014).
80. Büchel, C., Bornhovd, K., Quante, M., Glauche, V., Bromm, B. & Weiller, C. Dissociable neural

responses related to pain intensity, stimulus intensity, and stimulus awareness within the anterior cingulate cortex: a parametric single-trial laser functional magnetic resonance imaging study. *J. Neurosci.* **22**, 970-976 (2002).

81. Courville, T. & Thompson, B. Use of structure coefficients in published multiple regression articles: β is not enough. *Educ. Psychol. Meas.* **61**, 229-248 (2001).
82. Thompson, B. & Borrello, G. M. The importance of structure coefficients in regression research. *Educ. Psychol. Meas.* **45**, 203-209 (1985).
83. Yeo, B. T., et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125-1165 (2011).
84. Woo, C.-W., et al. Quantifying cerebral contributions to pain beyond nociception. *Nat. Commun.* **8**, 14211 (2017).