nature human behaviour

Article

https://doi.org/10.1038/s41562-024-01868-x

A neurofunctional signature of subjective disgust generalizes to oral distaste and socio-moral contexts

Received: 29 May 2023

Accepted: 19 March 2024

Published online: 19 April 2024

Check for updates

Xianyang Gan[1,2,10], Feng Zhou[3,10], 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] & Benjamin Becker[1,2,8,9]

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 socio-moral contexts. Here, in a series of studies, we combined functional magnetic resonance imaging with machine-learning-based predictive modelling to establish a comprehensive neurobiological model of subjective disgust. The developed 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 = 34 and n = 26), gustatory distaste (n = 30) and socio-moral (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 functional magnetic resonance imaging signature for disgust with a high potential to resolve ongoing evolutionary debates.

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 behaviours 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[1–4]. In response to bitter food, the distaste

[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, 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, China. [3]Faculty of Psychology, Southwest University, Chongqing, China. [4]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada. [5]Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands. [6]Sichuan Key Laboratory of Psychology and Behavior of Discipline Inspection and Supervision, Institute of Brain and Psychological Sciences, Sichuan Normal University, Chengdu, China. [7]Max Planck School of Cognition, Leipzig, Germany. [8]State Key Laboratory for Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China. [9]Department of Psychology, The University of Hong Kong, Hong Kong, China. [10]These authors contributed equally: Xianyang Gan, Feng Zhou. e-mail: ben_becker@gmx.de

mechanism directly triggers defensive reactions such as withdrawal and retching and distinct facial expressions across (some) mammalian species via brainstem circuits[5] or the posterior insular cortex[6], respectively (for the disgust output system[1], see also Fig. 1a). In humans, the distaste response was pre-adapted during evolution into core disgust, which is essentially characterized by the highly aversive feeling of revulsion in response to contaminated, infectious or poisonous stimuli, in turn, facilitating pathogen avoidance[7–10] (thus also termed as pathogen disgust according to Tybur et al.[11,12]). Core disgust is distinct from the distaste reaction given that the experience of revulsion can be evoked in the absence of actual sensory stimuli such as bitter taste and is strongly shaped by conscious appraisals (for example, food being rejected on the basis of where it might have been)[10]. Subsequently, core disgust may have been pre-adapted into the social context with revulsion being elicited by socio-moral transgressions[1,2,4,13], which may facilitate avoidance of potentially harmful interactions (socio-moral disgust)[14]. The pre-adaptation processes of disgust described above are further reflected in and supported by adaptationist perspectives[11,12] that posit that disgust has evolved in response to pathogen avoidance and was subsequently co-opted to serve adaptive purposes in the realm of socio-moral disgust.

While several neurobiological models on the exact brain mechanisms of the distaste reaction have been established in animal models[5,15,16], 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[8,17–21]. 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 (for example, fear) and whether the same mechanisms mediate the aversive experience in the gustatory distaste response or in socio-moral 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[7,8,22,23]; however, several studies did not observe insula engagement in response to disgust stimuli[24–28], and recent evidence suggests a broader role of the insula in interoceptive[29,30], central autonomic[31] and individual-level emotional information processing[32,33]. Subcortical regions like the amygdala and basal ganglia have also been considered as key disgust regions[8,22,34], probably due to their roles in early threat detection and defensive motor responses[8,17,18]. However, many neuroimaging studies failed to observe an involvement of the amygdala[26–28,35–37] while a recent meta-analysis did not confirm robust engagement of the basal ganglia[8] 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[38] 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[39], 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[40,41].

Recent methodological progress combined fMRI with machine-learning-based multivariate pattern analysis (MVPA) to develop more comprehensive and accurate neurofunctional models of internal mental processes, including sensitive and generalizable neuroaffective signatures for subjective emotional experiences (for example, negative affect or fear)[18,39,42,43] as compared with conventional brain imaging models. 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[44]. Briefly, MVPA examines neural responses across multiple brain regions and voxels[40,45–47], thus allowing to determine neural representations of mental processes with higher resolution and considerably larger effect sizes[40,48] than the conventional mass-univariate approach. The resultant fMRI-based neurofunctional signatures enabled to separate the subjective emotional experience from hard-wired defensive and physiological responses[18,20] and to explore common and separable neural representations of aversive experiences across domains[43,49]. 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[50,51]. Despite long-standing discussions on the neurobiological and cultural evolution of disgust and increasing interest in the role of disgust in a number of mental and neurological disorders (for example, obsessive–compulsive disorder, eating disorders and Huntington's disease)[7,9,34], as well as initial studies demonstrating the promise of MVPA for regional or categorical disgust decoding[52,53], an accurate and generalizable whole-brain signature predictive of momentary self-reported subjective disgust experience is currently lacking.

In this Article, we capitalized on recent methodological advances in MVPA-based neural decoding techniques to determine (1) whether 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) cohort that generalizes to data from an independent research team using a different disgust induction paradigm and MRI system (n = 26)[54] (Fig. 1b,c) as well as a modified disgust induction paradigm that uncouples neural activity related to the emotional experience and motor responses (n = 34); (2) whether 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 (Fig. 1d); (3) the neural representations underlying disgust experience utilizing forward (association), backward (prediction) encoding models[55] and isolated systems predictions (notably, the MVPA approach provides the opportunity to test the performance of local systems versus distributed whole-brain activity pattern[56]) (Fig. 1e). The developed visually induced disgust signature (VIDS) was next utilized to determine (4) the functional specificity of the subjective disgust signature by conducting predictive and topographical comparisons with established decoders for subjective fear[18] and negative affect[39], respectively (Fig. 1f); (5) whether the visual disgust decoder generalizes to the gustatory modality by applying the VIDS to data acquired while participants tasted (as well as swallowed) either disgusting or neutral liquids (n = 30) (Fig. 1g); (6) 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 (socio-moral disgust, n = 43) and in response to the suffering of other individuals in a pain empathy paradigm[49] (n = 238) (Fig. 1h). Finally, control analyses were employed in two additional independent datasets that examined neural reactivity in response to positive and negative visual emotional stimuli[57,58] (n = 30 and n = 150, 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 and to determine the extent of neurofunctional similarity with the experience of other negative emotions (fear and negative affect), as well as with the experience of distaste in the gustatory modality and unfairness in social situations (commonly referred to as socio-moral disgust) hypothesized in the disgust evaluation model (Fig. 1a).

https://doi.org/10.1038/s41562-024-01868-x

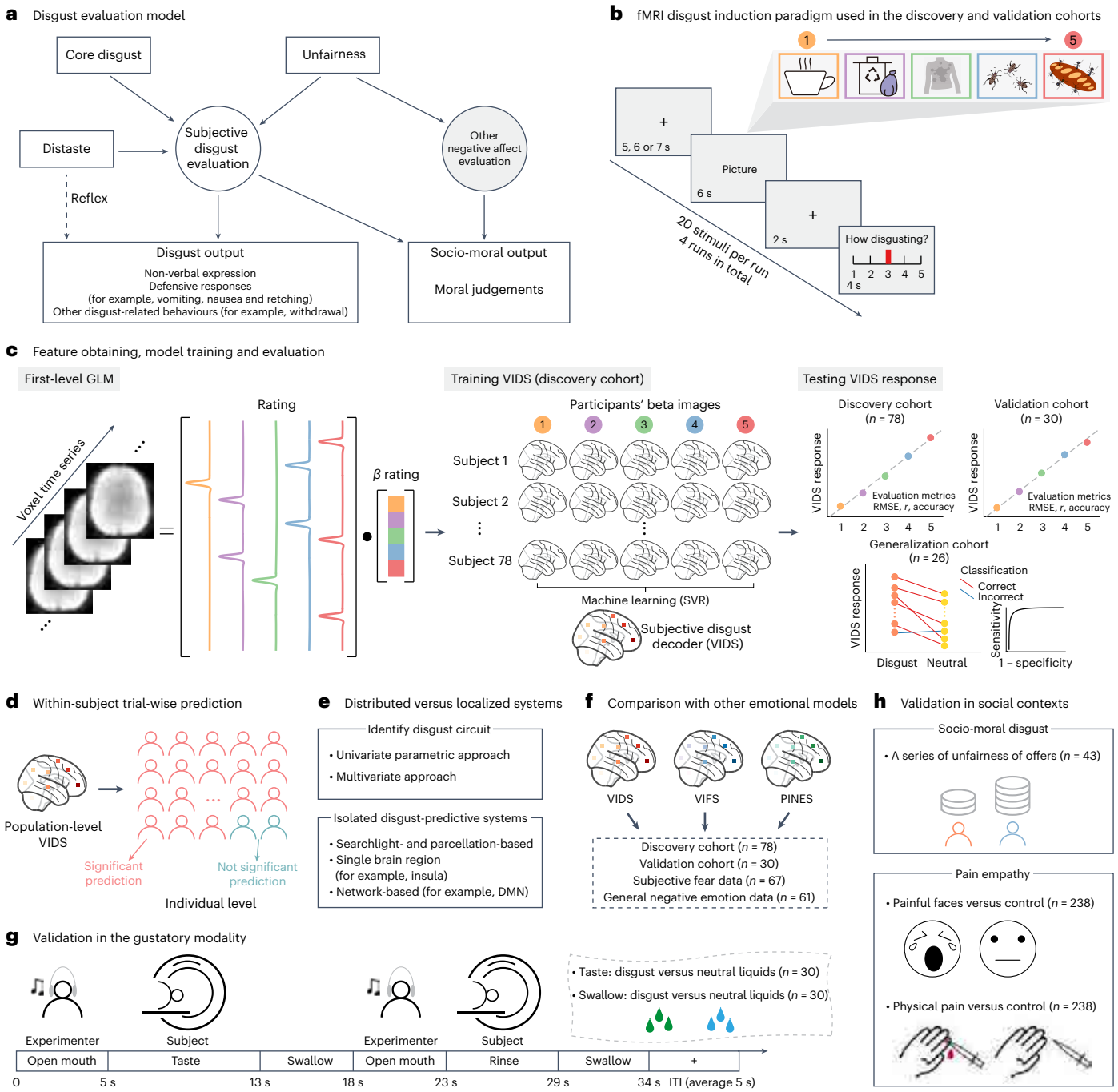


Fig. 1 | Disgust evaluation model, task design and analytic workflow. **a**, A disgust evaluation model proposed on the basis of previous studies[1,14]. **b**, Discovery and validation cohorts underwent the identical disgust induction fMRI paradigm. Of note, schematic figures were used for display purposes only to avoid copyright issues and were not included in the original stimulus set. **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 with subjective fear or general negative affect. **g**, Validation of disgust in the gustatory modality. The experimental flowchart depicts the gustatory procedures. The dotted region at the top right indicates validations in the gustatory modality performed for both the taste and swallow of disgust versus neutral liquids. **h**, Validation of disgust in social contexts. The VIDS was applied to two social contexts, that is, socio-moral disgust (unfairness) and pain empathy induced by physical (limb) and affective (face) pain-displaying stimuli. ITI, inter-trial intervals; VIDS, visually induced disgust signature developed in this study; VIFS, visually induced fear signature from Zhou et al.[18]; PINES, picture-induced negative emotion signature from Chang et al.[39]. See Methods and Supplementary Table 1 for further details.

Results

Subjective disgust experience induced by visual stimuli

We initially determined whether the visual disgust stimuli evoked different levels of subjective disgust experience. During fMRI, disgust was induced by means of validated disgust-specific affective pictures[59] (DIsgust-RelaTed-Images (DIRTI) disgust images database; for details, see 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). Eighty stimuli were presented and induced the entire intensity range of subjective disgust balanced across the most common disgust-related contexts (animal, human and scene; Extended Data Fig. 1). 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 five levels of subjective disgust (6 reported levels 1–4).

A brain signature for subjective disgust experience (VIDS)

We developed a brain signature sensitive to predict subjective disgust experience. Consistent with previous studies developing neuroaffective decoders[18,20,39], we employed a linear support vector regression (SVR) model to identify a whole-brain signature of fMRI activation predictive of subjective disgust experience (Fig. 2a and Extended Data Fig. 2). Notably, previous studies (for example, refs. 39,60) have shown that linear SVR exhibits nearly identical prediction performance to linear models with regulations and feature reduction. 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, five or four pairs of scalar values per subject: r = 0.88 ± 0.02 (mean ± s.e.m.); average root mean squared error (RMSE) of 1.43 ± 0.08; overall, that is, 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; Fig. 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; Fig. 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, and two-sided independent binomial tests were used to test whether the classification accuracies were higher than chance level (50%). 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 (n = 78 in the discovery cohort, n = 30 in the validation cohort, P < 0.001, Cohen's d = 1.22–1.96, 86–96% sensitivity (95% CI 78–100), 86–96% specificity (95% CI 77–100)) and high versus low with 99–100% accuracy in both cohorts (n = 78 in the discovery cohort, n = 30 in the validation cohort, P < 0.001, Cohen's d = 2.40–2.92, 99–100% sensitivity (95% CI 96–100), 99–100% specificity (95% CI 96–100)). Moreover, the VIDS response could distinguish each successive pair of disgust rating levels (for example, 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 = 0.008) (Fig. 2b,c).

Given that previous studies emphasized the contribution of the visual cortex in decoding emotional content[8,61], we re-trained 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 and Extended Data Fig. 3).

The interval between the picture presentation stage and the rating (button press) was fixed in the original disgust paradigm (Fig. 1b), which raises the possibility that the VIDS might partly decode the mapping of disgust onto motor responses. We therefore modified the disgust induction paradigm by including an additional jitter between the picture presentation and motor responding stage and further randomizing the order of rating 1 to rating 5 per trial (Supplementary Methods and Extended Data Fig. 4a). Applying the VIDS to the newly collected dataset (study 3; Supplementary Table 1) reveals that the disgust signature captures the experience of disgust rather than the mapping of disgust onto responses, such that the disgust signature yielded comparably high prediction–outcome correlations (within-subject r = 0.87 ± 0.02; overall prediction–outcome r = 0.54) and could further predict different levels of disgust experience with high accuracy (for details, see Supplementary Results and Extended Data Fig. 4b).

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 haemodynamic model, the VIDS response started approximately 4 s following the picture onset and increased during 6–12 s (Fig. 2d,e), suggesting that the VIDS captures disgust experience rather than expectation (pre-stimulus) or cognitive evaluation (response reporting) of disgust (see also Supplementary Results and Extended Data Fig. 4c).

Generalization of the disgust signature

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[54] (study 4; Supplementary Table 1) that examined emotional reactivity to disgusting stimuli. The VIDS classified disgust versus neutral with high accuracy (n = 26, accuracy 92% (±5.2% s.e.m.), area under the curve (AUC) 0.99, Cohen's d = 1.46, sensitivity 92% (95% CI 81–100), specificity 92% (95% CI 79–100), two-sided binomial test P < 0.001; Fig. 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[38,51]. 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; two-sided) 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 (Fig. 2g,h). 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[38,51], the VIDS could accurately and robustly predict the intensity of momentary disgust experience on the individual level across two independent populations.

https://doi.org/10.1038/s41562-024-01868-x

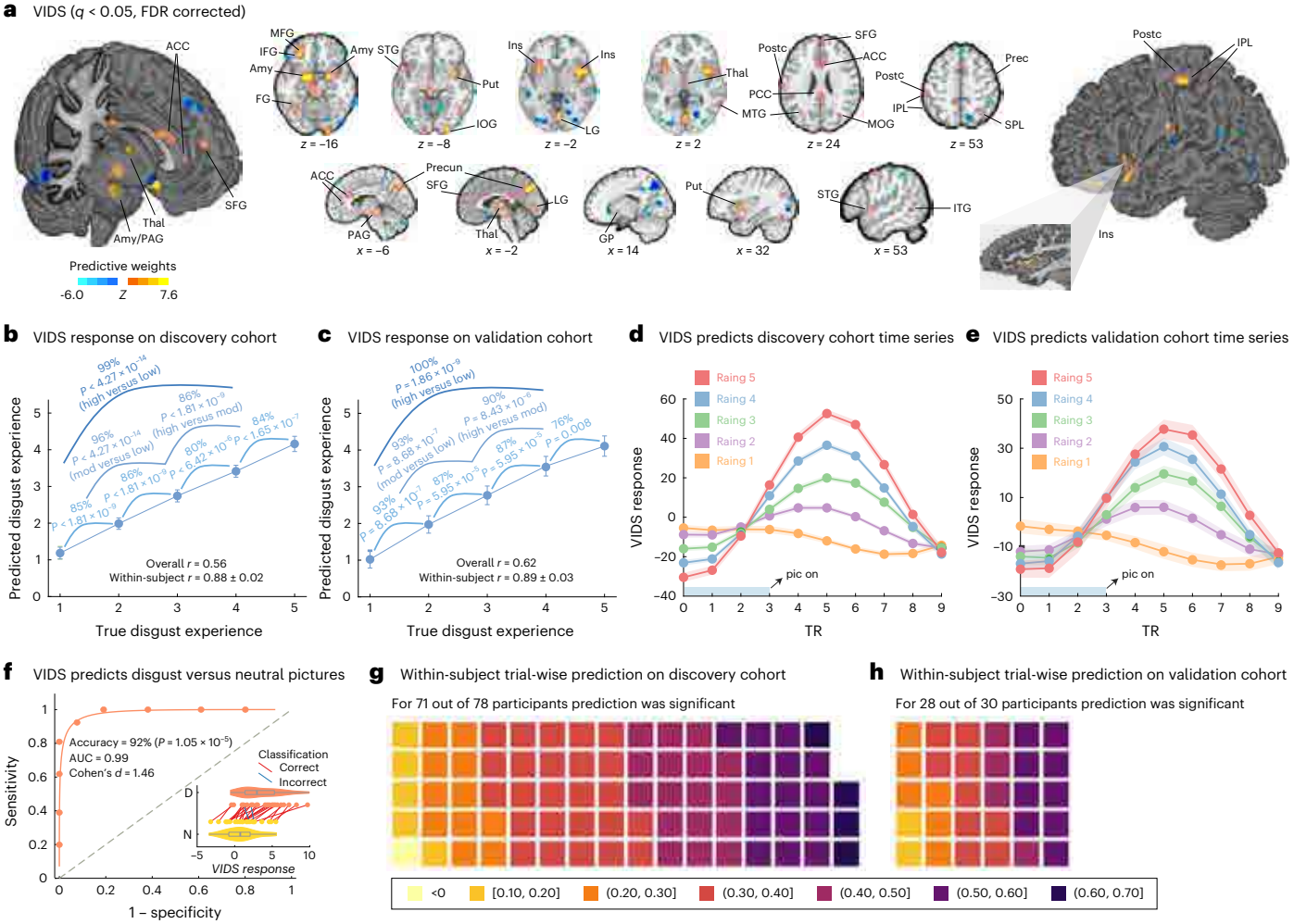


Fig. 2 | VIDS. a, The VIDS pattern thresholded at $q < 0.05$ (FDR corrected, 10,000-sample bootstrap). Extended Data Fig. 2 shows the topography of the unthresholded pattern in anatomical ROIs previously associated with disgust and emotional experience. **b**,**c**, The predicted disgust experience (subjective ratings; mean ± s.e.m.) compared with actual disgust ratings in the cross-validated discovery cohort (**b**; $n = 78$) and the independent validation cohort (**c**; $n = 30$), respectively. Accuracy provided for forced-choice comparisons. r indicates the Pearson correlation coefficient between predicted and true ratings. **d**,**e**, Averaged peristimulus plot (mean ± s.e.m.) of the VIDS response in the cross-validated discovery cohort (**d**) and the independent validation cohort (**e**), respectively, at every repetition time (TR; 2 s) for each disgust intensity rating separately. pic, picture. **f**, Generalization of the VIDS in an independent dataset (using two-alternative forced-choice to classify disgust versus neutral stimuli; $n = 26$). The figure shows the receiver operating characteristics (ROC), and the inset 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). The box plots in the inset for each condition present median (centre line), the 25th and 75th percentiles (lower and upper hinges) and 1.5 × interquartile range (IQR) (whiskers). **g**,**h**, The range of the within-subject trial-wise prediction–outcome correlation coefficients for each participant in discovery (**g**) and validation (**h**) cohorts, respectively; note that the grids represent the prediction performance on the individual level, with each cell representing prediction performance (in terms of prediction–outcome correlation) for an individual participant. The P values are based on two-sided independent binomial tests (uncorrected). The error bars and shaded regions indicate the s.e.m. Amy, amygdala; 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; Postc, postcentral gyrus; Prec, precentral gyrus; Precun, precuneus; Put, putamen; SFG, superior frontal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus; Thal, thalamus.

Regions associated with and predictive of subjective disgust

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 ($q < 0.05$, false discovery rate (FDR) corrected)[62]. Given that some brain features could contribute to controlling for noise in the data[55] rather than the emotional process per se, we next transformed the population-level VIDS into 'activation pattern' ('structure coefficient'; for details, see Methods). The results showed that a set of distributed subcortical and cortical brain systems exhibited significant model weights ($q < 0.05$, FDR corrected; Fig. 3a) and structure coefficients ($q < 0.05$, FDR corrected; Fig. 3b), including the bilateral amygdala, anterior and mid- and posterior insular cortex, hippocampus, dorsal anterior cingulate cortex (ACC), periaqueductal grey (PAG), thalamus and putamen as well as the right midcingulate cortex (MCC), left posterior cingulate cortex (PCC), left inferior frontal gyrus, left superior temporal gyrus, right precentral gyrus, and occipital and supplementary motor areas (Fig. 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, Supplementary Results and Extended Data Fig. 5). 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

https://doi.org/10.1038/s41562-024-01868-x

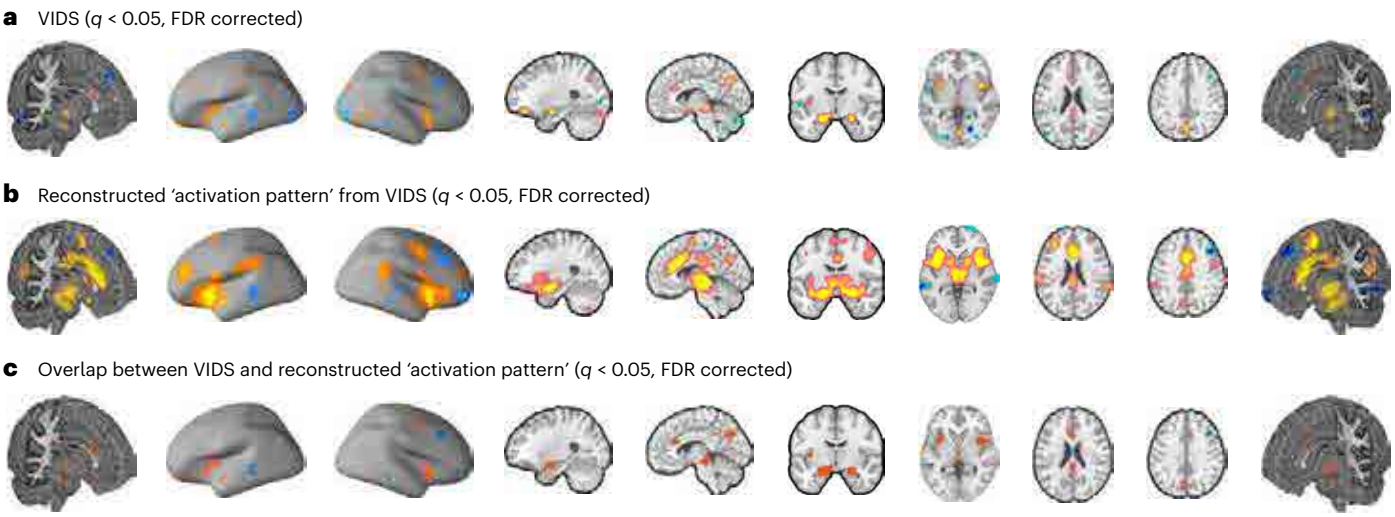


Fig. 3 | Subjective disgust experience is associated with and predicted by distributed brain regions. a, Thresholded VIDS. **b**, Thresholded transformed VIDS 'activation pattern'. **c**, The conjunction between VIDS and transformed 'activation pattern'. Images thresholded at $q < 0.05$, FDR corrected. Hot colour indicates positive weights (**a**) or associations (**b**), whereas cold colour indicates negative weights (**a**) or associations (**b**).

threats (amygdala, PAG, thalamus and putamen)[8,18], regions involved in interoceptive awareness, self-referential processing and explicit emotional processing such as the anterior insula, dorsal ACC and frontal regions[29,30,44,63], as well as regions that have been identified as core systems mediating the bitter taste reaction in animal models (posterior insula[6] and brainstem[5]). Negative associations with disgust were most consistently identified in superior frontal and middle temporal regions.

Evaluating the neurobiological validity of the signature

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 Brainnetome Atlas[64,65] (Methods). The top 15 atlas regions visualized in Fig. 4a included four amygdala subregions, five insula subregions, orbital gyrus, thalamus, rostral hippocampus, rostroventral cingulate gyrus, and brainstem and midbrain regions involved in defensive responses and conscious emotional experience (that is, PAG and substantia nigra[66]). Meta-analytic behavioural 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 (Fig. 4b,c). See Supplementary Results, Supplementary Table 2 and Extended Data Fig. 6 for the meta-analytic decoding results of the unthresholded VIDS pattern based on Neurosynth, which also supported the disgust signature as a biologically plausible model of disgust.

VIDS outperforms prediction based on local systems

Given the continuing debate on the contribution of specific brain regions, such as insula[8,23,67,68], amygdala[8,69] and the default mode network (DMN)[70,71], 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 Fig. 5a,b, subjective disgust experience could be significantly predicted by activations in widely distributed regions ($P < 0.001$, averaged across 10×10-fold cross-validation procedure). However, none of the local models predicted subjective disgust to the extent the VIDS did (see Extended Data Fig. 7a,b for predictions of models trained on discovery cohort in the 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 and (5) each of seven large-scale cerebral networks (Methods). The 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) and other brain networks (Fig. 5c–e, Extended Data Fig. 7c,d,e and Supplementary Table 3) could, to some extent, predict subjective disgust experience. Nonetheless, although statistically significant ($P < 0.001$), 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, Supplementary Results and Extended Data Fig. 8 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 (that is, the whole-brain model contains many 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 large-scale cerebral networks (averaged over 1,000 iterations)[18,40]. The asymptotic prediction when sampling from all brain systems as we did with the VIDS (black line in Fig. 5e and Extended Data Fig. 7e) was substantially higher than the asymptotic prediction within individual networks (coloured lines in Fig. 5e and Extended Data Fig. 7e; see also Supplementary Table 3 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 (that is, 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 (Extended Data Fig. 7e), indicating that models trained on approximately 10,000 randomly sampled voxels were robust enough (for similar findings, see ref. 18). Together, converging lines of evidence from the above systematic analyses point to the fact

https://doi.org/10.1038/s41562-024-01868-x

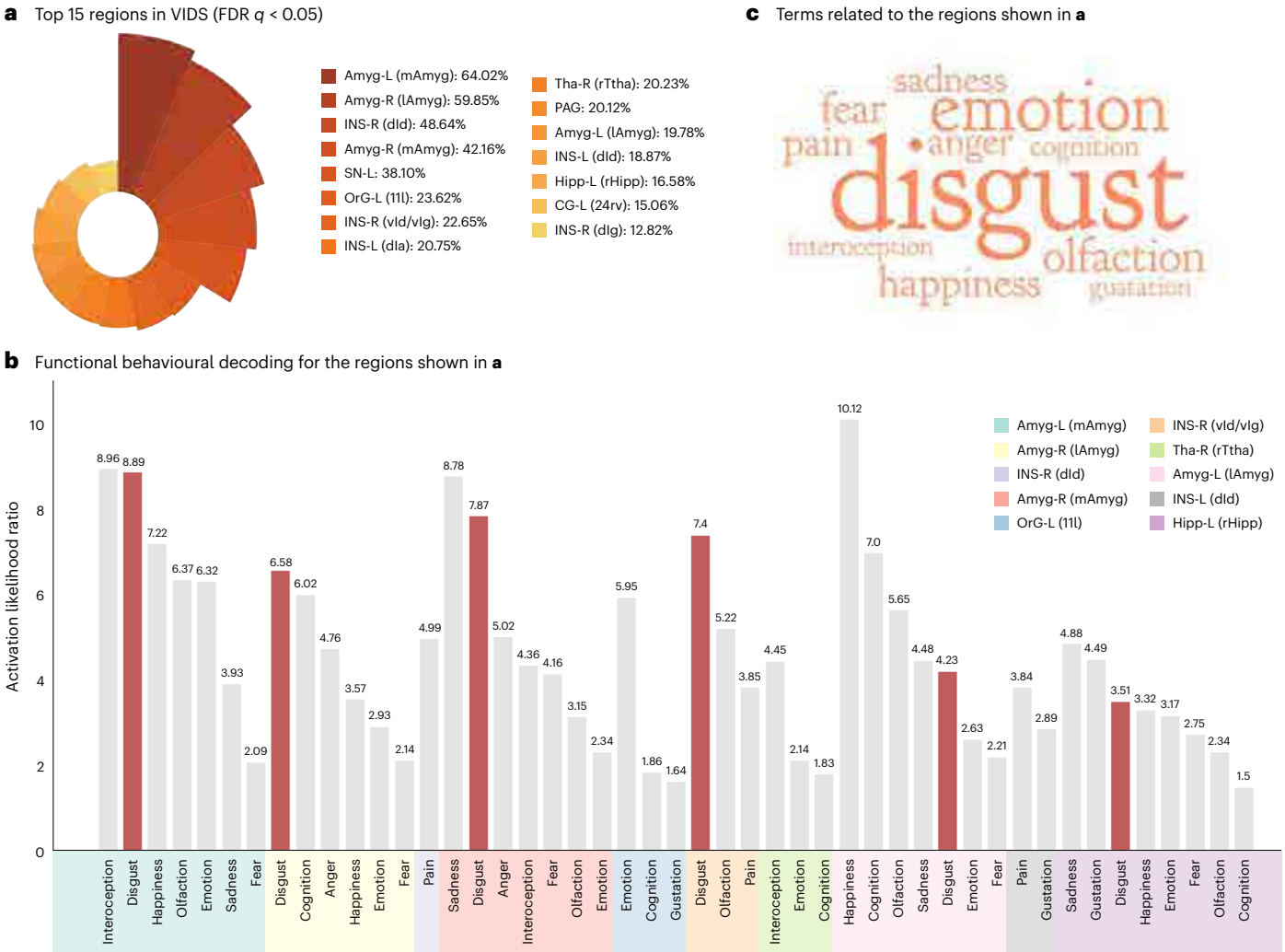


Fig. 4 | Neurobiological validity of the thresholded VIDS. a, The top 15 atlas regions overlapping with the thresholded VIDS (retaining positive values). **b**, Functional behavioural decoding for ten exemplary regions for the thresholded VIDS (provided by the Brainnetome website). **c**, Meta-analytic behavioural 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.

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.

Separable signatures of disgust, fear and negative affect

We next determined to what extent the subjective experience of disgust, fear and non-specific negative affect are encoded in shared or distinguishable neurofunctional representations. Disgust is inherently inter-related with general negative affect and functionally similar avoidance responses like fear[8]. 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 regions of interest (ROIs) as well as networks. Amygdala, anterior insula, basal ganglia, lateral prefrontal cortex (lPFC) and MCC contained stable predictive voxels in all three models, but VIDS showed larger contributions to the first four ROIs than visually induced fear signature (VIFS) and picture-induced negative emotion signature (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 (Fig. 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 more strongly to VIDS than to VIFS and PINES whereas the somatomotor network strongly contributed to PINES (Fig. 6b and Extended Data Fig. 9a). Second, we examined functional similarities between the VIDS and PINES[39] (general negative emotion experience) and VIFS[18] (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 (Fig. 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 (Fig. 6c). Particularly, additional analyses revealed that the three decoders were rather specific for their targeted emotion during discrimination at moderate versus low intensity (for example, 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 (Fig. 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 (Supplementary Table 4).

https://doi.org/10.1038/s41562-024-01868-x

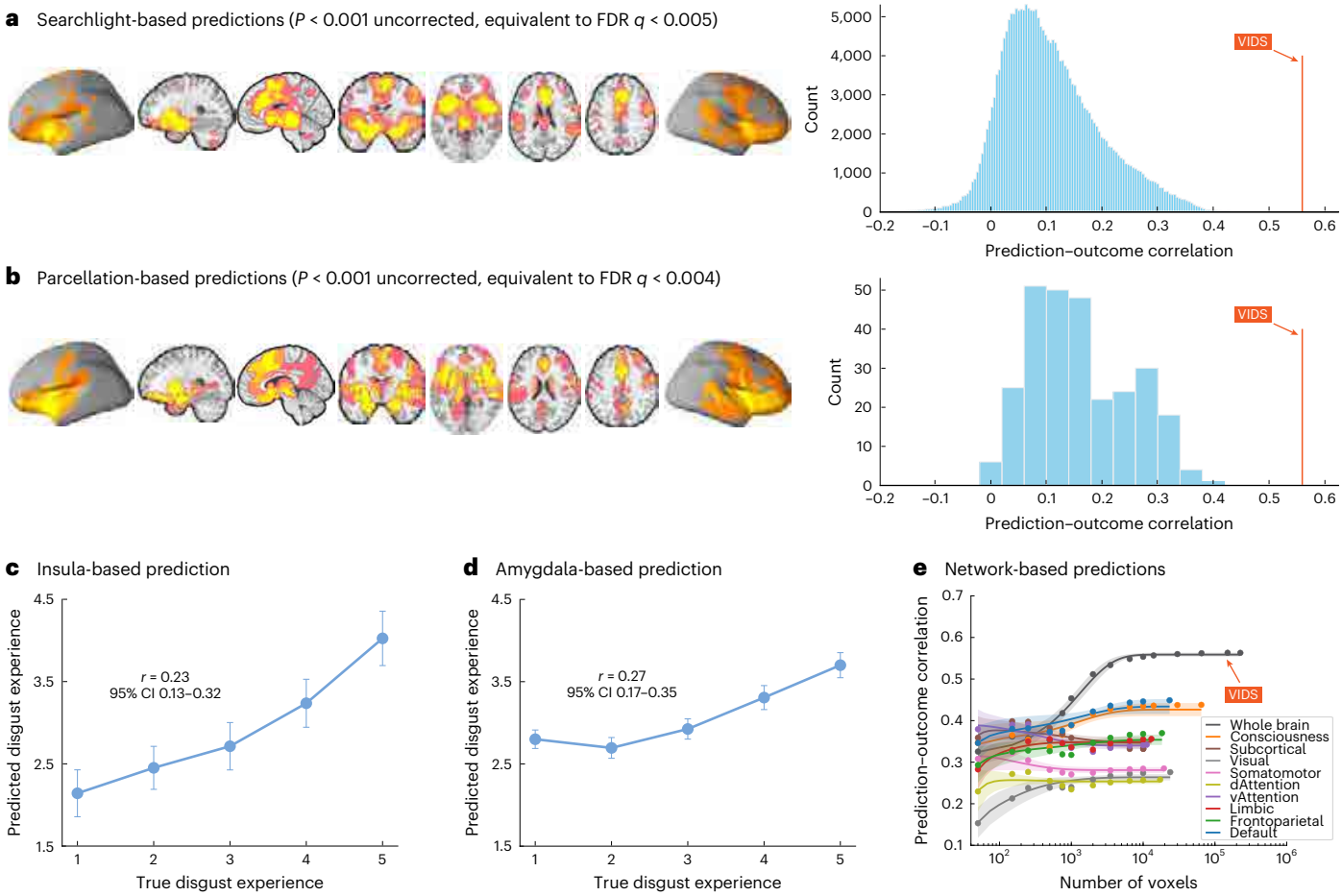


Fig. 5 | Local brain region and network predictions in the discovery cohort. **a**,**b**, Brain regions that significantly predict subjective disgust experience revealed by searchlight (**a**) and parcellation-based (**b**) analyses, respectively. Statistical significance was evaluated by prediction–outcome correlation (Pearson; two-sided). 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- (**a**) and parcellation-based (**b**) predictions, respectively) are displayed. Histograms: cross-validated predictions (correlations) from local searchlights (**a**) or parcellations (**b**). The orange lines indicate the prediction–outcome correlation from VIDS. **c**,**d**, Cross-validated predictions (mean ± s.e.m.) from insula- (**c**) and amygdala-based (**d**) prediction analyses, respectively. The error bar indicates the s.e.m.; *r* indicates overall (between- and within-subjects; that is, $n = 384$ pairs) prediction–outcome Pearson correlation coefficient. **e**, Subjective experience of disgust is distributed across multiple systems. The model performance was evaluated as increasing numbers of voxels/features (*x* axis) were selected to predict subjective disgust in different ROIs including the entire brain (black), consciousness network (light orange), subcortical network (brown) or individual large-scale cerebral networks (other coloured lines). The *y* axis denotes the cross-validated prediction–outcome correlation. Coloured dots indicate the mean correlation coefficients, solid lines indicate the mean parametric fit, and shaded regions indicate the s.d.

Finally, we employed multi-level mediation models, which examined whether the covariance between two variables (*X* and *Y*) can be explained by a third variable (*M*), 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 (Fig. 6e). In contrast, the PINES response failed to mediate the effect of VIDS response on disgust ratings (Fig. 6f). VIFS could predict subjective disgust (discovery cohort: $r = 0.35$; validation cohort: $r = 0.32$), and corresponding multi-level mediation models revealed that VIDS partially mediates the effect of VIFS response on disgust ratings and vice versa (Extended Data Fig. 9b,c); however, the first model (Extended Data Fig. 9b) had three times larger effect sizes than the second one (Extended Data Fig. 9c) 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 that may shape distinguishable subjective experiences.

Validating the disgust signature in oral distaste contexts

We examined that the visual subjective disgust signature predicts gustatory stimulus induced (taste) disgust. From an evolutionary perspective, core disgust is considered to have evolved from the ancestral oral distaste response rooted in chemosensory rejection[1–3,11,14]. To test whether the VIDS generalizes to other modalities and whether the hypothesized link between distaste and subjective disgust is reflected on the brain level, the VIDS—as well as the PINES and VIFS—was applied to fMRI data acquired while participants tasted (as well as swallowed) either disgust or neutral liquids (study 7, $n = 30$; Supplementary Methods and Supplementary Table 1). As shown in Fig. 7a, during the taste period, the VIDS could discriminate disgust taste from neutral

taste with high accuracy (n = 30, accuracy 80% (±7.3% s.e.m.), Cohen's d = 0.39, sensitivity 80% (95% CI 63–94), specificity 80% (95% CI 64–94), two-sided binomial test P = 0.001), confirming that the VIDS generalizes to disgust across modalities, while neither the VIFS (n = 30, accuracy 60% (±8.9% s.e.m.), Cohen's d = 0.26, sensitivity 60% (95% CI 42–79), specificity 60% (95% CI 42–78), two-sided binomial test P = 0.362) nor the PINES (n = 30, accuracy 53% (±9.1% s.e.m.), Cohen's d = 0.03, sensitivity 53% (95% CI 36–70), specificity 53% (95% CI 36–71), two-sided binomial test P = 0.856) could significantly predict disgust taste versus neutral taste. Moreover, during the swallow period (Fig. 7b), the VIDS could significantly classify swallowing of disgust versus neutral liquids (n = 30, accuracy 77% (±7.7% s.e.m.), Cohen's d = 0.68, sensitivity 77% (95% CI 61–91), specificity 77% (95% CI 60–90), two-sided binomial test P = 0.005), while both the VIFS (n = 30, accuracy 63% (±8.8% s.e.m.), Cohen's d = 0.43, sensitivity 63% (95% CI 45–79), specificity 63% (95% CI 46–80), two-sided binomial test P = 0.200) and PINES (n = 30, accuracy 50% (±9.1% s.e.m.), Cohen's d = 0.10, sensitivity 50% (95% CI 31–67), specificity 50% (95% CI 30–69), two-sided binomial test P = 1.000) failed to predict the swallowing of disgust versus neutral liquids, further corroborating the specificity and sensitivity of VIDS for tracking disgust experience. Together, these results suggested that the visual disgust signature (that is, VIDS) generalizes to the gustatory modality and exhibits a higher specificity and sensitivity compared with other signatures during both tasting and swallowing of disgust liquids.

Validating the disgust signature in the socio-moral context

Finally, we determined whether the neural signature of subjective disgust experience captures socio-moral disgust in terms of neural reactivity to unfairness. In everyday life, humans frequently refer to disgust experience in social (that is, socio-moral) contexts. 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[8], with some evidence showing that disgust experience evoked by moral transgressions (for example, unfair offers) may be similar to the disgust experience elicited by core disgust stimuli (for example, uncleanliness and pathogens)[3,72], while other evidence suggested that unfair offers may evoke a general negative emotional response rather than disgust experience[53]. Here we thus test whether the VIDS—which captures subjective disgust experience elicited by core disgust stimuli—could track socio-moral 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 8, n = 43; Supplementary Methods and Supplementary Table 1). As shown in Fig. 8a,b, VIDS could predict high versus low unfairness with high accuracy (n = 43, accuracy 74% (±6.7% s.e.m.), Cohen's d = 0.73, sensitivity 74% (95% CI 61–88), specificity 74% (95% CI 61–87), two-sided binomial test P = 0.002) and could also predict high versus moderate unfairness (n = 43, accuracy 70% (±7.0% s.e.m.), Cohen's d = 0.37, sensitivity 70% (95% CI 56–82), specificity 70% (95% CI 56–83), two-sided binomial test P = 0.014). PINES could predict high versus low unfairness (n = 43, accuracy 77% (±6.4% s.e.m.), Cohen's d = 0.54, sensitivity 77% (95% CI 64–88), specificity 77% (95% CI 63–89), two-sided binomial test P < 0.001) yet failed to predict high versus moderate unfairness (n = 43, accuracy 56% (±7.6% s.e.m.), Cohen's d = 0.20, sensitivity 56% (95% CI 41–70), specificity 56% (95% CI 42–71), two-sided binomial test P = 0.542). VIFS did not predict either high versus low unfairness (n = 43, accuracy 58% (±7.5% s.e.m.), Cohen's d = 0.05, sensitivity 58% (95% CI 43–73), specificity 58% (95% CI 43–73), two-sided binomial test P = 0.360) or high versus moderate unfairness (n = 43, accuracy 53% (±7.6% s.e.m.), Cohen's d = 0.10, sensitivity 53% (95% CI 39–68), specificity 53% (95% CI 38–69), two-sided binomial test P = 0.761). These results indicated that the neurofunctional reaction to unfair offers is highly similar to the emotional reaction towards core disgust stimuli.

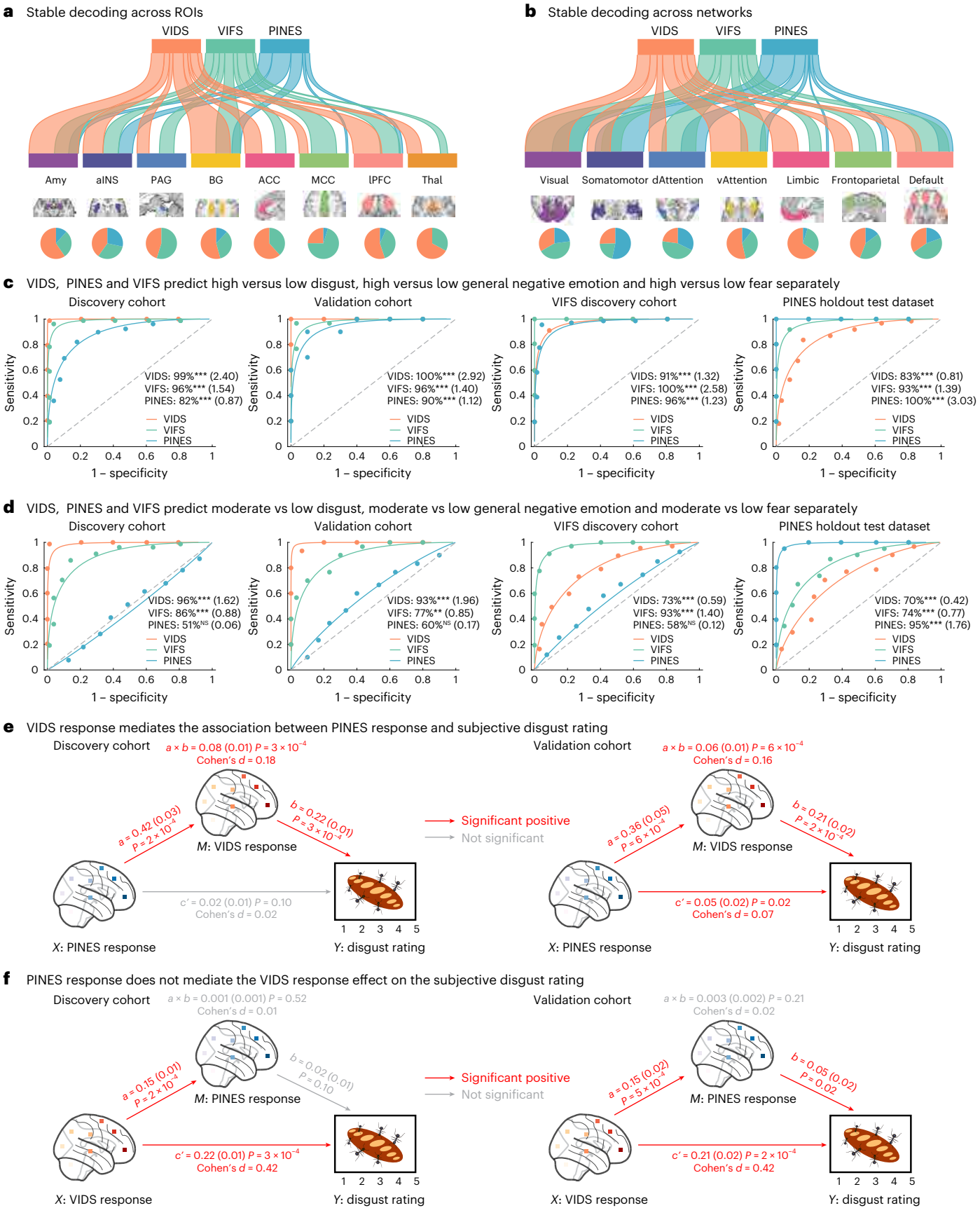
To further test whether VIDS simply captures negative emotional processing in high-arousing social contexts, we applied the VIDS to another independent fMRI dataset (study 9, n = 238; Supplementary Table 1) that acquired neural response during physical pain and painful facial stimuli as well as respective non-painful control stimuli. The results showed that the VIDS could not discriminate high-arousing painful faces from respective control faces (n = 238, accuracy 55% (±3.2% s.e.m.), Cohen's d = 0.18, sensitivity 55% (95% CI 49–62), specificity 55% (95% CI 48–61), two-sided binomial test P = 0.136; Fig. 8c). Although VIDS successfully distinguished physical pain versus corresponding control pictures (n = 238, accuracy 64% (±3.1% s.e.m.), Cohen's d = 0.34, sensitivity 64% (95% CI 58–70), specificity 64% (95% CI 58–70), two-sided binomial test P < 0.001; Fig. 8d), the effect size was substantially lower than that using VIDS to predict high versus low unfairness (Fig. 8a) (large effect sizes are commonly interpreted to indicate robust and replicable effects[73–75]); besides, the AUC of the VIDS for classifying high versus low unfairness (0.84) was larger than that for discriminating physical pain from corresponding control stimuli (0.68) (the AUC represents an important measure of classifier performance[46]). 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, Supplementary Results and Extended Data Fig. 10.

Discussion

Recent perspectives propose a paradigm shift towards subjective and conscious emotional experiences in neuroscience[18,44,50,76,77], 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 socio-moral contexts and underlie some psychopathological dysregulations (for example, 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

Fig. 6 | Comparing neurofunctional decoders for disgust, fear and negative affect (VIDS, VIFS and PINES). a, River plots showing spatial similarity (cosine similarity) between stable decoding maps and selected ROIs. **b**, River plots showing spatial similarity (cosine similarity) between stable decoding maps and selected networks. In **a** and **b**, ribbons are normalized by the max cosine similarity across all ROIs (**a**) and networks (**b**). 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. The pie charts show the relative contributions of each model to each ROI (**a**) or network (**b**) (that is, the percentage of voxels with the highest cosine similarity for each map). **c**,**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 (**c**) and moderate versus low (**d**) intensities, respectively. **P < 0.01, ***P < 0.001, NS not significant (binomial test, two-sided, uncorrected). **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. In **e** and **f**, the mediation analysis examines whether the observed covariance between the independent variable (X) and the dependent variable (Y) can be explained by the third variable (M, also mediator); for details, see Methods; two-sided P values are based on bootstrap tests with 10,000 samples, uncorrected. aINS, anterior insula; BG, basal ganglia.

https://doi.org/10.1038/s41562-024-01868-x



https://doi.org/10.1038/s41562-024-01868-x

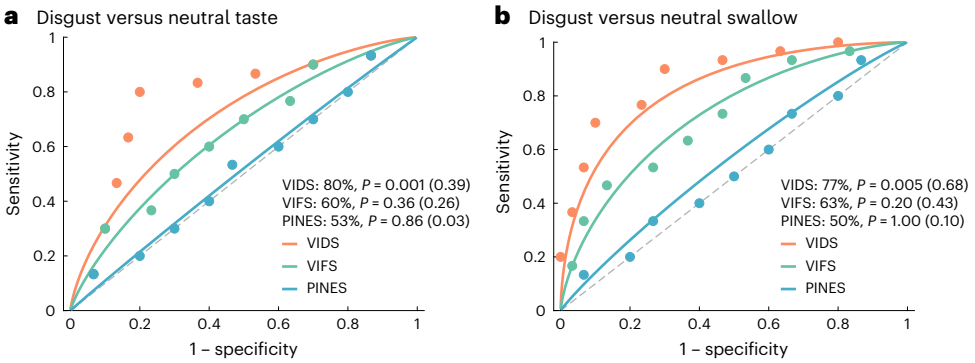


Fig. 7 | Validation tests of VIDS in the gustatory modality. a, During the taste stage, VIDS could predict disgust versus neutral liquids (shown as forced-choice classification accuracy, *P* value and Cohen's *d*), while VIFS and PINES failed to predict disgust versus neutral liquids. **b**, During the swallow stage, VIDS predicted disgust versus neutral liquids. However, neither PINES nor VIFS could predict disgust versus neutral liquids. *P* values were based on binomial tests, two-sided (uncorrected).

contexts with regard to accurately predicting disgust in response to core disgust stimuli, gustatory distaste and unfair offers, suggesting that the experience of disgust may extend from the gustatory over the subjective experience and into the socio-moral domain (see also evolutionary models proposed in ref. 14). 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 and VIFS) on subjective disgust ratings but not vice versa. Although the three neural signatures showed a certain extent of similarity in the range of intense emotional experiences (for a similar observation, see also ref. 44), the VIDS (as well as VIFS and PINES) were more sensitive to the target emotion in the moderate versus low intensity range. The VIDS moreover outperformed the other signatures in predicting disgust experience in response to both distaste and socio-moral 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 a comprehensive neurofunctional model of subjective disgust, and the resulting VIDS can predict subjective disgust experience with high robustness and generalization.

Previous studies emphasized the role of single brain systems or networks as core modules for disgust, including the insula[7,8,22,23] or the ventral attention network, which may facilitate detection and attentional processing of core disgust stimuli[78]. 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 and putamen)[8,18] and cortical systems engaged in interoceptive awareness and emotional appraisal such as the anterior insula, dorsal ACC and lateral frontal regions[29,30,44,63], as well as regions that have been identified as mediators of the bitter taste reaction in animal models (posterior insula[6] and brainstem[5]). However, neural activity in these isolated regions including systems such as insula or amygdala that have been typically implicated in disgust was only weakly associated with the subjective experience.

While no single network (for example, 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[63], 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[70] and to facilitate contextually sensitive escape decisions[79]. Consistent with previous predictive models for subjective emotional experiences[18], 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[80] and multivariate perspectives[18,39,42,43,52] that demonstrated that capturing affective experience requires integration across multiple distributed neural systems. The distributed representation perspective aligns with appraisal[81] and constructionist[51,80,82] theories of emotions that 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, for example, also Research Domain Criteria, RDoC[83]). For robust fMRI neuromarkers, a precise differentiation between subjective states is critical for model development and translation[44,46,48]. The VIDS exhibited partly overlapping yet also distinct neurofunctional representations compared with the PINES (non-specific negative affect)[39] or the VIFS (fear experience)[18]. 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.

According to Rozin et al., the distaste mechanism directly activates the disgust output system without a disgust evaluation stage[1], reflecting a reflexive hard-wired response to the chemosensory properties of gustatory stimuli. Mice do exhibit defensive responses (for example, retching-like behaviour)[5] and distinct facial expressions[6] after ingestion of potentially poisonous food, non-human primates may expel unpleasant tasting food[10] and human neonates exhibit a characteristic facial grimace towards bitter tastes[84]. The present study provides evidence

https://doi.org/10.1038/s41562-024-01868-x

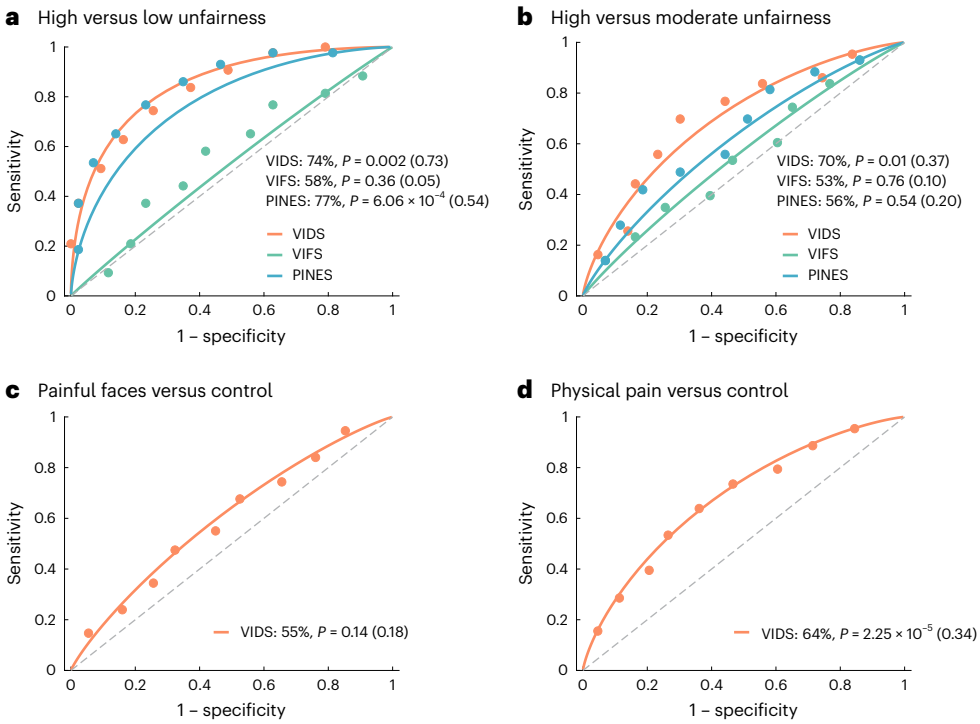


Fig. 8 | Validation tests of VIDS in social contexts. a, VIDS and PINES could predict high versus low unfairness (shown as forced-choice classification accuracy, *P* value and Cohen's *d*), while VIFS failed to predict high versus low unfairness. **b**, Neither PINES nor VIFS could predict high versus 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* values were based on binomial tests, two-sided (uncorrected).

for a link between the distaste response and the subjective core disgust experience by showing that the predictive potential of the signature for visually induced disgust generalizes to the neural representation of the gustatory distaste response (not only during the taste but also the swallow of disgust liquids) and provides supporting evidence for the proposed origin of disgust in oral distaste. Interestingly, meta-analytic functional decoding based on the Neurosynth repository not only revealed an association with affective states, including disgust (for example, disgust and feelings) but also an association with several gustation-related terms (for example, eating, food and taste; Extended Data Fig. 6).

We next utilized the disgust-predictive model to determine whether disgust experience extends into the socio-moral domain[3,53,72], in particular whether exposure to unfairness activates the VIDS. Evolutionary and disgust appraisal models have long hypothesized an association[1,14], 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 socio-moral disgust as it provides evidence for the activation of shared appraisal while it is concomitantly critical to decode other aversive states associated with withdrawal (for example, fear)[14]. Our results revealed that VIDS could predict different levels of unfairness with high accuracy (high, moderate and 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]. The authors showed that participants judged their experience as most similar to disgust when they received unfair offers, while other emotions such as 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 negative emotions, and anger and sadness may also play a role, albeit to a lesser 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 negative emotions (for example, disgust, $r = 0.94$; anger, $r = 0.94$; sadness, $r = 0.92$)[39], which explains why the PINES also captured unfairness to a certain extent as revealed in the present study. Unfair behaviour could thus induce a broad range of negative emotional responses in the observer. Therefore, we constructed a route from unfairness to socio-moral output via 'other negative affect evaluation' in Fig. 1a. Notably, a previous study employing MVPA also emphasized the role of general negative affect in response to unfairness[53] by demonstrating shared regional neural representations in the insular-mid ACC system for disgust and unfairness, a finding that is partly mirrored in the observation that the PINES could differentiate high versus low (yet not high versus moderate) unfairness (Fig. 8a). Other recent findings suggest that moral cognition interacts in a privileged manner with a neural representation of disgust and carefully controlled for aversiveness (that is, pain)[85]. The same study further underscores the role of functional interactions between regions as the common neural link between moral cognition and core disgust, such that enhanced coupling between separate subsystems (for example, PCC versus left ventral anterior insula) may link the neurofunctional pathway underlying disgust stimulation and moral appropriateness ratings of ethical dilemmas[85]. While the different paradigms used in the studies (directly experienced unfairness in the ultimatum game in the present study versus a cognitive evaluation of ethical dilemmas in ref. 85) may have contributed to the differences in determining shared neural activity patterns between disgust and socio-moral processes, future studies may include data acquired during different social cognitive and disgust processes and simultaneously account for shared and separable distributed activity and connectivity features (for a synergistic decoding approach, see also ref. 86). Altogether, our findings extend the initially proposed disgust appraisal model depicted in Fig. 1a, which combined several possibilities raised in previous disgust

appraisal models[1,14]. 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]. Here we demonstrate initial evidence for a route between the unfairness and the subjective disgust evaluation system. From the evolutionary perspective, these findings support the notion that disgust may have been pre-adapted to avoid harm and regulate behaviour in social contexts (see also adaptationist theory by Tybur et al.[11,12]).

From a biomarker perspective, it is imperative that a neuroaffective signature captures the respective mental process across variations of experimental contexts[40]. The VIDS generalized across cohorts, core and socio-moral disgust (as well as gustatory distaste) 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[49,53,87], we further tested a physical pain empathy pattern[49] on discovery and validation cohorts, yet the performance was substantially inferior to VIDS (Extended Data Fig. 8). 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 socio-moral domain; however, we could not provide a direct prediction of the degree of other negative emotions such as anger engaged throughout this process (for the role of anger in socio-moral disgust, see refs. 14,88). While the current study supports models that posit a strong link between core and socio-moral disgust, we focused on unfairness within a highly specific domain because aligning with fairness norms is deemed as a central part of human morality and sociality[3,89,90]. Nevertheless, moral disgust experience may extend to a range of social situations, and future studies may determine whether neurobiological models based on distributed activation patterns or coupling between regions[85] generalize across domains. Combining fMRI with MVPA enables us to detect and characterize fine-grained activity patterns with considerably higher precision compared with the conventional univariate approach, and can extend our understanding of how the brain engages in experimental tasks[56,91]. However, the fMRI–MVPA combination still requires cautious interpretation given that MVPA cannot compensate for the inherent limitations of fMRI, which indirectly probes neuronal activity, and distributed activation pattern cannot be viewed as reflecting the specific computational process of the neurons within a voxel[40,45,47]. Future studies should aim at providing more precise measurements of how fMRI aligns with detailed neuronal recordings at the microscopic scale (see also discussion in ref. 91).

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 (large-scale) brain systems rather than isolated brain regions. The robustness and specificity of the resulting VIDS were further tested with general negative emotion experience, subjective fear experience, gustatory modality (distaste) and social contexts (for example, pain empathy and socio-moral disgust). This study adds to the understanding of the neurobiological underpinnings of subjective disgust experience. Disgust dysregulations play a role in a number of mental disorders, and the resulting subjective disgust marker can be potentially useful to examine novel treatments (for initial applications of MVPA-based decoders in treatment and intervention evaluation, see refs. 21,92).

Methods

Ethics

The present paper includes 11 datasets (Supplementary Table 1). Among them, study 1 (discovery cohort), study 2 (validation cohort), study 3 (modified disgust induction task) and study 7 (gustatory experiment) were new and original experiments designed and implemented by the authors. Informed consent was obtained before each of these experiments. Corresponding experimental protocols were approved by the University of Electronic Science and Technology of China (UESTC) Ethics Board (study 1 and study 2 with the approval number 1061421093006220; study 3 and study 7 with the approval number 1061423092027488). The studies were implemented in line with the latest revision of the Declaration of Helsinki. Participants in study 1 and study 2 were compensated with 110 RMB, those in study 3 with 60 RMB, and those in study 7 with 90 RMB. The current work also includes secondary analysis of previously acquired experiments based on anonymized data (that is, the remaining seven studies; Supplementary Table 1). All participants in these studies provided informed consent in line with local ethics and institutional review boards. Detailed descriptions of the ethics approval and information on participant compensation are available through the corresponding references.

Participants in the discovery cohort (study 1)

Discovery cohort (study 1) included eighty participants recruited from the UESTC. All participants were right-handed, reported having normal or corrected to normal vision, had no history of mental or physical disorders, had 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 (>3 mm) during fMRI scanning, therefore leading to a final sample of $n = 78$ participants (44 females; mean ± s.d. age 22.10 ± 2.68 years).

Participants in the validation cohort (study 2)

To validate the performance of the disgust decoder (VIDS) developed using the discovery cohort (that is, 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 31 healthy participants recruited from the UESTC. Enrolment criteria were identical to study 1. Data from one male participant were excluded due to excessive head movement (>3 mm) during fMRI scanning, leading to a final sample of $n = 30$ participants (16 females; mean ± s.d. age 21.13 ± 2.18 years).

Stimuli and paradigm used in the discovery cohort (study 1)

The disgust rating task incorporated 80 pictorial stimuli from the DIRTI database[59] (67 pictures), Nencki Affective Picture System (NAPS)[93] (5 pictures) and International Affective Picture System (IAPS)[94] (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). Participants were instructed to pay attention to the pictures and naturally experience the induced emotion. Each trial consisted of a 6-s presentation of the picture followed by a 2-s fixation cross separating the stimuli from the rating period. Participants then had 4 s 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 7 s) before the presentation of the next picture (Fig. 1b). All participants reported '1–4' in their responses, while 6 out of 78 participants did not use rating '5'.

Stimuli and paradigm used in the validation cohort (study 2)

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). Structural images were collected using a high-resolution T1 spoiled gradient recall sequence (repetition time 6 ms, echo time 2 ms, flip angle 9°, field of view 256 × 256 mm, 1 × 1 × 1 mm voxels, 256 × 256 acquisition matrix, 156 slices, 1 mm 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 sequence (repetition time 2,000 ms, echo time 30 ms, flip angle 90°, field of view 240 × 240 mm, 3.75 × 3.75 × 4 mm voxels, 64 × 64 resolution, 39 slices, 3 mm slice thickness, 1 mm gap).

fMRI data pre-processing in discovery and validation cohorts

In line with our previous studies[18,49], all fMRI data were pre-processed and analysed using SPM12 (Statistical Parametric Mapping[95]). The first five volumes of the fMRI data in each run were removed to allow for T1 equilibration. Briefly, the pre-processing steps included correction for differences in slice timing, realignment of head motion, and unwarp to correct for magnetic field inhomogeneity, followed by the tissue segmentation of high-resolution anatomical image (into grey 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 space (interpolated to $2 \times 2 \times 2$ mm^3 voxels) and spatial smoothing with an 8-mm full width at half maximum Gaussian kernel. In line with earlier studies (for example, refs. 18,49,60,96,97), we used smoothed data (8-mm full width at half maximum Gaussian kernel) for MVPA as previous studies indicated that this smoothing level could improve inter-subject functional alignment without impairing the sensitivity to mesoscopic activity patterns that are consistent across participants[98,99]. Before first-level analysis, global motion outliers in each run were identified according to any of the following criteria: (1) signal intensity exceeding three s.d. from the global mean or (2) signal intensity and Mahalanobis distances exceeding ten mean absolute deviations[100]. Each timepoint 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 (GLMs). 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 (that is, 1–5), enabling us to model brain activity in response to each disgust level separately. The model also included one boxcar regressor indicating the 4-s 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 modelling the picture presentation period and the other modelling the subjective disgust rating period. The self-reported disgust ratings (that is, 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 haemodynamic response function and a high-pass filter of 128 s was applied. The four runs of the fMRI task were concatenated for each participant (via SPM spm_fmri_concatenate.m function). Nuisance variables encompassed: (1) 'dummy' regressors representing each run (intercept for each run); (2) 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); (3) vectors indicating motion outlier timepoints.

MVPA

We used whole-brain multivariate machine-learning pattern analysis[18,39,43,62] to determine a pattern of weighted brain activity that best predicted the self-reported disgust ratings. Here we used the SVR algorithm (linear kernel with $C = 1$) implemented in the Spider toolbox[101]. Whole-brain data masked with a grey matter mask from the discovery cohort was used. Features (that is, predictors) constituted 384 beta maps (one per rating for each participant, that is, $72 \times 5 + 6 \times 4$), 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 and to rule out the possibility of model overfitting[21,39,102], we used a rigorous 10×10-fold cross-validation procedure on the discovery cohort during which all participants were randomly assigned to ten subsamples of seven or eight participants using MATLAB's cvpartition function. The optimal hyperplane was computed on the basis of the multivariate pattern of 70 or 71 participants (training set) and evaluated by the excluded eight or seven 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[18,49,103].

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. 43). 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 (that is, 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 (and to rule out the possibility of overfitting[102]), we applied the VIDS to multiple separate test datasets, for example, the validation cohort, the dataset from the modified disgust induction task (study 3; Supplementary Methods), the generalization cohort (see 'Generalization cohort' section below), gustatory disgust dataset (see 'Validation in the gustatory context' section below) and ultimatum game dataset (see 'Validation in social contexts' 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 4 (ref. 54)). In line with the

aim of the current study, we focused on the disgust experience condition. Briefly, 26 participants (10 females; mean ± s.d. 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 ten blocks (five disgust and five neutral blocks), and each block encompassed three consecutive pictures (2 s 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[18,104,105], which was achieved by constructing a GLM design matrix with separate regressors for each stimulus. Each task regressor was convolved with the canonical haemodynamic 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 (that is, beta estimates) for a given trial could be strongly contaminated by artefacts (for example, participant's head movement, scanner pulse artefacts and so on) 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 (that is, 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 s.d. of the bootstrap distribution, on the population-level disgust-predictive pattern (that is, the VIDS). After correction for multiple comparisons (FDR $q < 0.05$), those regions with significant and consistent model weights were obtained[62]. Secondly, as in previous studies that computed model encoding maps[18,43], we transformed the population-level pattern (that is, VIDS, also backward model) into 'activation pattern' (forward model) that maps each voxel to the response (fitted values) in the multivariate model[55]. The reconstructed 'activation pattern' is known as 'structure coefficient' in the statistical literature[106,107]. 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 (that is, 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 adopted two approaches. First, 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[64,65]. 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[108] were reported through probabilistic maps indicating the behavioural domain based on meta data labels of the BrainMap database[64,109]. Second, to complement the BrainMap functional decoding, we further performed an exploratory decoding analysis of the disgust pattern using data derived from Neurosynth and NiMARE[110,111], which contain a large pool of automatically generated meta-analytic activation maps across a multitude of terms/topics. This approach enables us to discuss our results in relation to these terms/topics, without relying on acquiring data from a wide range of functional neuroimaging tasks in the same cohort. Consistent with the recommended input specifications as well as recent publications[62,112–114], we subjected the unthresholded VIDS pattern for the meta-analytic decoding. Here we capitalized on the BrainStat toolbox[115,116] to run the analysis. For every meta-analytic map in the database, a voxel-wise Pearson product-moment correlation between the meta-analytic map and input map was calculated. We only displayed the top 100 most correlated functional terms/topics for clarity. Additionally, we excluded topics such as dorsolateral prefrontal, anterior cingulate and so on.

Furthermore, we tested whether disgust processing could be reducible to activations in a single brain region (for example, insula and amygdala) or network (for example, the DMN). To examine this hypothesis, we performed whole-brain searchlight- (three-voxel radius spheres)—and parcellation- (279 regions[64,65])—based analyses to identify local regions predictive of disgust and compared model performances of local regions with the whole-brain model (that is, the VIDS). Additionally, we compared the prediction performances of insula and amygdala (based on the 279-region version of the Brainnetome Atlas[64,65]) as well as large-scale networks to the whole-brain model. The networks of interest comprised seven large-scale cerebral networks[117], a subcortical network (including the striatum, thalamus, hippocampus and amygdala) and a cortical 'consciousness network'[63]. 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 (PINES[39]) and subjective fear experience (VIFS[18]). To compare the performance of VIDS with the PINES and the VIFS, we applied the three decoders to the discovery, validation, PINES holdout test[39] (study 5, $n = 61$; see Supplementary Table 1 for details) and VIFS discovery[18] (study 6, $n = 67$; Supplementary Table 1) 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 ROIs as well as networks of interest

River plots were created to illustrate spatial similarity between stable decoding maps derived from bootstrap tests and a priori ROIs previously documented as regions linking to disgust, fear and negative

affect processes[8,18,39,43]. We further depicted spatial similarity between stable decoding maps and seven large-scale cerebral networks, the subcortical and consciousness networks. In line with a recent study[43], 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.

Multi-level two-path mediation analysis

To explore the relationship between VIDS response, disgust rating and PINES response, multi-level two-path mediation analyses were performed using the Mediation Toolbox, available via ref. 118 (see also refs. 18,119). 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, that is, $a \times b$). A significant mediation effect is obtained when a, b and $a \times b$ are all significant. Furthermore, when c' is significant, M (that is, the mediator) is considered to have a partial mediation effect; otherwise, M plays a full mediation role. In this study, we constructed two multi-level 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 the 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 the gustatory context

To test whether the VIDS—developed on visual stimuli—can generalize to the gustatory modality, we designed and implemented a new fMRI experiment (study 7, $n = 30$; Supplementary Methods and Supplementary Table 1) employing gustatory stimuli (design based on previous similar studies[3,53,120–122]). Next, we applied the VIDS—as well as the VIFS and PINES—to the gustatory fMRI data. Specifically, we calculated forced-choice classification accuracies between disgust and neutral taste (as well as between swallowing of disgust and neutral liquids) based on the pattern responses.

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[7,8,14])—could be extended into the socio-moral disgust domain (for example, unfairness[3,72]), 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 8, $n = 43$; Supplementary Methods and Supplementary Table 1). 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. It is noteworthy that disgust is not the only emotion evoked in response to unfairness; other negative emotions might also be involved[3]. Thus, we compared the performance of PINES and VIFS with VIDS in the socio-moral context. Next, to further examine how specific the VIDS is for unfairness as compared with other social contexts (for example, pain empathy), we tested the performance of VIDS on another independent dataset (study 9, $n = 238$, fMRI brain responses to physical pain and respective non-painful control images as well as to painful faces and respective non-painful control images[49]; Supplementary Table 1). We calculated forced-choice classification accuracies between physical pain and control (as well as between painful faces and control) based on the pattern responses.

Reporting summary

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

Data availability

fMRI data used to train and validate the signature are available via figshare at https://figshare.com/articles/dataset/Discovery_dataset_disgust/22827974 (ref. 123) (study 1) and https://figshare.com/articles/dataset/validation_dataset_disgust/22841117 (ref. 124) (study 2). fMRI data of the modified disgust induction task are available via figshare at https://figshare.com/articles/dataset/Dataset_of_the_modified_disgust_induction_experiment/25284895 (ref. 125) (study 3). The data of study 4 were provided by the authors of a previous study[54]. The data of study 5 are from a previous study[39] and are available via NeuroVault at https://neurovault.org/collections/1964 (ref. 126). The data of study 6 are from a previous study[18] and are available via figshare at https://figshare.com/articles/dataset/Subjective_fear_dataset/13271102?file=25556276 (ref. 127). The data from the gustatory experiment (study 7) as well as the ultimatum game experiment (study 8) are from two independent ongoing projects of our research team and are available from the corresponding author upon request. The data of pain empathy task (study 9) from our previous study[49] are available via figshare at https://figshare.com/articles/dataset/Vicarious_pain_dataset/11994498 (ref. 128). Data from the negative/neutral experiment (study 10) are from an ongoing project from our team[57] and available upon request. The data of study 11 were provided by the authors of a previous study[58] (note that we applied for a subset of randomly selected data from 150 participants (with gender ratio balanced) from the authors). The VIDS and the thresholded statistical maps are available via figshare at https://figshare.com/articles/dataset/Brain_models_and_maps_zip/22827950 (ref. 129). Data from the functional characterizations of the Brainnetome Atlas based on the BrainMap database are available at https://atlas.brainnetome.org/bnatlas.html (ref. 108).

Code availability

Data were analysed using CANLab neuroimaging analysis tools available via GitHub at https://github.com/canlab (ref. 130) and https://github.com/ganxianyang/fMRI-studies/tree/main/Subjective_disgust_experience_signature (ref. 131). Glass brain was drawn using the open-source Python package Nilearn[132].

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. Rozin, P. & Haidt, J. The domains of disgust and their origins: contrasting biological and cultural evolutionary accounts. *Trends Cogn. Sci.* **17**, 367–368 (2013).

5. Xie, Z. et al. The gut-to-brain axis for toxin-induced defensive responses. *Cell* **185**, 4298–4316.e4221 (2022).
6. 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).
7. 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).
8. 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).
9. Amoroso, C. R. et al. Disgust theory through the lens of psychiatric medicine. *Clin. Psychol. Sci.* **8**, 3–24 (2019).
10. Jones, D. The depths of disgust. *Nature* **447**, 768–771 (2007).
11. Tybur, J. M., Lieberman, D., Kurzban, R. & DeScioli, P. Disgust: evolved function and structure. *Psychol. Rev.* **120**, 65–84 (2013).
12. Tybur, J. M., Lieberman, D. & Griskevicius, V. Microbes, mating, and morality: individual differences in three functional domains of disgust. *J. Pers. Soc. Psychol.* **97**, 103–122 (2009).
13. 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).
14. 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).
15. Zhang, C., Vincelette, L. K., Reimann, F. & Liberles, S. D. A brainstem circuit for nausea suppression. *Cell Rep.* **39**, 110953 (2022).
16. Zhang, C. et al. Area postrema cell types that mediate nausea-associated behaviors. *Neuron* **109**, 461–472.e465 (2021).
17. Becker, B. et al. Fear processing and social networking in the absence of a functional amygdala. *Biol. Psychiatry* **72**, 70–77 (2012).
18. Zhou, F. et al. A distributed fMRI-based signature for the subjective experience of fear. *Nat. Commun.* **12**, 6643 (2021).
19. 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).
20. 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).
21. Reddan, M. C., Wager, T. D. & Schiller, D. Attenuating neural threat expression with imagination. *Neuron* **100**, 994–1005.e1004 (2018).
22. Chapman, H. A. & Anderson, A. K. Understanding disgust. *Ann. N. Y. Acad. Sci.* **1251**, 62–76 (2012).
23. Wicker, B. et al. Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* **40**, 655–664 (2003).
24. Stark, R. et al. Erotic and disgust-inducing pictures—differences in the hemodynamic responses of the brain. *Biol. Psychol.* **70**, 19–29 (2005).
25. 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).
26. Baumann, O. & Mattingley, J. B. Functional topography of primary emotion processing in the human cerebellum. *NeuroImage* **61**, 805–811 (2012).
27. 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).
28. Viol, K. et al. Erroneously disgusted: fMRI study supports disgust-related neural reuse in obsessive-compulsive disorder (OCD). *Front. Behav. Neurosci.* **13**, 81 (2019).
29. 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).
30. Yao, S. et al. Oxytocin modulates attention switching between interoceptive signals and external social cues. *Neuropsychopharmacology* **43**, 294–301 (2018).
31. 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).
32. Levine, S. M., Wackerle, A., Rupprecht, R. & Schwarzbach, J. V. The neural representation of an individualized relational affective space. *Neuropsychologia* **120**, 35–42 (2018).
33. Levine, S. M. & Schwarzbach, J. V. Individualizing representational similarity analysis. *Front. Psychiatry* **12**, 729457 (2021).
34. 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).
35. Shapira, N. A. et al. Brain activation by disgust-inducing pictures in obsessive-compulsive disorder. *Biol. Psychiatry* **54**, 751–756 (2003).
36. 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).
37. 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).
38. Stark, R. et al. Hemodynamic brain correlates of disgust and fear ratings. *Neuroimage* **37**, 663–673 (2007).
39. 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).
40. 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).
41. Haxby, J. V. Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage* **62**, 852–855 (2012).
42. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**, 1437–1448 (2015).
43. 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).
44. Wager, T. D., Krishnan, A. & Hitchcock, E. in *The Nature of Emotion. Fundamental Questions* (eds Fox, A. S. et al.) 112–118 (Oxford Univ. Press, 2018).
45. Weaverdyck, M., Lieberman, M. & Parkinson, C. Multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. *Soc. Cogn. Affect. Neurosci.* **15**, 487–509 (2020).
46. Kragel, P. A. & LaBar, K. S. Decoding the nature of emotion in the brain. *Trends Cogn. Sci.* **20**, 444–455 (2016).
47. Haynes, J.-D. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* **87**, 257–270 (2015).
48. 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).
49. Zhou, F. et al. Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. *eLife* **9**, e56929 (2020).
50. LeDoux, J. E. & Brown, R. A higher-order theory of emotional consciousness. *Proc. Natl Acad. Sci. USA* **114**, E2016–E2025 (2017).
51. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **12**, 1–23 (2017).

52. Saarimäki, H. et al. Discrete neural signatures of basic emotions. *Cereb. Cortex* **26**, 2563–2573 (2015).
53. 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).
54. Chen, S. et al. Functional decoupling of emotion coping network subsides automatic emotion regulation by implementation intention. *Neural Plast.* **2021**, 6639739 (2021).
55. Haufe, S. et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **87**, 96–110 (2014).
56. Peelen, M. V. & Downing, P. E. Testing cognitive theories with multivariate pattern analysis of neuroimaging data. *Nat. Hum. Behav.* **7**, 1430–1441 (2023).
57. 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).
58. Fastenrath, M. et al. Human cerebellum and corticocerebellar connections involved in emotional memory enhancement. *Proc. Natl Acad. Sci. USA* **119**, e2204900119 (2022).
59. 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).
60. Wager, T. D. et al. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
61. 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).
62. Kohoutová, L. et al. Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat. Protoc.* **15**, 1399–1435 (2020).
63. 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).
64. Fan, L. et al. The Human Brainnetome Atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
65. Lee, J.-J. et al. A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* **27**, 174–182 (2021).
66. Damasio, A. & Carvalho, G. B. The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* **14**, 143–152 (2013).
67. Kirby, L. A. J. & Robinson, J. L. Affective mapping: an activation likelihood estimation (ALE) meta-analysis. *Brain Cogn.* **118**, 137–148 (2017).
68. 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).
69. 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).
70. Satpute, A. B. & Lindquist, K. A. The default mode network's role in discrete emotion. *Trends Cogn. Sci.* **23**, 851–864 (2019).
71. 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).
72. 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).
73. Han, X. et al. Effect sizes and test–retest reliability of the fMRI-based neurologic pain signature. *NeuroImage* **247**, 118844 (2022).
74. Poldrack, R. A. et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* **18**, 115–126 (2017).
75. Reddan, M. C., Lindquist, M. A. & Wager, T. D. Effect size estimation in neuroimaging. *JAMA Psychiatry* **74**, 207–208 (2017).
76. 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).
77. Kyzar, E. J. & Denfield, G. H. Taking subjectivity seriously: towards a unification of phenomenology, psychiatry, and neuroscience. *Mol. Psychiatry* **28**, 10–16 (2023).
78. Pujol, J. et al. Mapping the sequence of brain events in response to disgusting food. *Hum. Brain Mapp.* **39**, 369–380 (2018).
79. Qi, S. et al. How cognitive and reactive fear circuits optimize escape decisions in humans. *Proc. Natl Acad. Sci. USA* **115**, 3186–3191 (2018).
80. 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).
81. Clore, G. L. & Ortony, A. in *Handbook of Emotions* 3rd ed. (eds Lewis, M., Haviland-Jones, J. M. & Barrett, L. F.) 628–642 (The Guilford Press, 2008).
82. Wager, T. D. et al. A Bayesian model of category-specific emotional brain responses. *PLoS Comput. Biol.* **11**, e1004066 (2015).
83. 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).
84. Steiner, J. E. The gustofacial response: observation on normal and anencephalic newborn infants. in *Fourth Symposium on Oral Sensation and Perception* (ed. Bosma, J. F.) 254–278 (US Department of Health, Education and Welfare, 1973).
85. Sharvit, G., Lin, E., Vuilleumier, P. & Corradi-Dell'Acqua, C. Does inappropriate behavior hurt or stink? The interplay between neural representations of somatic experiences and moral decisions. *Sci. Adv.* **6**, eaat4390 (2020).
86. Zhou, F. et al. Capturing dynamic fear experiences in naturalistic contexts: an ecologically valid fMRI signature integrating brain activation and connectivity. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.18.553808 (2023).
87. 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).
88. 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).
89. Henrich, J. et al. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford Univ. Press, 2004).
90. Sheskin, M. in *Interdisciplinary Perspectives on Fairness, Equity, and Justice* (eds Li, M. & Tracer, D. P.) 33–49 (Springer, 2017).
91. Finn, E. S., Poldrack, R. A. & Shine, J. M. Functional neuroimaging as a catalyst for integrated neuroscience. *Nature* **623**, 263–273 (2023).
92. Zhang, R. et al. Angiotensin II regulates the neural expression of subjective fear in humans: a precision pharmaco-neuroimaging approach. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **8**, 262–270 (2023).
93. 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. *Behav. Res. Methods* **46**, 596–610 (2014).

94. 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, 2008).
95. SPM12. *SPM* https://www.fil.ion.ucl.ac.uk/spm/software/spm12/ (2020).
96. Woo, C.-W. et al. Separate neural representations for physical pain and social rejection. *Nat. Commun.* **5**, 5380 (2014).
97. Losin, E. A. R. et al. Neural and sociocultural mediators of ethnic differences in pain. *Nat. Hum. Behav.* **4**, 517–530 (2020).
98. Op de Beeck, H. P. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage* **49**, 1943–1948 (2010).
99. Shmuel, A., Chaimow, D., Raddatz, G., Ugurbil, K. & Yacoub, E. Mechanisms underlying decoding at 7 T: ocular dominance columns, broad structures, and macroscopic blood vessels in V1 convey information on the stimulated eye. *NeuroImage* **49**, 1957–1964 (2010).
100. CanlabCore. *GitHub* https://github.com/canlab/CanlabCore (2014).
101. MPI Biol. Cybernetics. The Spider. http://people.kyb.tuebingen.mpg.de/spider (2006).
102. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* **77**, 534–540 (2020).
103. 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).
104. 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).
105. Büchel, C. et al. 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).
106. Courville, T. & Thompson, B. Use of structure coefficients in published multiple regression articles: β is not enough. *Educ. Psychol. Meas.* **61**, 229–248 (2001).
107. Thompson, B. & Borrello, G. M. The importance of structure coefficients in regression research. *Educ. Psychol. Meas.* **45**, 203–209 (1985).
108. Brainnetome Atlas. https://atlas.brainnetome.org/bnatlas.html (2023).
109. BrainMap Taxonomy. *brainmap.org* http://www.brainmap.org/taxonomy/ (2022).
110. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
111. Salo, T. et al. NiMARE: neuroimaging meta-analysis research environment. *Aperture Neuro* **3**, 1–32 (2023).
112. Schurz, M. et al. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* **147**, 293–327 (2021).
113. Maliske, L. Z., Schurz, M. & Kanske, P. Interactions within the social brain: co-activation and connectivity among networks enabling empathy and theory of mind. *Neurosci. Biobehav. Rev.* **147**, 105080 (2023).
114. Pankey, B. S. et al. Extended functional connectivity of convergent structural alterations among individuals with PTSD: a neuroimaging meta-analysis. *Behav. Brain Funct.* **18**, 9 (2022).
115. BrainStat. *GitHub* https://github.com/MICA-MNI/BrainStat (2022).
116. Larivière, S. et al. BrainStat: a toolbox for brain-wide statistics and multimodal feature associations. *NeuroImage* **266**, 119807 (2023).
117. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
118. MediationToolbox. *GitHub* https://github.com/canlab/MediationToolbox (2014).
119. Woo, C.-W. et al. Quantifying cerebral contributions to pain beyond nociception. *Nat. Commun.* **8**, 14211 (2017).
120. Eskine, K. J., Kacinik, N. A. & Prinz, J. J. A bad taste in the mouth: gustatory disgust influences moral judgment. *Psychol. Sci.* **22**, 295–299 (2011).
121. Jabbi, M., Swart, M. & Keysers, C. Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage* **34**, 1744–1753 (2007).
122. Jabbi, M., Bastiaansen, J. & Keysers, C. A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. *PLoS ONE* **3**, e2939 (2008).
123. Disgust discovery cohort dataset. *figshare* https://figshare.com/articles/dataset/Discovery_dataset_disgust/22827974 (2023).
124. Disgust validation cohort dataset. *figshare* https://figshare.com/articles/dataset/validation_dataset_disgust/22841117 (2023).
125. Dataset of the modified disgust induction experiment. *figshare* https://figshare.com/articles/dataset/Dataset_of_the_modified_disgust_induction_experiment/25284895 (2024).
126. PINES holdout test dataset. *NeuroVault* https://neurovault.org/collections/1964 (2015).
127. VIFS discovery cohort dataset. *figshare* https://figshare.com/articles/dataset/Subjective_fear_dataset/13271102?file=25556276 (2021).
128. Pain empathy task dataset. *figshare* https://figshare.com/articles/dataset/Vicarious_pain_dataset/11994498 (2020).
129. The VIDS and thresholded stastistical maps. *figshare* https://figshare.com/articles/dataset/Brain_models_and_maps_zip/22827950 (2023).
130. Cognitive and Affective Neuroscience Laboratory, CANLab. *GitHub* https://github.com/canlab (2014).
131. Subjective_disgust_experience_signature. *GitHub* https://github.com/ganxianyang/fMRI-studies/tree/main/Subjective_disgust_experience_signature (2023).
132. Abraham, A. et al. Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14 (2014).

Acknowledgements

We thank D. Coynel and D. J.-F. de Quervain as well as S. Chen for sharing their data. We also thank the CANLab for providing the PINES signature and the PINES holdout dataset. In addition, we thank X. Tian and Q. Xie (both majored in pharmacy) who provided us with the essential background knowledge on how to calculate the concentration of different gustatory liquids and the appropriate medical-grade equipment for the gustatory experiment. Any opinions, findings, conclusions or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission. This work was partly supported by the China MOST2030 Brain Project (grant no. 2022ZD0208500) to D.Y., the National Natural Science Foundation of China (grant nos. 32250610208 and 82271583 to B.B., and 32300862 to F.Z.), National Key Research and Development Program of China (grant no. 2018YFA0701400) to B.B., the Fundamental Research Funds for the Central Universities (SWU2309733) to F.Z. and a start-up grant from the University of Hong Kong to B.B. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

X.G., F.Z. and B.B. conceived and designed the experiment. X.G., F.Z. and B.B. analysed the data and were responsible for interpretation of data. T.X., X.L., R.Z., Z.Z., X.Y., X.Z., F.Y., J.L. and R.C. provided important suggestions during formal analysis. X.G., R.Z., T.X. and L.W. conducted

the experiment. X.G. and R.C. were responsible for visualization. X.G. and B.B. drafted the paper; F.Z., J.Y. and D.Y. provided feedback and revised the paper. B.B. supervised the project and acquired the funding. All authors meet the four ICMJE authorship criteria and were responsible for revising the paper, for approving the final version for publication and for accuracy and integrity of the work.

Competing interests
The authors declare no competing interevsts.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41562-024-01868-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-024-01868-x.

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

Peer review information *Nature Human Behaviour* thanks Corrado Corradi-Dell'Acqua, Peter de Jong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

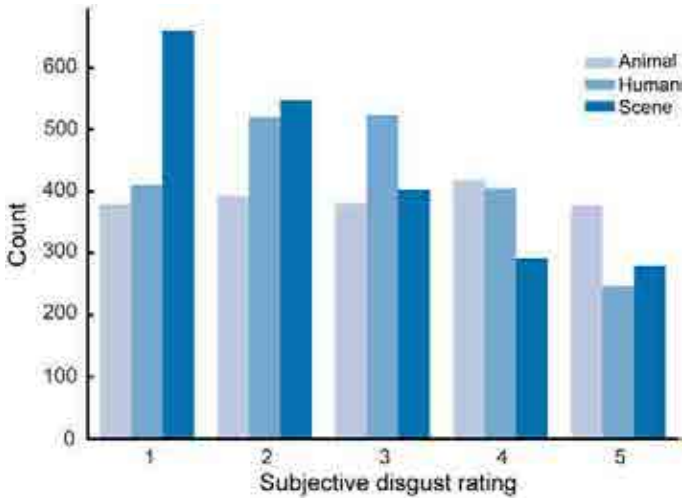
Reprints and permissions information is available at www.nature.com/reprints.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

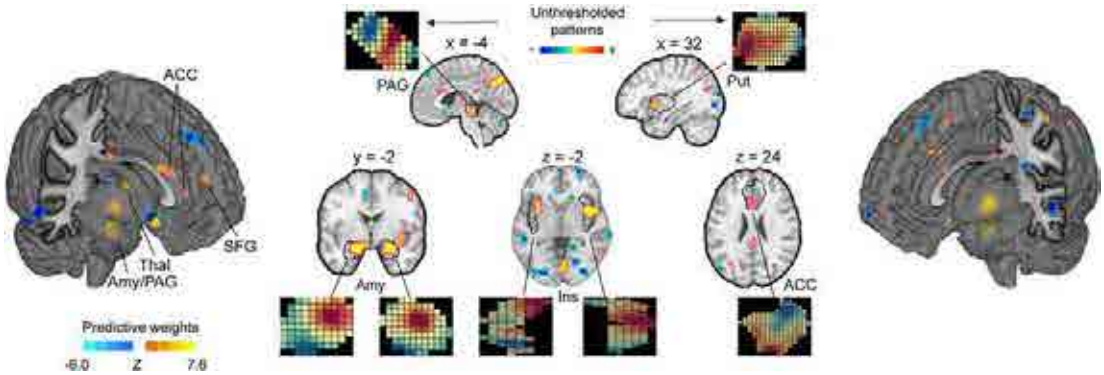
© The Author(s), under exclusive licence to Springer Nature Limited 2024

https://doi.org/10.1038/s41562-024-01868-x



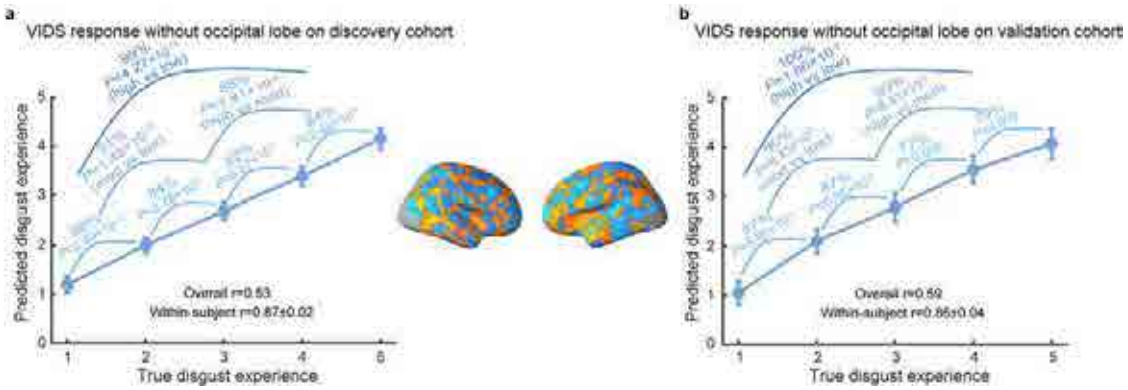
Extended Data Fig. 1 | The distribution of subjective disgust ratings for each category (that is, animal, human and scene), respectively.

https://doi.org/10.1038/s41562-024-01868-x



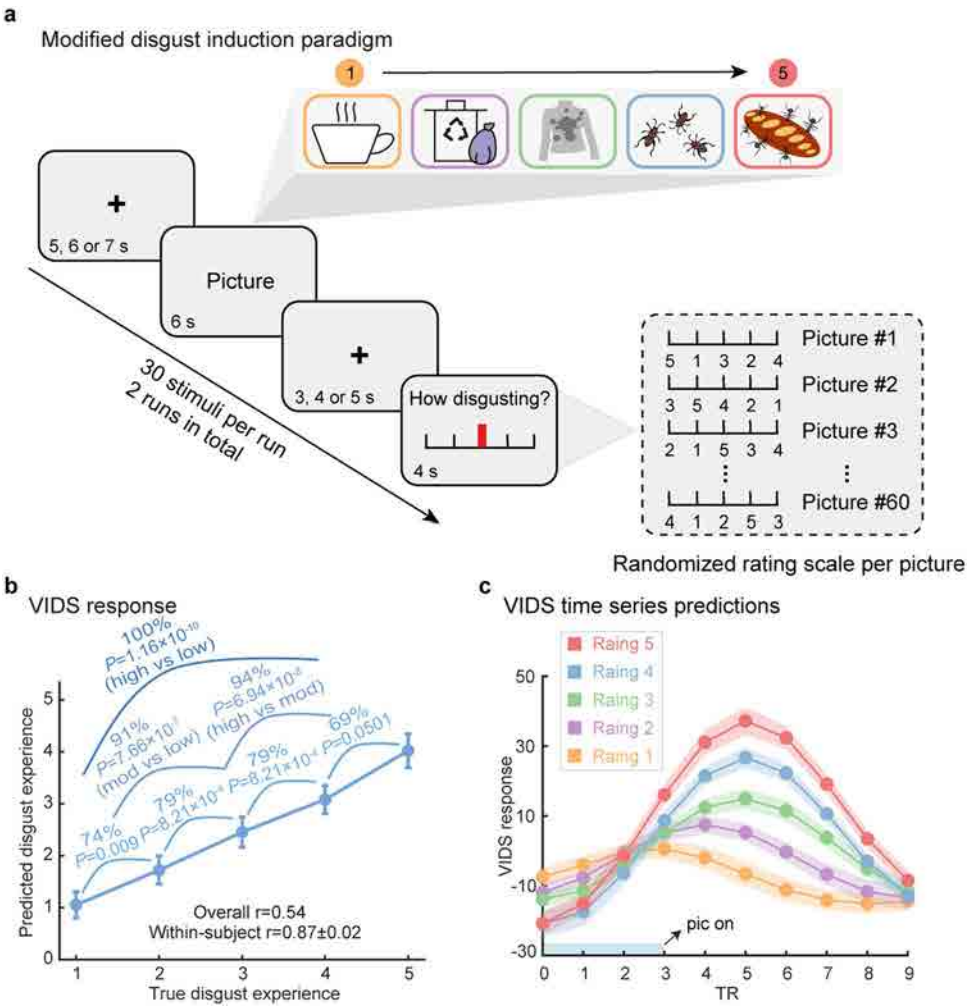
Extended Data Fig. 2 | The spatial topography of the unthresholded patterns in some anatomical regions of interest (ROIs). This panel illustrates the VIDS pattern thresholded using a 10,000-sample bootstrap procedure at $q < 0.05$, FDR corrected. Inserts show the spatial topography of the unthresholded patterns in some anatomical ROIs. ACC=anterior cingulate cortex, Amy=amygdala, Ins=insula, PAG=periaqueductal gray, Put=putamen, SFG=superior frontal gyrus, Thal=thalamus.

https://doi.org/10.1038/s41562-024-01868-x



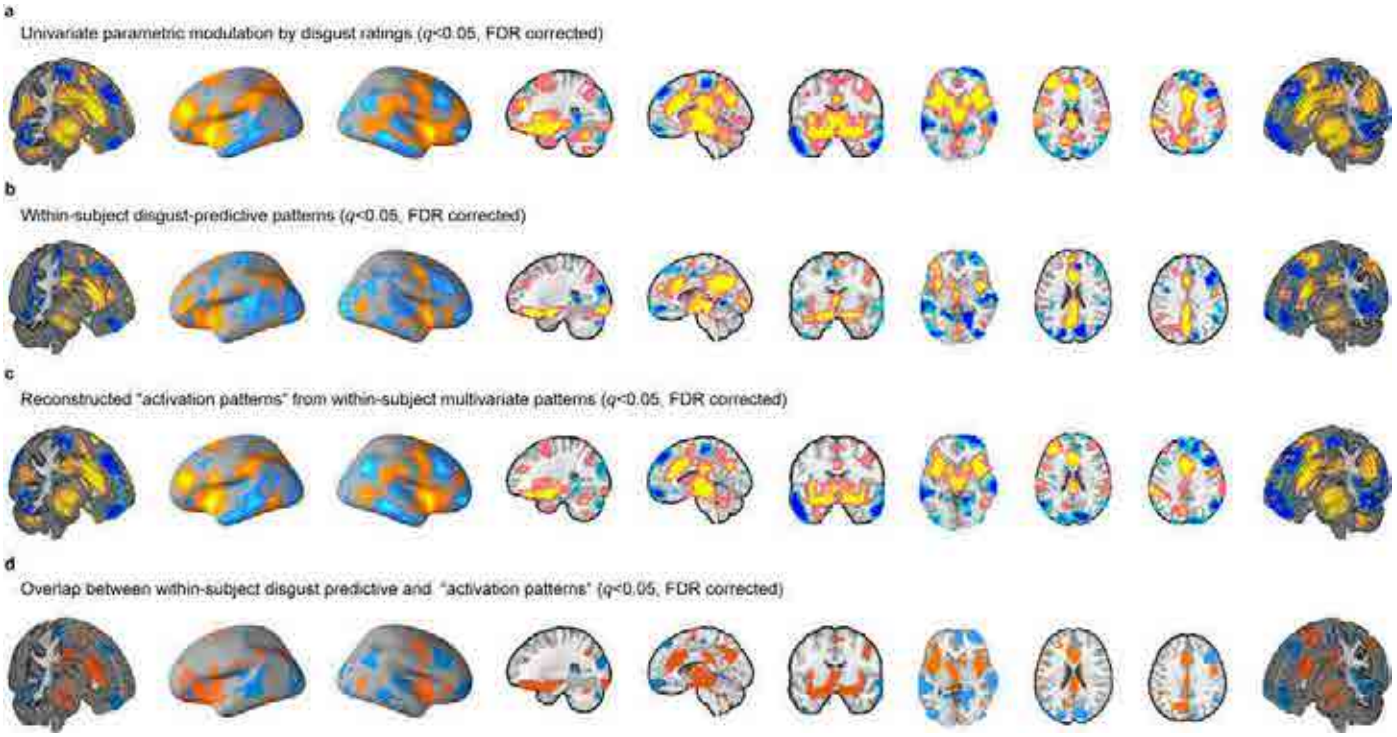
Extended Data Fig. 3 | VIDS pattern response without occipital lobe. a, The predicted disgust ratings compared to the true ratings for the cross-validated discovery cohort ($n = 78$). **b**, The predicted disgust ratings compared to the true ratings for the independent validation cohort ($n = 30$). Accuracies reflect forced-choice comparisons. *P* values were based on binomial tests, two-sided (uncorrected). *r* indicates the Pearson correlation coefficient between predicted and true ratings. Error bars reflect the s.e.m.

https://doi.org/10.1038/s41562-024-01868-x



Extended Data Fig. 4 | The VIDS tracks disgust experience independent of the motor responses. a, The modified disgust induction paradigm included a jittered period between the stimulus and the rating, and the rating numbers were provided in a randomized order. This allowed to better uncouple the motor and emotional response. Of note, schematic figures were used for display purpose only to avoid copyright issues and were not included in the original stimulus set. **b**, Predicted disgust experience (subjective ratings; mean ± s.e.m.) compared to actual disgust ratings using data from the modified disgust induction task (acquired in $n = 34$ individuals). Accuracy provided for forced-choice comparisons. *P* values were based on two-sided independent binomial tests (uncorrected). *r* indicates Pearson correlation coefficient between predicted and true ratings. **c**, Averaged peristimulus plot (mean ± s.e.m.) of the VIDS response using data from the modified disgust induction task at every repetition time (TR; 2 s) for each disgust intensity rating separately. pic, picture. Error bars and shaded regions indicate the s.e.m.

https://doi.org/10.1038/s41562-024-01868-x

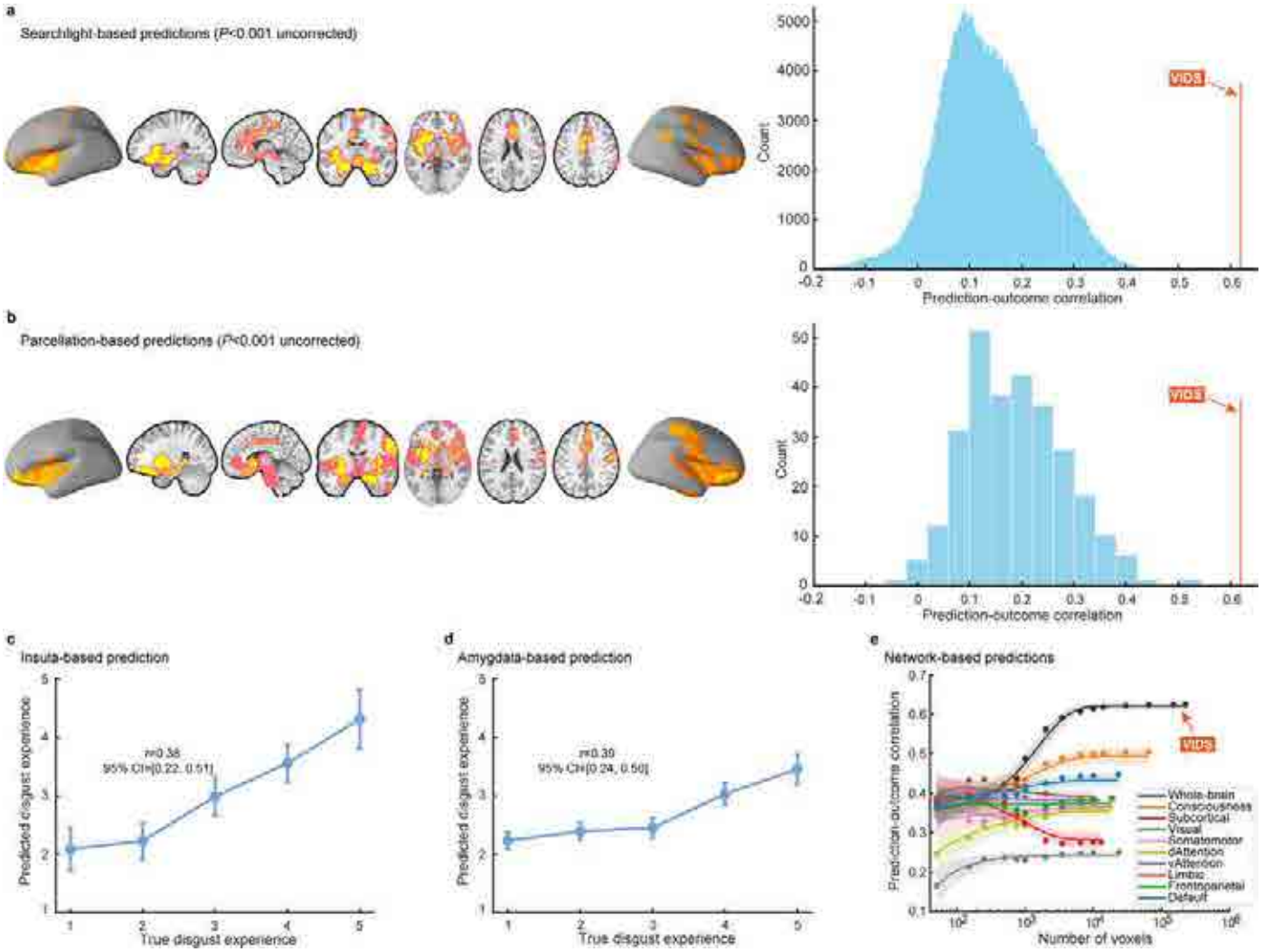


Extended Data Fig. 5 | Subjective experience of disgust is associated with and predicted by distributed brain regions. a, The univariate parametric effects of disgust ratings. **b**, Multivariate patterns trained on individual subjects and depicts brain regions consistently predictive of subjective disgust across participants. **c**, Thresholded transformed 'activation patterns' from within-subject disgust-predictive patterns. **d**, Overlapping (that is, from a conjunction analysis) brain regions between (**b** and **c**). Hot color indicates positive associations (**a** and **c**) or weights (**b**) whereas cold color indicates negative associations (**a** and **c**) or weights (**b**).

https://doi.org/10.1038/s41562-024-01868-x

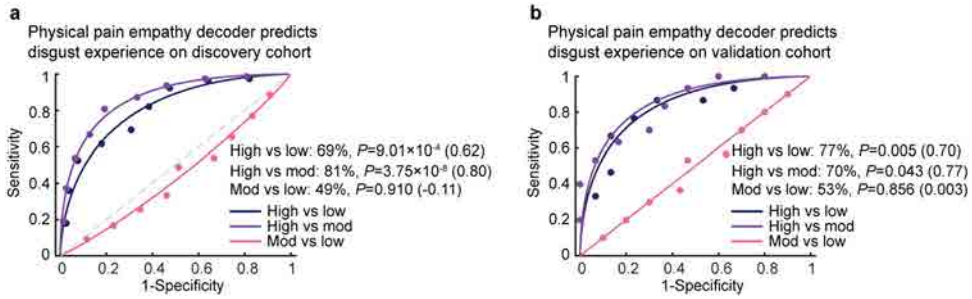
Extended Data Fig. 6 | Neurosynth functional decoding of the unthresholded VIDS. Here, the 100 most strongly correlated terms were displayed, with a larger font size indicating a larger Pearson correlation coefficient.

https://doi.org/10.1038/s41562-024-01868-x



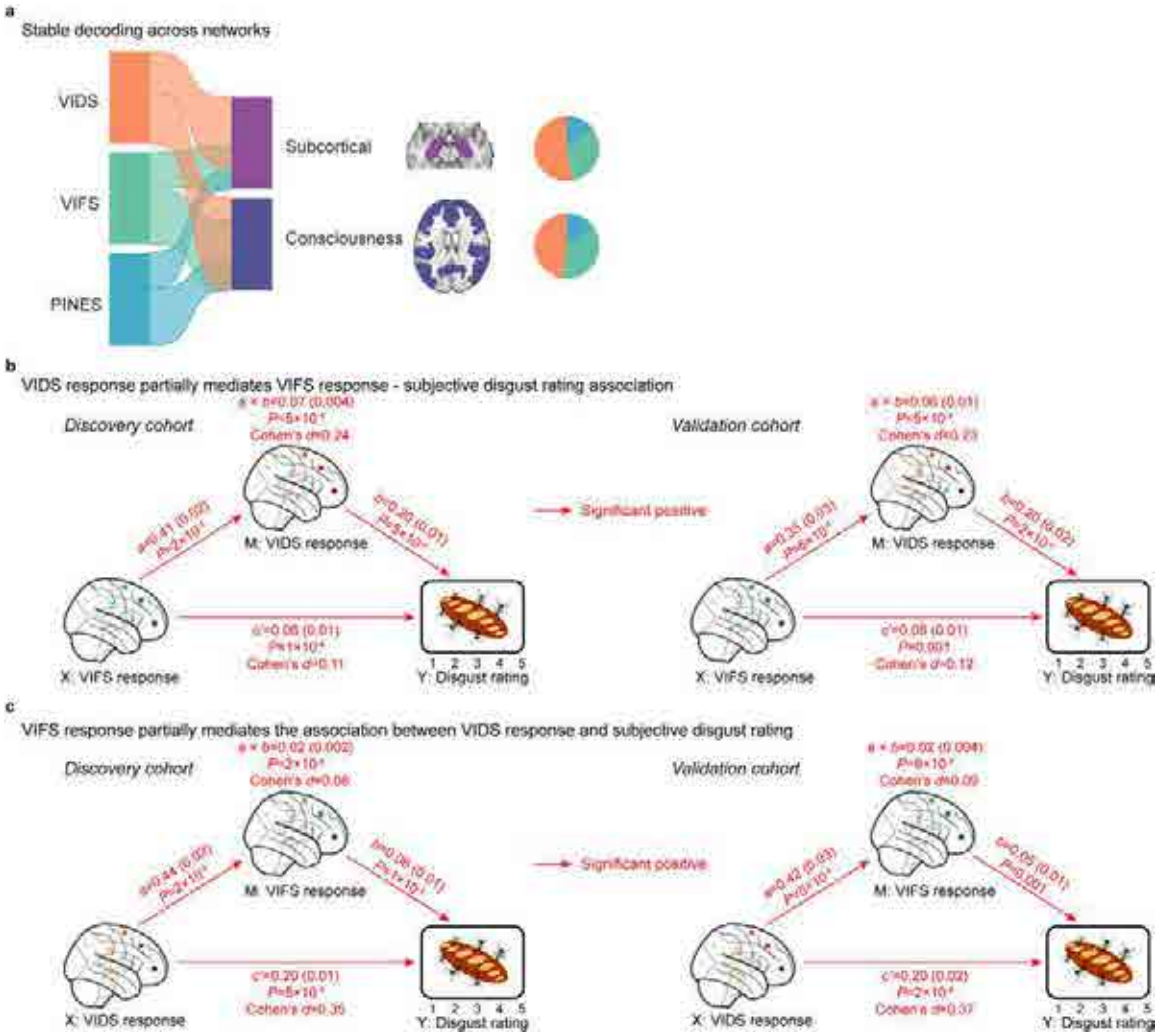
Extended Data Fig. 7 | Predictions of models trained on discovery cohort on validation cohort. a,b, Brain regions that can significantly predict subjective disgust revealed by searchlight- (Panel **a**) and parcellation-based (Panel **b**) analyses, respectively. Statistical significance was evaluated by prediction–outcome correlation (Pearson; two-sided; $P < 0.001$, uncorrected). Histograms: Predictions (correlations) from searchlights (Panel **a**) and parcellations (Panel **b**), respectively. The orange line indicates the prediction-outcome correlation from VIDS. **c,d**, Predictions (mean ± s.e.m.) from insula- (Panel **c**) and amygdala-based (Panel **d**) prediction analyses, respectively. Error bar indicates the s.e.m.; r indicates overall (between- and within-subjects; that is, $n = 149$ pairs) prediction-outcome Pearson correlation coefficient. **e**, The information about subjective experience of disgust is distributed across multiple systems. Model performance was evaluated as increasing numbers of voxels/features (x-axis) were used to predict subjective disgust in different regions of interest including the entire brain (black), consciousness network, subcortical regions or large-scale cerebral networks. The y-axis denotes the prediction-outcome correlation. Colored dots indicate the mean correlation coefficients, solid lines indicate the mean parametric fit and shaded regions indicate the s.d.

https://doi.org/10.1038/s41562-024-01868-x



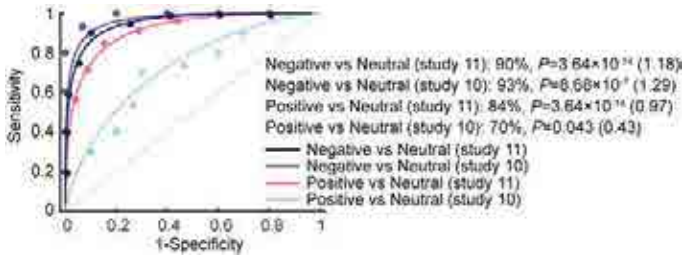
Extended Data Fig. 8 | Physical pain empathy decoder predicts disgust experience. a, The physical pain empathy decoder could predict high versus low (shown as forced-choice classification accuracy, *P* value and Cohen's d) and high versus moderate disgust, nonetheless, it fails to discriminate moderate versus low disgust in the discovery cohort. **b**, The classification results of the physical pain empathy decoder in the validation cohort, which replicates the findings as shown in (**a**). *P* values were based on binomial tests, two-sided (uncorrected).

https://doi.org/10.1038/s41562-024-01868-x



Extended Data Fig. 9 | Comparing VIDS, PINES and VIFS. a, River plots displaying spatial similarity (calculated as cosine similarity) between the stable decoding maps and the subcortical as well as consciousness network. Ribbons are normalized by the max cosine similarity across networks. Stable decoding models were thresholded at FDR $q < 0.05$ and positive voxels were retained only for similarity calculation and interpretation. Ribbon locations in relation to the boxes are arbitrary. Pie charts show relative contributions of each model to each network (that is, percentage of voxels with highest cosine similarity for each map). **b**, The multilevel mediation analytic results showing that VIDS response partially mediates the association between VIFS response and subjective disgust rating in both discovery and validation cohorts. **c**, The VIFS response plays a partial mediation role in the effect of VIDS response on the disgust rating. In **b** and **c**, the mediation analysis examines whether the observed covariance between the independent variable (X) and the dependent variable (Y) can be explained by the third variable (M, also mediator), details see Methods section. Two-sided P values are based on bootstrap tests with 10,000 samples, uncorrected.

https://doi.org/10.1038/s41562-024-01868-x



Extended Data Fig. 10 | VIDS predicts negative/positive versus neutral emotion. The VIDS reacted somehow to negative/positive versus neutral emotion (shown as forced-choice classification accuracy, *P* value and Cohen's d). *P* values were based on two-sided independent binomial tests (uncorrected).