Interoception of breathing and its relationship with anxiety

Olivia K. Harrison[1,2,3*], Laura Nanz[1], Stephanie Marino[1], Roger Lüchinger[4], Franciszek Hennel[4], Alexander J. Hess[1], Stefan Frässle[1], Sandra Iglesias[1], Fabien Vinckier[1,5,6], Frederike Petzschner[1], Samuel J. Harrison[1,3], Klaas E. Stephan[1,7]

[1] Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Switzerland
[2] School of Pharmacy, University of Otago, New Zealand
[3] Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom
[4] Institute for Biomedical Engineering, University of Zürich and ETH Zürich, Switzerland
[5] Université de Paris, France
[6] Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, France
[7] Max Planck Institute for Metabolism Research, Germany

Article type: Research article

Keywords: interoception, anxiety, breathing, inspiratory resistance

* Corresponding author:
Dr Olivia Harrison (née Faull)
Translational Neuromodeling Unit
Institute for Biomedical Engineering
University of Zurich and ETH Zurich

Abstract

Interoception, the perception of bodily states, is thought to be inextricably linked to affective qualities such as anxiety. While interoception spans sensory to metacognitive processing, it is not clear whether anxiety is differentially related to these processing levels. Here we investigated this question in the domain of breathing, using computational modelling and high-field (7 Tesla) fMRI to assess brain activity relating to dynamic changes in respiratory resistance of varying predictability. Notably, the anterior insula was associated with both interoceptive prediction certainty and prediction errors, suggesting an important role in representing and updating models of the body. Individuals with low vs. moderate anxiety traits showed differential anterior insula activity for prediction certainty. Multimodal analyses of data from fMRI, computational assessments of metacognition, and questionnaires demonstrated that anxiety-interoception links span all levels, from perceptual sensitivity to metacognition, with the largest effects seen at higher levels of interoceptive processes.

Introduction

We perceive the world through our body. While questions regarding how we sense and interpret our external environment (exteroception) have been highly prominent across centuries of research, the importance and cognitive mechanisms of monitoring our internal environment have only more recently gained traction within the neuroscience community[1–4]. 'Interoception', the perception of our body and inner physiological condition[2], constitutes a fundamental component of cerebral processes for maintaining bodily homeostasis[5–9]. However, it has also been suggested to play a wider role within systems governing emotion, social cognition and decision making[4,10]. An impaired ability to monitor bodily signals has also been hypothesised to exist across a host of psychiatric illnesses[11,12], and in particular for anxiety[13,14]. As sympathetic arousal is a reflexive response to a perceived threat, many symptoms associated with anxiety manifest themselves in the body (such as a racing heart or shortness of breath). Conversely, perceiving bodily states compatible with sympathetic arousal in the absence of external triggers can itself induce anxiety[14]. Miscommunications between the brain and body are thus thought to represent a key component of anxiety, where bodily sensations may be under-, over- or mis-interpreted[13], to initiate and perpetuate symptoms.

Studying interoception is not without significant challenges, as bodily signals are both noisy and difficult to safely manipulate[11]. Controlled manipulations of respiratory processes represent a promising way to address these challenges: suitable experimental setups allow for dynamic yet safe changes in respiration[15–22]; furthermore, given the vitally important role of breathing for survival, respiratory changes are highly salient. Indeed, laboured, unsatisfied, unexpected or uncontrolled breathing can itself be perceived as a dangerous and debilitating interoceptive threat[23–25]. Beyond respiratory diseases[24–29], aversive breathing symptoms have been noted to be particularly prevalent in many individuals suffering from psychiatric conditions such as anxiety and panic disorder[14,30–34].

Work towards conceptualising interoceptive dimensions has provided us with a framework to integrate the growing body of interoception research. Instead of treating interoception as a single entity, studies now consider both different sensory channels (e.g., organ-specific and humoral signals) and cognitive layers of interoceptive processing[35]. These layers encompass multiple levels, ranging from metrics of afferent signal strength at 'lower' levels (using techniques such as heartbeat evoked potentials)[36,37] and psychophysical properties (such as measuring perceptual sensitivity[38–40]) to psychological and cognitive components at 'higher' levels[35]. Notable domains within these higher levels include attention toward bodily signals[15,41,42], static and dynamic beliefs and models of body state[2,4,35], and insight into both our interoceptive abilities[43–46] and the accuracy of our interoceptive beliefs[6,40] ('metacognition'). Importantly, research into dynamic models of body state has also connected the interoceptive literature to that of learning, where influential (Bayesian) theories of inference about the external world, e.g. predictive coding[47–50], have been extended to interoception and propose how brains may build models of the changing internal environment[2,3,6,51,52].

Here, we build on these conceptual advances and assess the relationship between anxiety and breathing-related interoception across the multiple hierarchical levels of processing. While previous work has investigated links between anxiety and lower-level breathing sensitivity[44,53], higher-level beliefs[13,45,54,55] or metacognition[56] in isolation, a unifying perspective is yet to emerge and the relative size of anxiety-associated effects across these hierarchical levels is not known. Similarly, we lack insights into dynamic (trial-by-trial) interoceptive processes and underlying neurophysiological mechanisms. To address these issues, we adopted a multimodal experimental approach: we investigated multiple levels of breathing-related interoceptive processing, including low-level perceptual sensitivity and related higher-level metacognition via the Filter Detection Task (FDT)[46], subjective interoceptive beliefs via questionnaires, and trial-by-trial behaviour and brain activity in a novel Breathing Learning Task (BLT). Both the FDT and trial-by-trial behavioural and functional magnetic resonance imaging (fMRI) data from the BLT were analysed with separate computational models. All tasks were performed by two matched groups of low and moderate anxiety individuals, allowing us to evaluate the relationship between anxiety and each level of interoceptive processing across the hierarchy, from interoceptive sensitivity to metacognition.

Methods

Participants

Thirty individuals (pre-screened online for MRI compatibility, right handedness, non-smoking status, and no history of major somatic or psychological conditions) were recruited into each of two groups, either with very low anxiety (score of 20-25 on the Spielberger State-Trait Anxiety Inventory [57]; STAI-T), or moderate anxiety (score>=35 STAI-T). The resulting mean (±std) trait anxiety score for the low anxiety group was 23.2±1.8 and for the moderate anxiety group 38.6±4.6. Groups were matched for age and sex (15 females in each group), with mean (±std) ages of 25.4±3.9 and 24.2±5.0 years for low and moderate anxiety groups, respectively. Behavioural data (not used in any other analyses) from an additional 8 participants (four from each group, two females in each) served to determine model priors. All participants signed a written, informed consent, and the study was approved by the Cantonal Ethics Committee Zurich (Ethics approval BASEC-No. 2017-02330). Each participant completed three tasks over two testing sessions: a behavioural session that included questionnaires and a task probing interoceptive sensitivity and metacognition (the Filter Detection Task, or FDT), and a brain imaging session where the Breathing Learning Task (BLT) was paired with fMRI. Each of these tasks and analyses are described below, and all analyses were pre-specified in time-stamped analysis plans (https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety).

Questionnaires

The main questionnaire set employed was designed to firstly capture subjective affective measures, and secondly both general and breathing-specific subjective interoceptive beliefs. The assignment of participants to groups was based on the Spielberger Trait Anxiety Inventory (STAI-T)[57]. Affective qualities that were additionally assessed included state anxiety (Spielberger State Anxiety Inventory; STAI-S[57]), symptoms that are part of anxiety disorder (Generalised Anxiety Disorder Questionnaire; GAD-7[58]), anxiety sensitivity (anxiety regarding the symptoms of anxiety; Anxiety Sensitivity Index; ASI-3[59]), and symptoms of depression (Centre for Epidemiologic Studies Depression Scale; CES-D[60]). To obtain self-reports of body awareness we used the Body Perception Questionnaire (BPQ)[61], while the Multidimensional Assessment of Interoceptive Awareness Questionnaire (MAIA)[62] was used to measure positive and 'mindful' attention towards body symptoms. We also measured breathing-related catastrophising using the Pain Catastrophising Scale (PCS-B)[63], and breathing-related vigilance using the Pain Vigilance Awareness Questionnaire (PVQ-B)[64] (in both questionnaires, the word 'breathing' was substituted for 'pain'). Finally, the following supplementary questionnaires were included to explore possible contributing factors (e.g. general positive and negative affect, resilience, self-efficacy and fatigue): Positive Affect Negative Affect Schedule (PANAS-T)[65], Connor-Davidson Resilience Scale[66], General Self-Efficacy Scale[67], Fatigue Severity Scale (FSS)[68]. The STAI-T and CES-D were completed online as part of the pre-screening process; all other questionnaires were completed in the behavioural session at the laboratory.

Analysis: Group differences were tested individually for the 13 scores resulting from the 12 questionnaires, with all questionnaires included except the trait anxiety score that was used to screen participants and assign them to groups. The data that was used for group comparisons across all modalities were first tested for normality (Anderson-Darling test, with $p<0.05$ rejecting the null hypothesis of normally distributed data), and group differences were determined using either two-tailed independent t-tests or Wilcoxon rank sum tests. For the questionnaires, Bonferroni correction for the 13 tests was applied, requiring $p<0.004$ for a corrected significant group difference. Results with $p<0.05$ not surviving correction are reported as exploratory for questionnaires as well as all other data. In a secondary exploratory step, group difference analyses were then conducted on the questionnaires' subcomponent scores (22 scores); please see Supplementary Material.

Filter detection task

Stimuli and task description: To systematically test properties of breathing perception and related metacognition, we utilised a perceptual threshold breathing task (the Filter Detection Task; FDT)[46]. The FDT was used to determine interoceptive perceptual sensitivity, decision bias, metacognitive bias (self-reported confidence) and metacognitive

performance (congruency between performance and confidence scores) regarding detection of very small variations in an inspiratory load. In this task (outlined in Figure 2A), following three baseline breaths either an inspiratory load was created via the replacement of an empty filter with combinations of clinical breathing filters, or the empty filter was removed and restored onto the system (sham condition) for three further breaths. All filter changes were performed behind participants, out of their field of view. After each trial of six breaths, participants were asked to decide whether or not a load had been added, as well as reporting their confidence in their decision on a scale of 1-10 (1=not at all confident in decision, 10=extremely confident in decision). An adapted staircase algorithm was utilised to alter task difficulty until participants were between 60-85% accuracy[46], and 60 trials were completed at the corresponding level of filter load. Respiratory threshold detection[45], metacognitive bias[69] and interoceptive metacognitive performance[56] have previously been linked to anxiety symptomology.

Analysis: Breathing-related interoceptive sensitivity (i.e. perceptual threshold) was taken as the number of filters required to keep task performance between ~60-85% accuracy. Both decision bias and metacognitive performance from the FDT were analysed using the hierarchical HMeta-d statistical model[70]. This model firstly utilises signal detection theory[71] to provide single subject parameter estimates for task difficulty (d'; not analysed as performance is fixed between 60-85% by design) and decision bias (*c*, akin to over- or under-reporting the presence of resistance with values below and above zero, respectively), as well as using a hierarchical Bayesian formulation of metacognitive performance (Mratio, calculated by fitting metacognitive sensitivity meta-d', then normalising by single subject values for d'). Finally, metacognitive bias was calculated as the average confidence scores across all analysed trials. Hypothesised group differences were based on previous findings, where respiratory threshold detection level was hypothesised to be higher[44,53], perceptual decisions biased towards 'yes' (or deciding the resistance was present; denoted by more negative values for *c*), metacognitive bias to be lower[69] and interoceptive metacognitive performance to be lower[56] with greater anxiety. All hypotheses had been pre-specified in an analysis plan (https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety). Bonferroni correction for the four tests was applied, requiring $p<0.013$ for a corrected significant group difference.

Breathing Learning Task

Stimuli and task description: To measure behaviour and brain activity concerning the dynamic updating of interoceptive beliefs or expectations under uncertainty, a novel associative learning task was developed and employed during functional magnetic resonance imaging (fMRI). In this Breathing Learning Task (BLT), 80 trials were performed where on each trial two visual cues were paired with either 80% or 20% chance of a subsequent inspiratory resistive load. The visual information for the task was presented through the VisualStim system (Resonance Technology, Northridge, CA, US). As outlined in Figure 1, participants were required to explicitly predict (via button press) whether they would experience a breathing resistance following the presentation of one of the cues. Following this prediction and a short (2.5s) pause, a circle appeared on the screen to indicate the stimulus period (5s), where participants either experienced inspiratory resistance (70% of their maximal inspiratory resistance, measured in the laboratory, delivered via a PowerBreathe KH2; PowerBreathe International Ltd, Warwickshire, UK) or no resistance was applied. Rest periods of 7-9s were pseudo-randomised between trials. For the inspiratory resistances we used a mechanical breathing system that allows for remote administration and monitoring of inspiratory resistive loads (for technical details on resistance administration see Supplementary Figure 1 and previous work[22]). The cue presentations were balanced such that half of all trials delivered the inspiratory resistance. Following an initial stable period of 30 trials, the stimulus-association pairing was swapped four times during the remainder of the 80 trials (i.e., repeated reversals; Figure 1). Participants were explicitly told that the cues acted as a matched pair that could only swap in probability. The trial sequence was pseudorandom and fixed across subjects to ensure comparability of the induced learning process. Following every stimulus, participants were asked to rate 'How difficult was it to breathe?", on a visual analogue scale (VAS) from "Not at all difficult" to "Extremely difficult". Immediately following the final trial of the task, participants were also asked to rate "How anxious were you about your breathing" on a VAS from "Not at all anxious" to "Extremely anxious".

Two representations of trial-wise quantities were employed for subsequent analyses of data from this task. First, a computational model (see below) provided dynamic estimates of both predictions and prediction errors on

each trial. Second, a standard categorical approach represented trial-by-trial whether the subjects' prediction decisions indicated the anticipated presence or absence of an upcoming inspiratory resistance, as well as unsurprising (i.e. following correct predictions) and surprising (i.e. following incorrect predictions) respiratory stimuli. The latter results are presented in the Supplementary Material.

Computational modelling of behavioural data: For the trial-by-trial analysis of behavioural data from the BLT, we considered three computational models that are routinely used for associative learning tasks. This included a Rescorla Wagner (RW) model (Equation 1) and 2 variants of the Hierarchical Gaussian Filter (HGF) with 2 or 3 levels (HGF2 and HGF3). While the RW model assumes a fixed learning rate, the HF allows for online adaption of learning rate as a function of volatility. All learning models were paired with a unit-square sigmoid response model (Equation 2) and were implemented using the Hierarchical Gaussian Filter Toolbox[72,73] (version 5.3) from the open-source TAPAS software[74] (http://www.translationalneuromodeling.org/tapas/). The alternative models were formally compared using random effects Bayesian model selection (BMS) as implemented in SPM12[75,76]. BMS utilises the log model evidence (LME) to determine the most likely amongst a set of competing hypotheses (i.e. models) that may have generated observed data, and is robust to outliers[75]. Our analysis plan had specified that a model would be chosen as the 'winning' model if it demonstrated a protected exceedance probability (PXP) greater than 90%. As explained in the Results section, none of our models reached this criterion (although simulations indicated that the proposed models could in principle be differentiated; see Supplementary Material for details). We therefore applied the simplest of the models considered (i.e. the RW model), as pre-specified in our analysis plan (https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety).

In our application of the RW model as a perceptual model, the update equation corresponded to a simple delta-learning rule with a single free parameter, the learning rate[77]:

$$v_{(k+1)} = v_{(k)} + \alpha\delta_{(k)}$$

(Equation 1)

where $v_{(k+1)}$ is the predicted probability for a specific outcome (encoded as 0 or 1) on trial $(k+1)$, $v_{(k)}$ is the estimated outcome probability on the k^{th} trial, $\alpha \in [0, 1]$ is a constant learning rate parameter, and $\delta_{(k)}$ is the prediction error magnitude at trial k.

The above perceptual model was paired with a unit-square sigmoid response model[72]. This response model accounts for decision noise by mapping the predicted probability $v_{(k)}$ that the next outcome will be 1 onto the probabilities $p(y_{(k)} = 1)$ and $p(y_{(k)} = 0)$ that the agent will choose response 1 or 0, respectively:

$$p(y_{(k)}|v_{(k)}, \zeta) = \left(\frac{{v_{(k)}}^{\zeta}}{{v_{(k)}}^{\zeta} + (1 - v_{(k)})^{\zeta}}\right)^{y_{(k)}} \left(\frac{(1 - v_{(k)})^{\zeta}}{{v_{(k)}}^{\zeta} + (1 - v_{(k)})^{\zeta}}\right)^{1-y_{(k)}}$$

(Equation 2)

Here, $y_{(k)}$ represents the expressed decision of a subject given the cue (contingency pairs) on trial k. The parameter ζ captures how deterministically y is associated with v. The higher ζ, the more likely the agent is to choose the option that is more in line with its current prediction. The decision model uses the perceptual model indirectly via its inversion[72], given the trajectories of trial-wise cues and responses (see Figure 1).

In our paradigm, trial-wise outcomes are categorical (resistance vs. no resistance), which raises the question of how outcomes should be coded in the computational model. One way would be to model two trajectories, separately for resistance and no resistance outcomes, and indicate on any given trial whether the respective outcome has occurred (1) or not (0). However, due to the fixed coupling of contingencies in our paradigm (see above) – which the participants were explicitly instructed about – a computationally more efficient approach that requires only a single model is to code the outcome in relation to the cue. Here, we adopted this coding in "contingency space", following the same procedure as in the supplementary material of Iglesias and colleagues[78]. Specifically, due to the

fixed coupling of contingencies in our paradigm (see above), we represented the occurrence of "no resistance" given one cue and the occurrence of "resistance" given the other cue as 1, and both other cue-outcome combinations as 0 (note that under the subsequent transformations described below, the resulting trajectories of predictions and prediction errors would remain identical if the opposite choice had been made).

Maximum a posteriori (MAP) parameter estimates were obtained using the Brayden-Fletcher-Goldfarb-Shanno algorithm, as implemented in the HGF toolbox. Prior means and variances were determined using the distribution of maximum likelihood estimates fit across 8 pilot participants who were distinct from the participants of our study (see Supplementary Material for prior means and variances determined from the pilot data).

Group differences in model parameter estimates of learning rate (α) and inverse decision temperature (ζ), as well as perception measures of stimulus intensity (averaged across all trials), breathing-related anxiety (rated immediately following the task) and prediction response times were tested. Bonferroni correction for five tests was applied, requiring $p<0.01$ for a corrected significant group difference. Results from additional exploratory models encompassing anxiety, depression and gender are reported in the Supplementary Material.

Following random effects Bayesian model selection (BMS[75,76]), the chosen model was examined in each participant with regard to whether it demonstrated an adequate fit. To this end, model fit in each individual was compared to the likelihood of obtaining the data by chance[79] using the likelihood ratio test (*lratiotest* function) provided in MATLAB. The final behavioural and brain imaging analyses presented here were run without two subjects in which non-significant ($p>0.05$) differences to randomness were encountered.

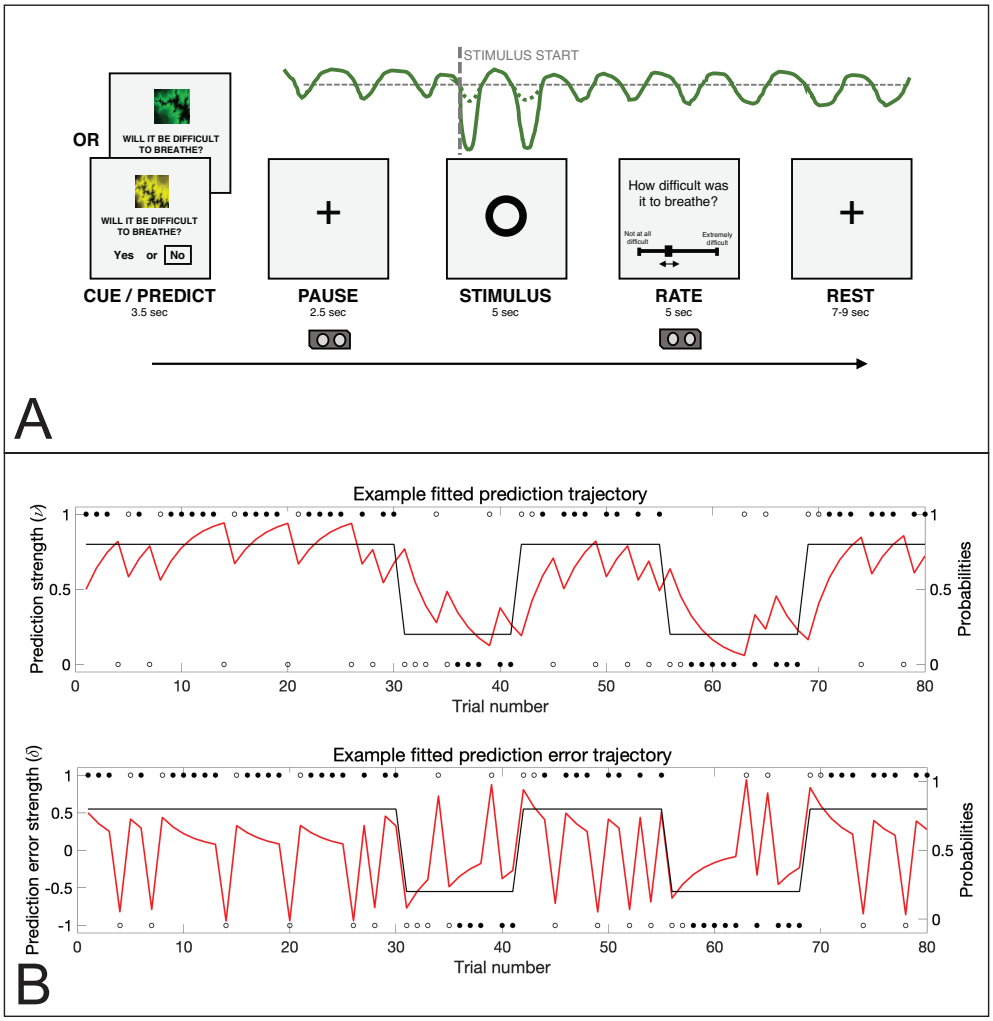


Figure 1. The 'Breathing Learning Task' (BLT), used to measure dynamic learning of breathing-related stimuli. A) An overview of the single trial structure, where one of two cues were presented and participants were asked to predict (based on the cue) whether they thought that an inspiratory breathing resistance would follow. When the circle appeared on the screen, either an inspiratory resistance or no resistance was applied for 5 seconds, with the resistance set to 70% of the individual's maximal

inspiratory resistance. After every trial, participants were asked to rate the intensity of the previous stimulus. The trace in green is an example of a pressure trace recorded at the mouth. B) The 80-trial trajectory structure of the probability that one cue predicts inspiratory resistance (black trace), where the alternative cue has an exactly mirrored contingency structure, together with example responses (circles). Filled black circles represent stimuli that were correctly predicted, and open black circles represent stimuli that were not correctly predicted. Example fitted prediction certainty (top) and prediction error (bottom) trajectories are overlaid (red traces). The example trajectories were taken from the participant with the closest learning rate to the mean value across all participants.

Constructing computationally informed regressors: The trajectories of predictions and prediction errors estimated by the RW model were used to construct regressors representing computational trial-by-trial quantities of interest for subsequent GLM analyses. In order to investigate the salient effects of inspiratory resistance as an interoceptive stimulus, we separated trials into "negative" (occurrence of resistance) and "positive" (no resistance) events and represented these events by separate regressors in the GLM (see Figure 4). To achieve this, we first transformed both the original prediction and prediction error values (estimated in contingency space) back into the stimulus space, according to the cue presented at each trial:

$$v_{(k)}^{stim} \stackrel{\text{def}}{\equiv} \begin{cases} v_{(k)}, & if\ cue\ type = 1 \\ 1 - v_{(k)}, & if\ cue\ type = 2 \end{cases}$$

(Equation 3)

$$\delta_{(k)}^{stim} \stackrel{\text{def}}{\equiv} \begin{cases} \delta_{(k)}, & if\ cue\ type = 1 \\ -\delta_{(k)}, & if\ cue\ type = 2 \end{cases}$$

(Equation 4)

Here, $v_{(k)}^{stim}$ and $\delta_{(k)}^{stim}$ now represent the prediction and prediction error values in stimulus space, with $v_{(k)}^{stim} = 1$ representing maximal predictions of no resistance and $v_{(k)}^{stim} = 0$ maximal predictions of resistance. Similarly, $\delta_{(k)}^{stim} = 1$ represents maximal prediction errors of no resistance and $\delta_{(k)}^{stim} = -1$ maximal prediction errors of resistance (see Supplementary Figure 5 for details).

Secondly, trial-wise prediction values were then transformed to represent the deviation from maximally uninformed predictions (i.e., guessing), by taking the distance from 0.5 (see Equations 5 and 6). In the RW model prediction values are probabilities bounded by 0 and 1, hence the distance from 'guessing' (at 0.5) reflects the 'certainty' by which the absence or presence of respiratory resistance was predicted. This simple transformation enabled us to take into account the role of (un)certainty of predictions – which plays a crucial role in interoception-oriented theories of anxiety[13,80] but, in contrast to Bayesian models, is not represented explicitly in the RW model. Specifically, separately for the two event types, we defined certainty of positive predictions (no resistance) and of negative predictions (resistance) as the absolute deviation from a prediction with maximum uncertainty (i.e., 0.5):

$$\text{If } v_{(k)}^{stim} > 0.5 \quad v_{(k)}^{pos} \stackrel{\text{def}}{\equiv} v_{(k)}^{stim} - 0.5$$

(Equation 5)

$$\text{If } v_{(k)}^{stim} < 0.5 \quad v_{(k)}^{neg} \stackrel{\text{def}}{\equiv} 0.5 - v_{(k)}^{stim}$$

(Equation 6)

Here, both $v_{(k)}^{pos}$ and $v_{(k)}^{neg}$ exist between 0 and 0.5, with values closer to zero indicating less certain predictions.

Like predictions, prediction errors were also divided between positive (no resistance) and negative events (resistance) values. This was again determined as the absolute deviation from the mid-point of the prediction errors (i.e., 0):

$$\text{If } \delta_{(k)} > 0 \qquad \delta_{(k)}^{pos} \stackrel{\text{def}}{=} \delta_{(k)}$$

(Equation 7)

$$\text{If } \delta_{(k)} < 0 \qquad \delta_{(k)}^{neg} \stackrel{\text{def}}{=} -\delta_{(k)}$$

(Equation 8)

Here, both $\delta_{(k)}^{pos}$ and $\delta_{(k)}^{neg}$ exist between 0 and 1, with values closer to zero indicating smaller prediction errors. Note that this derivation gives prediction error values identical to those that would have been obtained by modelling two separate trajectories for resistance and no resistance outcomes (see above and [78]).

Physiological data processing: Physiological data were recorded at a sampling rate of 1000 Hz, and included heart rate, chest distension, pressure of expired carbon dioxide ($P_{ET}CO_2$) and oxygen ($P_{ET}O_2$), and pressure at the mouth (for equipment details see[22] and Supplementary Material). In addition to the task, small boluses of a CO_2 gas mixture (20% CO_2; 21% O_2; balance N_2) were administered during some rest periods, allowing for de-correlation of any changes in $P_{ET}CO_2$ from task-related neural activity, as previously described[18,20,21,81].

Physiological noise regressors were prepared for inclusion into single-subject general linear models (GLMs, described below). Linear interpolation between $P_{ET}CO_2$ peaks was used to form an additional CO_2 noise regressor, which was convolved using a response function based on the haemodynamic response function (HRF) provided by SPM with delays of 10s and 20s for the overshoot and undershoot, respectively[82]. Temporal and dispersion derivatives of this CO_2 noise regressor were also included. An additional three cardiac- and four respiratory-related waveforms (plus one interaction term) were created using PhysIO[83]. Four respiratory volume per unit time (RVT) regressors (delays: -5, 0, 5, 10) were created using the Hilbert-transform estimator in PhysIO[84], and convolution with a respiratory response function[83].

Magnetic resonance imaging: MRI was performed using a 7 Tesla scanner (Philips Medical Systems: Achieva, Philips Healthcare, Amsterdam, The Netherlands) and a 32 channel Head Coil (Nova Medical, Wilmington, Massachusetts, United States of America). A T2*-weighted, gradient echo EPI was used for functional scanning, using a reduced field of view (FOV) with an axial-oblique volume centred over the insula and midbrain structures. The FOV comprised 32 slices (sequence parameters: TE 30ms; TR 2.3s; flip angle 75°; voxel size 1.5x1.5x1.5mm; slice gap 0.15mm; SENSE factor 3; ascending slice acquisition), with 860 volumes (scan duration 33 mins 9s). A matched whole-brain EPI scan (96 slices) was immediately acquired following the task scan for registration purposes. Additionally, a whole-brain T1-weighted structural scan with 200 slices was acquired (MPRAGE, sequence parameters: TE 4.6ms; TR 10ms; segment-TR 3000ms; TI 1000ms; flip angle 8°; voxel size 0.8x0.8x0.8mm; bandwidth; 153.1Hz/Px; sagittal slice orientation). Finally, a task-free (resting-state) functional scan (250 volumes) was obtained, with participants instructed to keep their eyes open and fixating a white fixation cross on a black screen.

MRI preprocessing: MRI data analysis was performed using a combination of FSL version 6.0.1 (the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain Software Library, Oxford, UK[85]) and SPM12 (Statistical Parametric Mapping software, London, UK) as prespecified in our analysis plan (https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety). Image preprocessing was performed using FSL, including motion correction (MCFLIRT[86]), removal of non-brain structures (BET[87]), and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting; 100s cut-off period)[88]. Independent component analysis (ICA) was used to identify noise due to motion, scanner and cerebrospinal fluid artefacts[89], and the timeseries of these noise components were entered into single-subject GLMs (described below) as nuisance regressors. The functional scans were registered to the MNI152 (1x1x1mm) standard space using a three-step process: 1) Linear registration (FLIRT) with 6 degrees of freedom (DOF) to align the partial FOV scan to the whole-brain EPI image[90]; 2) Boundary-based registration (BBR; part of the FMRI Expert Analysis Tool, FEAT) with 12 DOF and a weighting mask of the midbrain and insula cortex to align the whole-brain EPI to T1 structural image; and 3) Non-linear registration using a combination of FLIRT and FNIRT[91] to align the T1 structural scan to 1mm standard space. Functional MRI scans were resampled once into

standard space with a concatenated warp from all three registration steps, and then spatial smoothing in standard space was performed using a Gaussian kernel with 3mm full-width half-maximum using the fslmaths tool.

Single-subject general linear model: Single-subject estimates of the general linear model (GLM) were performed using SPM. A GLM was constructed for each of the participants, with a design matrix informed by trial-wise estimates from the RW model of each participant (see above). An additional analysis, using a more classical (non-computational model-based) design matrix, is presented in the Supplementary Material. Alongside the task regressors described below, rigorous de-noising was performed by the inclusion of the following regressors (all described above): the convolved end-tidal CO_2 regressor plus temporal and dispersion derivatives, six motion regressors trajectories plus their first-order derivatives, physiological noise regressors (provided by the PhysIO toolbox) and ICA components identified as noise.

The regressors of interest in the design matrix were as follows (compare Figure 1A):

1) A 'Cue' regressor (80 repeats), with onsets and durations (2.5s) determined by the presentations of visual cues and a magnitude of 1;
2) A 'Positive prediction' regressor, with onsets given by the presentation of each corresponding visual cue (when no-resistance was predicted), durations of 0.5s and magnitudes given by $v_{(k)}^{pos}$ in Equation 5;
3) A 'Negative prediction' regressor, with onsets given by the presentation of each corresponding visual cue (when resistance was predicted), durations of 0.5s and magnitudes given by $v_{(k)}^{neg}$ in Equation 6;
4) A 'No resistance' stimulus regressor, with onset timings according to the first inspiration that occurred after the presentation of the visual cue, and durations as the remaining time of the potential resistance period (circle in Figure 1), with a magnitude of 1;
5) A 'Resistance' stimulus regressor, with onset timings according to the initiation of the inspiratory resistance (identified from the downward inflection of the inspiratory pressure trace) after the presentation of the visual cue, and durations as the remaining time of the resistance period (circle in Figure 1), with a magnitude of 1;
6) A 'Positive prediction error' regressor, with onsets given by the start of each corresponding no resistance period, durations of 0.5 s and magnitudes given by $\delta_{(k)}^{pos}$ in Equation 7;
7) A 'Negative prediction error' regressor, with onsets given by the start of each corresponding resistance period, durations of 0.5 s and magnitudes given by $\delta_{(k)}^{neg}$ in Equation 8;
8) A 'Rating period' noise regressor, with onsets and durations covering the period where participants were asked to rate the difficulty of the previous stimulus, and with a magnitude of 1.

Regressors 1-8 were included in the design matrix after convolution with a standard HRF in SPM12, together with their temporal and dispersion derivatives. Contrasts of interest from this design examined brain activity associated with the average across positive and negative valence for both predictions and prediction errors, as well as the difference due to valence (i.e. positive vs. negative) for both predictions and prediction errors.

Group analysis: Firstly, for the analysis of our entire field of view, contrasts of interest were assessed using random effects group-level GLM analyses based on the summary statistics approach in SPM12. The group-level GLM consisted of a factorial design with both a group mean and group difference regressor. The analyses used a significance level of $p<0.05$ with family-wise error (FWE) correction at the cluster-level, with a cluster-defining threshold of $p<0.001$. Secondly, for our region of interest (ROI) analysis, we used FSL's non-parametric threshold-free cluster enhancement[92] within a combined mask of the anterior insula and periaqueductal gray (PAG). This analysis employed a significance level of $p<0.05$, with FWE correction across the joint mask. While the anterior insula and PAG have previously been shown to be involved in both conditioned anticipation and perception of inspiratory resistances[15,16,18,20,21,93] as well as prediction errors[94] and precision[95] towards pain perception, our current analysis considers computational trial-by-trial estimates of interoceptive predictions and prediction errors for the first time. The mask of the anterior insula was taken from the Brainnetome atlas[96] (bilateral ventral and dorsal anterior insula regions), and the PAG incorporated an anatomically-defined mask that has been used in previous fMRI publications[20,21].

Multi-modal analysis

Data: The different task modalities were then combined into a multi-modal analysis to assess both the relationships between and shared variance amongst measures. The data entered into this analysis consisted of:

1) The scores from the four main affective questionnaires that were not used to pre-screen the participants (STAI-S[57], GAD-7[58], ASI-3[59] and CES-D[60]);
2) The four interoceptive questionnaires (BPQ[61], MAIA[62], PCS-B[63] and PVQ-B[64]);
3) The four FDT measures (breathing sensitivity, decision bias *c*, metacognitive bias, metacognitive performance Mratio); and
4) The individual peak anterior insula activity associated with both positive and negative predictions, as well as positive and negative prediction errors. Activity was extracted from a 4mm sphere, centred on each participant's maximal contrast estimate within a Brainnetome atlas mask of the anterior insula[96], using the first eigenvariate of the data.

Multi-modal correlations and shared variance: A correlation matrix of all 16 included measures was calculated in order to visualise the relationships between all variables. The significance values of the correlation coefficients were taken as $p<0.05$ (exploratory), and a false discovery rate (FDR) correction for multiple comparisons was applied (using the *mafdr* function in MATLAB).

To assess the shared variance across measures and delineate which measures were most strongly associated with affective qualities, we entered all specified data into a principal component analysis (PCA), following normalisation using z-scoring within each variable. The number of significant components were determined by comparing the variance explained of each component to a null distribution, created by randomly shuffling (n=1000) the measures from each variable across participants. Components were considered significant if the variance explained was above the 95% confidence interval of the corresponding component's null distribution.

To assess the relationship between each of the significant components and anxiety, the component scores for low and moderate anxiety were compared using either independent t-tests or Wilcoxon rank sum tests (following Anderson-Darling tests for normality). The significance values of the group differences in component scores were taken as $p<0.05$ (exploratory), and a false discovery rate (FDR) correction for multiple comparisons (number of significant components) was applied.

An independent code review was performed on all data analysis procedures, and the analysis code is available on GitLab (https://gitlab.ethz.ch/tnu/code/harrison_breathing_anxiety_code).

Results

Questionnaire results

The group summaries and comparisons for each of the affective and interoceptive questionnaires (excluding the trait anxiety score that was used for group allocation) are displayed in Figure 2. The group summary values and statistics presented are either mean±standard error (ste) when values were normally distributed and thus compared using unpaired T-tests, or median±inter-quartile range (iqr) when values were not normally distributed and thus compared using Wilcoxon rank sum tests. Scores from all questionnaires of affective symptoms employed were found to be highly significantly different between low and moderate trait anxiety groups: Individuals with moderate levels of trait anxiety demonstrated higher state anxiety (STAI-S mean±ste; low anxiety=25.7±0.7; moderate anxiety=34.1±1.2; $T=-6.1$; $p<0.01$), higher levels of anxiety disorder symptoms (GAD-7 median±iqr; low anxiety=1.0±2.0; moderate anxiety=4.0±3.0; $Z=-5.9$; $p<0.01$), greater anxiety sensitivity (ASI mean±ste; low anxiety=6.8±0.8; moderate anxiety=18.4±1.5; $T=-6.9$; $p<0.01$), and higher levels of depression symptoms (CES-D median±iqr; low anxiety=6.5±3.0; moderate anxiety=14.0±6.0; $Z=-6.0$; $p<0.01$).

The interoceptive questionnaires we used measured 'positively-minded' interoceptive awareness, overall body awareness, breathing symptom catastrophising and breathing symptom vigilance. All except breathing-related

vigilance were also found to be highly significantly different between groups. Individuals with moderate levels of trait anxiety demonstrated reduced 'positively-minded' interoceptive awareness (MAIA mean±ste; low anxiety=109.1±4.6; moderate anxiety=84.6±3.7; T=4.2; p<0.01) and greater reports of overall body awareness (BPQ median±iqr; low anxiety=66.0±68.0; moderate anxiety=104.0±52.0; Z=-2.5; p=0.01) in line with previous research[13,45,54,55]. Additionally, elevated levels of breathing-related catastrophising were observed in the moderate anxiety group (PCS-B median±iqr; low anxiety=3.5±11.0; moderate anxiety=14.0±17.0; Z=-3.3; p<0.01), while no statistically significant difference was observed for breathing-related vigilance (PVQ-B mean±ste; low anxiety=16.3±2.2; moderate anxiety=21.4±2.5; T=-1.5; p=0.13). Results for sub-component scores and additional questionnaires can be found in the Supplementary Material.

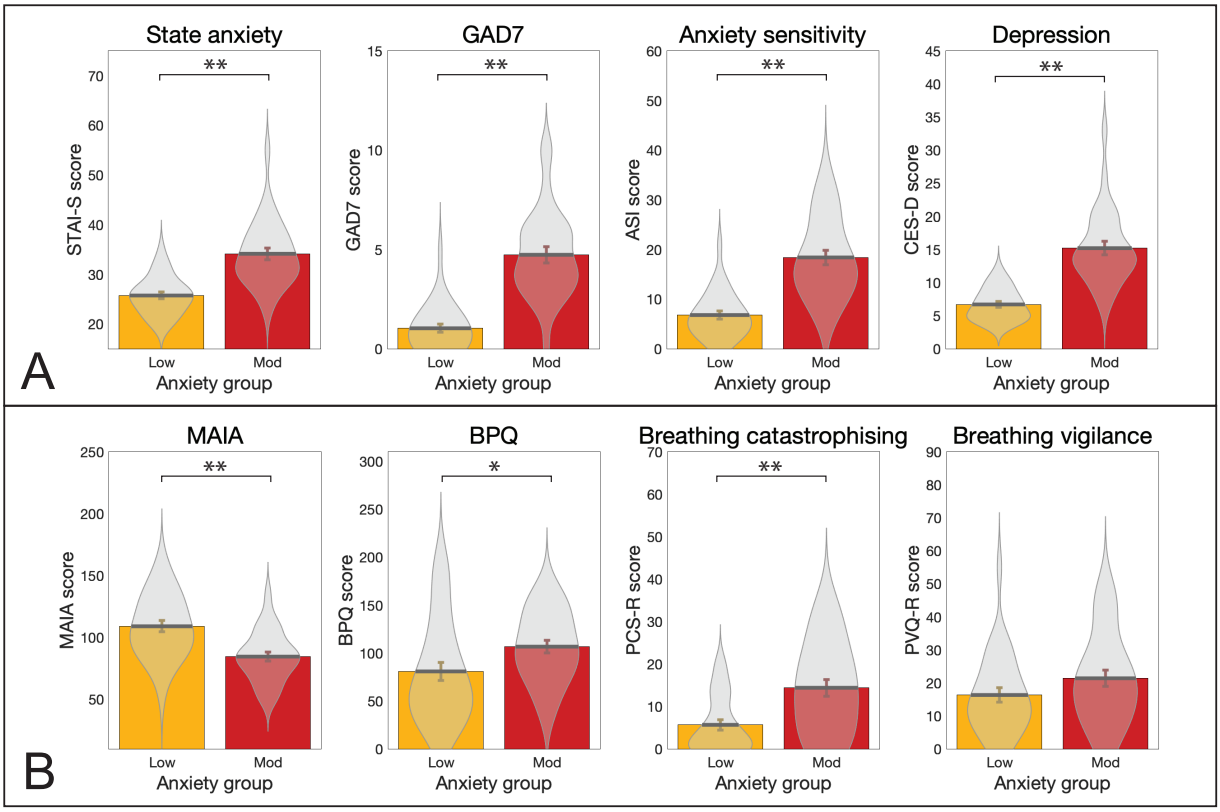


*Figure 2. Results from the affective questionnaires (A) and interoceptive questionnaires (B) measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Affective questionnaires (A): 'State anxiety', Spielberger State Anxiety Inventory; 'GAD-7', Generalised Anxiety Disorder Questionnaire; 'Anxiety sensitivity', Anxiety Sensitivity Index; 'Depression', Centre for Epidemiologic Studies Depression Scale. Interoceptive questionnaires (B): 'MAIA', Multidimensional Assessment of Interoceptive Awareness Questionnaire; 'BPQ', Body Perception Questionnaire; 'Breathing catastrophising', Pain Catastrophising Scale (with the word 'pain' substituted for 'breathing'); 'Breathing vigilance', Pain Vigilance Awareness Questionnaire (with the word 'pain' substituted for 'breathing'). *Significant at p<0.05; **Significant following Bonferroni correction for multiple comparisons across all 8 questionnaires. Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*

Filter Detection Task results

The group summaries and comparisons for each of the FDT measures are displayed in Figure 3. The FDT output includes the number of filters at perceptual threshold (indicative of perceptual sensitivity, where a greater number of filters indicates lower perceptual sensitivity), decision bias (with *c*<0 indicating a tendency to report the presence of a filter), metacognitive bias (calculated from average confidence scores) and metacognitive performance (reflecting the congruence between confidence scores and performance accuracy). Individuals with moderate levels of trait anxiety demonstrated both lower perceptual sensitivity (in line with previous findings[44,53]) (filter number median±iqr;

low anxiety=3.0±2.0; moderate anxiety=4.0±2.0; Z=-2.4; p=0.01) and lower metacognitive bias (average confidence score median±iqr; low anxiety=6.7±2.2; moderate anxiety=6.2±2.1; Z=2.0; p=0.02) than those with low levels of anxiety, with a similar level of metacognitive performance (Mratio median±iqr; low anxiety=0.8±0.0; moderate anxiety=0.8±0.1; Z=0.7; p=0.23). Decision bias was not found to be different between the groups (decision bias *c* parameter mean±ste; low anxiety=0.15±0.06; moderate anxiety=0.05±0.06; T=1.1; p<0.14) The relationship between greater anxiety and reduced confidence is consistent with results previously observed in the exteroceptive (visual) domain, where decreased confidence related to individual levels of both anxiety and depression[69].

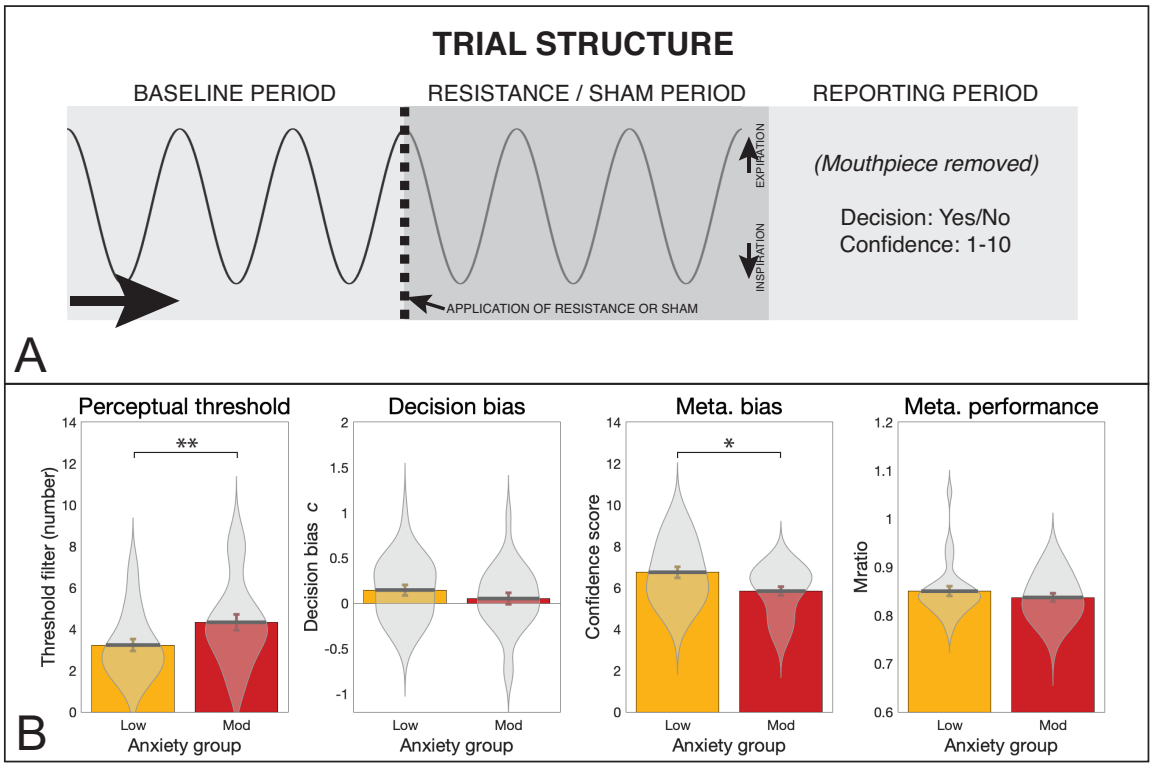


*Figure 3. A) Trial structure of the 'Filter Detection Task' (FDT). For each trial participants first took three breaths on the system ('baseline period'), before either an inspiratory resistance or sham was applied. Following three further breaths, participants removed the mouthpiece and reported their decision as to whether a resistance was added (Yes/No), and their confidence in their decision (1-10, 1=not at all confident / guessing; 10=maximally confident in their decision). B) Results from the FDT: Individuals with moderate anxiety (score of 35+ on the STAI-T[57]) demonstrated a higher (less sensitive) perceptual threshold and lower metacognitive bias (lower average confidence, independent of task accuracy) when compared to individuals with low levels of anxiety (score of 20-25 on the STAI-T[57]). No difference was found between groups for decision bias (where c values below zero indicate a tendency to report the presence of resistance) nor metacognitive performance (where higher values indicate better metacognitive performance). *Significant at p<0.05; **Significant following Bonferroni correction for multiple comparisons across all FDT measures. Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*

Breathing Learning Task results

Behavioural data modelling: When comparing the plausibility of the three alternative models (a Rescorla Wagner, RW; a 2-level Hierarchical Gaussian Filter, HGF2; and a 3-level Hierarchical Gaussian Filter, HGF3) using random effects Bayesian model selection[75], no single model was found to have a protected exceedance probability (PXP) greater than 90% (RW:HGF2:HGF3 PXP=0.30:0.40:0.30, Supplementary Table 6). Therefore, as specified in our analysis plan (https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety), we conducted our model-based analysis using the conceptually most simple model (the RW model), in accordance with Occam's Razor. Two participants (one from each anxiety group) were excluded from any further model-based analyses and comparisons since their model fit was not significantly different to a model based on chance.

Both model-based and behavioural parameter comparisons are presented in Table 1. For the estimated model parameters, no difference was observed between the groups for either learning rate (α) and inverse decision temperature (ζ). Results from parameter comparisons between groups including the excluded participants can be found in the Supplementary Material. For the subjective measures, no difference was observed between the groups for breathing difficulty ratings, while the task-induced anxiety ratings were significantly greater in those with moderate anxiety (Table 1). Additionally, no difference in any physiological measures were observed (Supplementary Table 3), nor in relative head motion during the task (average root mean square displacement (±std) for low anxiety=0.17±0.10 mm; moderate anxiety=0.18±0.07 mm; Wilcoxon rank sum p=0.91).

Computational modelling of brain activity: The overall and between-group BLT brain activity analysis results are displayed in Figures 4 and 5. In the analysis for the entire field of view, activations related to interoceptive prediction certainty and prediction errors across all participants is shown in Figure 4. Dorsolateral prefrontal cortex (dlPFC), anterior insula (aIns), anterior cingulate cortex (ACC) and middle frontal gyrus (MFG) demonstrated significant deactivations with overall prediction certainty (i.e. averaged across trials with positive and negative prediction certainty; Figure 4A). In contrast, aIns, ACC, MFG and PAG demonstrated significant activations with overall prediction error values (i.e. averaged across trials with positive and negative prediction errors; Figure 4B). A small number of differences due to valence (differences between positive and negative outcomes) were found for prediction errors but not prediction certainty, with negative prediction errors associated with deactivations of left dlPFC and activations of left posterior insula (Figure 4B). While no main effect of anxiety group was observed, an interaction effect was found using the ROI analysis between valence and groups for predictions in the bilateral aIns (Figure 5). In contrast, no group or interaction effects were found for prediction errors. Brain activity associated with inspiratory resistance is provided in the Supplementary Material for comparison with previously published results[15–21].

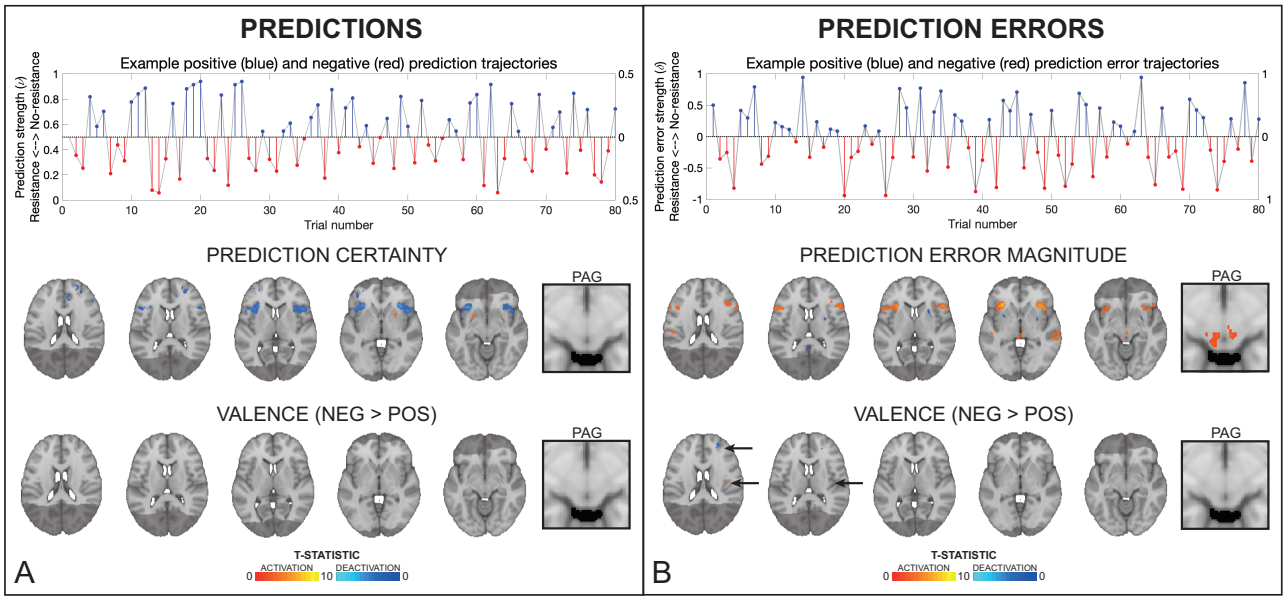


Figure 4. The upper plots demonstrate how estimated prediction (in A) and prediction error (in B) trajectories are encoded as positive (i.e. towards no resistance) and negative (i.e. towards resistance) prediction certainty values and prediction error magnitudes. The example trajectories were taken from the participant with the closest learning rate to the mean value across all participants. The solid grey lines demonstrate the estimated prediction or prediction error traces (in stimulus space). Positive trial values are demonstrated in blue and the negative trial values in red, encoded as distance from 0 (i.e. absolute values; right axes). The brain images represent significant activity across both groups for prediction certainty (averaged over trials with positive and negative prediction certainty) and the influence of valence on prediction certainty (difference between negative and positive predictions), prediction error magnitude (averaged over trials with positive and negative prediction errors) and the influence of valence on prediction error magnitude (difference between negative and positive prediction errors). The images consist of a colour-rendered statistical map superimposed on a standard (MNI 1x1x1mm) brain. The bright grey region represents the coverage of the coronal-oblique functional scan. Significant regions are displayed with a cluster threshold of

p<0.05, FWE corrected for multiple comparisons across all voxels included in the functional volume. Abbreviations: PAG, periaqueductal gray.

*Table 1. Behavioural and model-based group comparison results from the 'Breathing Learning Task' (BLT). All parameters are presented as median±inter-quartile range, and include the model parameter estimates (learning rate, α; inverse decision temperature ζ), the subjective ratings of breathing difficulty (average of the ratings provided following each resistance stimulus) and anxiety (rating provided immediately following the end of the task), and the response times for the predictions made during the task. Abbreviations: Wxn, Wilcoxon rank sum test. **Significant difference between groups at p<0.05 with multiple comparison correction for the number of behavioural parameters.*

	Total	Low	Moderate	P-value	Test
Learning rate (α)	0.25 (0.19)	0.24 (0.15)	0.25 (0.22)	0.58	Wxn
Inverse decision temp. (ζ)	2.66 (3.35)	2.71 (3.15)	2.37 (3.65)	0.88	Wxn
Breathing difficulty rating (%)	82.6 (18.4)	80.5 (19.9)	83.8 (15.8)	0.61	Wxn
Breathing anxiety rating (%)	10.0 (42.0)	**0.0 (10.0)**	**34.0 (48.0)**	**< 0.001****	Wxn
Response time (ms)	1.28 (0.33)	1.23 (0.36)	1.29 (0.30)	0.73	Wxn

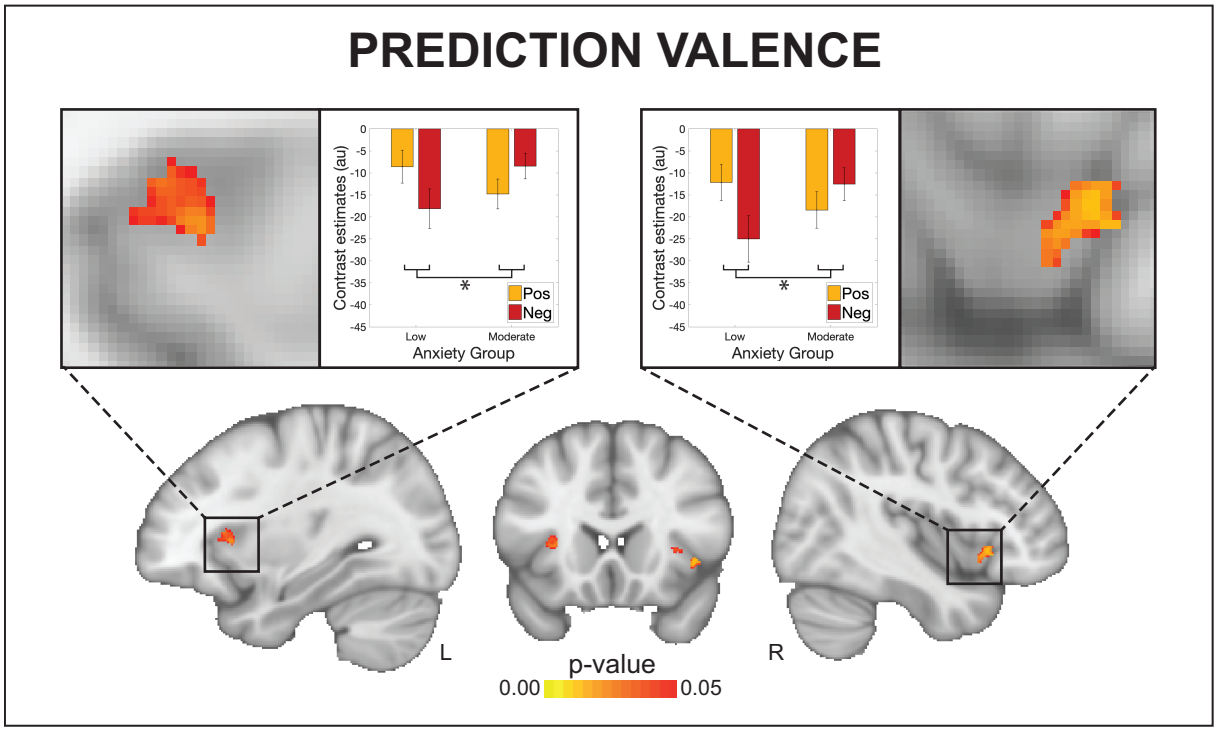


Figure 5. An interaction effect was observed between valence (i.e. trials with positive vs. negative predictions) and anxiety group (low vs. moderate) for activity in the anterior insula related to prediction certainty. The images consist of a colour-rendered statistical map superimposed on a standard (MNI 1x1x1mm) brain. Voxel-wise statistics were performed using non-parametric permutation testing within a mask of the anterior insula and periaqueductal gray, with significant results determined by p<0.05 (corrected for multiple comparisons within the mask).

Multi-modal analysis results

First, the key measures from each of the different modalities were combined into a multi-modal correlation matrix. This analysis allowed us to assess the relationships both within and across task modalities and across levels of breathing-related interoceptive processing. The full correlation matrix of all 16 included measures is displayed in Figure 6A and Supplementary Table 7. To briefly summarise, the strongest across-task modality correlations were found between all affective and interoceptive questionnaires (Figure 6A). Concerning affective and the FDT measures, state anxiety was weakly correlated with the FDT perceptual threshold, decision bias and metacognitive

bias, while anxiety sensitivity was additionally weakly related to metacognitive bias. Depression scores were also weakly related to the FDT perceptual threshold. Between the interoceptive and the FDT measures, breathing-related catastrophising was weakly related to the FDT metacognitive performance. Lastly, between the FDT and aIns activity, metacognitive performance was strongly related to the peak aIns associated with negative (i.e. resistance-related) prediction errors, while metacognitive bias was weakly related to aIns associated with negative (i.e. resistance-related) prediction certainty.

Principal component analysis: Finally, to assess the extent of shared variance across interoceptive measures, the multimodal data matrix was then subjected to a PCA. This analysis allowed us to delineate how many underlying dimensions may exist within the data, as well as which measures were most strongly associated with trait anxiety. Two principal components (PC) were found to be significant, where the variance explained with each component was above the 95% confidence interval of its null distribution (Supplementary Figure 14). The properties of each of these significant components are displayed in Figure 6B and 6C. The first PC demonstrated a highly significant ($p<1x10^{-11}$) difference in scores between the anxiety groups. Correspondingly, the greatest coefficients within the first PC were from the affective measures of depression scores, state anxiety, anxiety sensitivity and anxiety disorder scores. Additionally, breathing-related catastrophising and negative interoceptive awareness also had strong coefficient values, followed by negative metacognitive bias (i.e. lower confidence scores), body perception scores (from the BPQ) and negative metacognitive performance (i.e. lower metacognitive performance). In contrast, the second PC demonstrated a weak difference ($p=0.05$) in scores between the anxiety groups. This component had the highest coefficient scores from the peak aIns activity related to positive and negative prediction certainty, as well as negative coefficients for negative prediction errors, metacognitive performance and positive prediction errors.

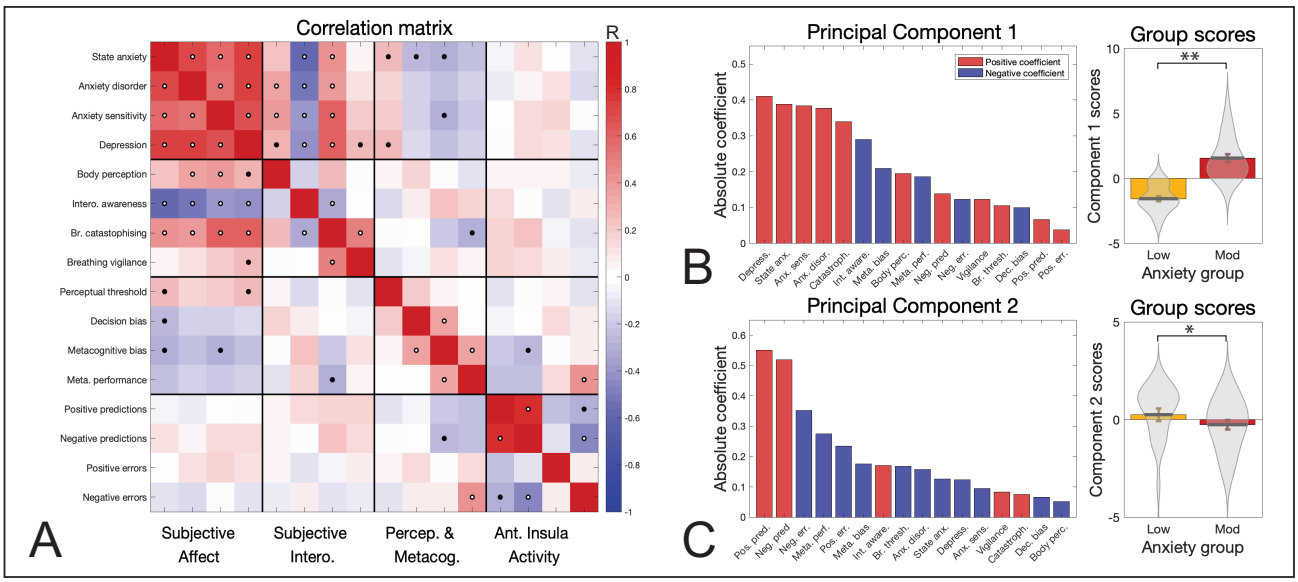


*Figure 6. A) Correlation matrix results for the 16 included measures in the multi-modal analysis. Black dots represent significant values at $p<0.05$, while white dots denote significance with correction for multiple comparisons. B) The weights and group scores of the first significant principal component, where a strong anxiety group difference in component scores is observed. C) The weights and group scores of the second significant principal component, where a weak anxiety group difference in principal component scores is observed. *Significant difference between groups at $p<0.05$. **Significant difference between groups at $p<0.05$ with multiple comparison correction for the two significant components. Bar plots (rightmost panels) represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*

Discussion

Main findings

Interoceptive abilities are thought to be tightly linked to affective properties such as anxiety. Here we have characterised this relationship across multiple domains, including novel findings of altered brain activity within the aIns when processing dynamic interoceptive predictions. Furthermore, our multi-modal approach revealed that not only is the relationship between interoception and trait anxiety broad, it is largest (i.e. greatest PCA weights: Figure 6) at the higher levels of interoceptive processing, which includes specific subjective measures of interoceptive beliefs (often termed 'interoceptive sensibility'[43]) followed by metacognitive aspects of breathing perceptions. Notably, interoceptive sensibility differences with greater levels of anxiety included reduced 'positively-minded' interoceptive awareness, greater reports of overall body awareness and elevated levels of breathing-related catastrophising (as measured using interoceptive questionnaires). Additionally, those with greater trait anxiety demonstrated differences in metacognitive bias (i.e. reduced confidence in perceptual decisions), as well as decreased metacognitive performance.

This study is one of the first to simultaneously tackle multiple levels of interoceptive processing, using breathing as a salient and accessible channel of body perception. The tasks employed reflected the broad range of targeted processes; not only were questionnaires employed that spanned affect and body perceptions, but behavioural data from two different tasks were assessed by separate computational models. These analyses allowed for formal assessments of both interoceptive learning and metacognition, including the first computational assessment of trial-by-trial learning in the interoceptive domain as well as applying state-of-the-art models of metacognition to interoception of breathing. Finally, the learning task was paired with high-field (7 Tesla) to maximise signal-to-noise ratio for a detailed investigation of key brain areas thought to be vital for dynamic interoceptive processing.

Affect and levels of interoception

Beyond consequences at single levels of interoceptive processing, here we aimed to assess how the relationship with trait anxiety may cross multiple interoceptive levels related to breathing. Employing PCA (with permutation testing) allowed us to identify any components that share common variance within our multi-modal dataset, and additionally assess the relative contribution of our measures to each dimension (Figure 6B). Here we found all affective qualities loaded strongly onto the first principal component, with the greatest additional contributions from subjective measures of negative interoceptive awareness and breathing-related catastrophising. General body awareness and the metacognitive measures (bias and performance) were the next largest contributors to this shared variance, followed by the perceptual sensitivity and decision bias parameters, and lastly peak aIns activity from the BLT. These results suggest that the relationship with anxiety is potentially largest at the level of subjective interoceptive beliefs, thought to exist in the higher levels of interoceptive space[35], followed by metacognitive insight into breathing perception. In comparison, the relationship of trait anxiety to lower-level properties such as interoceptive sensitivity[35,43–45] appear to be present but less prominent in the breathing domain.

Although strong relationships were observed between affective qualities and many of our interoceptive measures, a sparse number of correlations were found between interoceptive measures themselves, and in particular across task modalities (Figure 6A). These findings support the idea that there are potentially separable levels of interoception as proposed[35]. The only notably strong cross-modal relationship was found between metacognitive performance and aIns activity, where greater insight into interoceptive sensitivity correlated with greater aIns activity during negative interoceptive prediction errors. This relationship may reflect a previously proposed contribution of error-processing towards metacognitive awareness, where deviations between actual and predicted bodily inputs are propagated to metacognitive areas via interoceptive brain structures such as the aIns[6].

Novel measures of dynamic interoceptive predictions and prediction errors

Many theories surrounding anxiety have hypothesised an important role of altered predictions regarding upcoming threat[97–99], and in particular interoceptive threat[13,80] in the anterior insula[100–104]. While numerous studies have employed inspiratory resistive loads to evoke threatening interoceptive stimuli[15,16,18–21,93,105–108], the BLT approach

presented here is, to the authors' knowledge, the first investigation of dynamic (trial-by-trial) brain activity associated with model-based interoceptive predictions and prediction errors in the respiratory domain. By fitting an associative learning model to each participant's behavioural responses, we could quantify both the certainty of predictions and magnitude of prediction errors on each trial. We could then compare both the parameter estimates and the brain activity associated with these computational quantities, with a particular focus on the aIns and PAG[94,95,100,104,109] (Figure 4). Here we observed evidence for a relationship between anxiety and aIns reactivity to threat valence in the prediction domain (Figure 5). Specifically, while individuals with low trait anxiety demonstrated greater aIns deactivation that scaled with predictions of breathing resistance compared to no resistance, the opposite was true in individuals with moderate trait anxiety (creating an interaction effect). This demonstrates a shift in the aIns processing of threat valence with different levels of anxiety, in line with hypothesised differences in brain prediction processing[13,80,100]. In comparison, no anxiety group differences or interactions were found in the prediction error domain.

Beyond the anterior insula and independent of anxiety, multiple (and largely consistent) proposals have been made, inspired by predictive coding and related theories of brain function, regarding which brain networks might be involved in processing interoceptive predictions and prediction errors[2,3,5,6,11,26,80,104,110–117]. These proposed networks are loosely hierarchical in structure and typically assign metacognitive processes to higher cortical areas (e.g. in prefrontal cortex; PFC) while interoceptive predictions are thought to originate from regions that may serve as interface between interoceptive and visceromotor function (e.g. aIns and anterior cingulate cortex; ACC). In these concepts, prediction errors have two different roles: on the one hand, they are thought to be sent up the cortical hierarchy of interoceptive regions in order to update predictions in aIns and ACC[3,5,51]; on the other hand, they are thought to determine regulatory signals, sent from visceromotor regions and brainstem structures like the PAG to the autonomic nervous system and bodily organs[6,40].

While these theories have received considerable attention, there has been little empirical evidence thus far. In particular, we are not aware of any studies that have demonstrated, using a concrete computational model, trial-by-trial prediction and prediction error activity in interoceptive areas. Here, we report evidence of relevant computational quantities being reflected by activity within several areas of a putative interoceptive network. While activity related to trial-wise prediction certainty was found in higher structures such as dorsolateral PFC, ACC and aIns, prominent prediction error responses were not only found in aIns and ACC, but also in the midbrain PAG (Figure 4). Importantly, concerning predictions, widespread brain activity was found to be mainly related to prediction *un*certainty, where BOLD activity was decreased in proportion to increases in the certainty of predictions[47,48]. Furthermore, it is notable that the anterior insula displayed both deactivation for more certain predictions and activation for greater prediction errors. This might reflect the proposed critical role of aIns in the representation and updating of models of bodily states[2,6,80,93,100,104,110,118], given that greater certainty (precision of beliefs) reduces and greater prediction errors increase belief (model) updating[40].

Our PAG findings are of particular interest. While the PAG has been previously noted during anticipation of certain breathing resistance stimuli[20,21] and has been related to the precision of prior beliefs about placebo-induced reductions in pain intensity[95], here we observed that PAG activity did not appear to be related to the extent of prediction certainty towards upcoming breathing stimuli (Figure 4). Concerning prediction error activity in the PAG, this has previously been demonstrated in relation to pain[94]; here, we found PAG activity in relation to the magnitudes of trial-wise interoceptive prediction errors (Figure 4B), consistent with a role of PAG in homeostatic control[6].

Conclusions

The relationship between anxiety and breathing crosses multiple levels of the interoceptive hierarchy. In particular, anxiety and associated affective dimensions appear to be most strongly related to subjective negative body awareness and catastrophising about breathing symptoms, followed by metacognitive measures related to breathing perception. Furthermore, a novel interaction between trait anxiety group and valence was found within the aIns associated with dynamic prediction certainty (but not prediction errors) of interoceptive processing. These results provide new and comprehensive insights how anxiety is related to levels of interoceptive processing, and provide first evidence of brain activity associated with trial-wise predictions and prediction errors about bodily states in interoceptive networks.

Acknowledgements

The authors would like to thank Professor Klaas Pruessmann and Dr Lars Kasper for their work establishing and supporting the MRI protocol. OKH (née Faull) was supported by a Marie Skłodowska-Curie Postdoctoral Fellowship from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 793580. SF was supported by the UZH Forschungskredit Postdoc FK-18-046, as well as the ETH Zurich Postdoctoral Fellowship Program and the Marie Curie Actions for People COFUND program FEL-49 15-2. FV was supported by the Fondation Deniker, the Fondation pour la Recherche Médicale, and the Fondation Bettencourt Schueller. SJH was supported by the grant #2017-403 of the Strategic Focus Area "Personalized Health and Related Technologies (PHRT)" of the ETH Domain. KES was supported by the René and Susanne Braginsky Foundation and the University of Zurich.

Conflict of interest statement

FV been invited to scientific meetings, consulted and/or served as speaker and received compensation by Lundbeck, Servier, Recordati, Janssen, Otsuka, LivaNova and Chiesi. None of these links of interest are related to this work. The authors have no other conflicts of interest to declare.

References

1. Craig, A. D. How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience* **3**, 655–666 (2002).
2. Seth, A. K. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences* **17**, 565–573 (2013).
3. Barrett, L. F. & Simmons, W. K. Interoceptive predictions in the brain. *Nature Reviews Neuroscience* **16**, 419–429 (2015).
4. Tsakiris, M. & Critchley, H. Interoception beyond homeostasis: affect, cognition and mental health. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20160002–6 (2016).
5. Pezzulo, G., Rigoli, F. & Friston, K. Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in neurobiology* **134**, 17–35 (2015).
6. Stephan, K. E. *et al.* Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression. *Frontiers in human neuroscience* **10**, 49–27 (2016).
7. Berntson, G. G. & Khalsa, S. S. Neural Circuits of Interoception. *Trends Neurosci* **44**, 17–28 (2021).
8. Chen, W. G. *et al.* The Emerging Science of Interoception: Sensing, Integrating, Interpreting, and Regulating Signals within the Self. *Trends Neurosci* **44**, 3–16 (2021).
9. Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F. & Tsakiris, M. Functions of Interoception: From Energy Regulation to Experience of the Self. *Trends Neurosci* **44**, 29–38 (2021).
10. Adolfi, F. *et al.* Convergence of interoception, emotion, and social cognition: A twofold fMRI meta-analysis and lesion approach. *Cortex* **88**, 124–142 (2017).
11. Khalsa, S. S. *et al.* Interoception and Mental Health: a Roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1–57 (2017) doi:10.1016/j.bpsc.2017.12.004.
12. Bonaz, B. *et al.* Diseases, Disorders, and Comorbidities of Interoception. *Trends Neurosci* **44**, 39–51 (2021).
13. Paulus, M. P. & Stein, M. B. Interoception in anxiety and depression. *Brain Structure and Function* **214**, 1–13 (2010).
14. Paulus, M. P. The breathing conundrum - Interoceptive sensitivity and anxiety. *Depression and Anxiety* **30**, 315–320 (2013).
15. Berner, L. A. *et al.* Altered interoceptive activation before, during, and after aversive breathing load in women remitted from anorexia nervosa. *Psychological Medicine* **48**, 142–154 (2017).
16. Paulus, M. P. *et al.* Subjecting Elite Athletes to Inspiratory Breathing Load Reveals Behavioral and Neural Signatures of Optimal Performers in Extreme Environments. *PLOS ONE* **7**, e29394 (2012).

17. DeVille, D. C. *et al.* The Neural Bases of Interoceptive Encoding and Recall in Healthy and Depressed Adults. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1–42 (2018) doi:10.1016/j.bpsc.2018.03.010.
18. Faull, O. K., Cox, P. J. & Pattinson, K. T. S. Cortical processing of breathing perceptions in the athletic brain. *Neuroimage* **179**, 92–101 (2018).
19. Hayen, A. *et al.* Opioid suppression of conditioned anticipatory brain responses to breathlessness. *Neuroimage* **150**, 383–394 (2017).
20. Faull, O. K. & Pattinson, K. T. The cortical connectivity of the periaqueductal gray and the conditioned response to the threat of breathlessness. *eLife* **6**, (2017).
21. Faull, O. K., Jenkinson, M., Ezra, M. & Pattinson, K. T. S. Conditioned respiratory threat in the subdivisions of the human periaqueductal gray. *eLife* **5**, (2016).
22. Rieger, S. W., Stephan, K. E. & Harrison, O. K. Remote, Automated, and MRI-Compatible Administration of Interoceptive Inspiratory Resistive Loading. *Frontiers in human neuroscience* **14**, (2020).
23. Schwartzstein, R. M., Manning, H. L., Weiss, J. W. & Weinberger, S. E. Dyspnea - a Sensory Experience. *Lung* **168**, 185–199 (1990).
24. Herigstad, M., Hayen, A., Wiech, K. & Pattinson, K. T. S. Dyspnoea and the brain. *Respiratory medicine* **105**, 809–817 (2011).
25. Hayen, A., Herigstad, M. & Pattinson, K. T. S. Understanding dyspnea as a complex individual experience. *Maturitas* **76**, 45–50 (2013).
26. Marlow, L. L., Faull, O. K., Finnegan, S. L. & Pattinson, K. T. S. Breathlessness and the brain: the role of expectation. *Current Opinion in Supportive and Palliative Care* **13**, 200–210 (2019).
27. Parshall, M. B. *et al.* An Official American Thoracic Society Statement: Update on the Mechanisms, Assessment, and Management of Dyspnea. *American journal of respiratory and critical care medicine* **185**, 435–452 (2012).
28. Janssens, T. *et al.* Dyspnea perception in COPD: association between anxiety, dyspnea-related fear, and dyspnea in a pulmonary rehabilitation program. *CHEST Journal* **140**, 618–625 (2011).
29. Carrieri-Kohlman, V. *et al.* Additional evidence for the affective dimension of dyspnea in patients with COPD. *Research in Nursing & Health* **33**, n/a-n/a (2009).
30. Giardino, N. D. *et al.* The impact of panic disorder on interoception and dyspnea reports in chronic obstructive pulmonary disease. *Biological Psychology* **84**, 142–146 (2010).
31. Mallorqui-Bague, N., Bulbena, A., Pailhez, G., Garfinkel, S. N. & Critchley, H. D. Mind-Body Interactions in Anxiety and Somatic Symptoms. *Harvard Review of Psychiatry* **24**, 53–60 (2016).
32. McNally, R. J. & Eke, M. Anxiety sensitivity, suffocation fear, and breath-holding duration as predictors of response to carbon dioxide challenge. *Journal of abnormal psychology* **105**, 146–149 (1996).
33. Woods, S. W. *et al.* Carbon Dioxide Sensitivity in Panic Anxiety: Ventilatory and Anxiogenic Response to Carbon Dioxide in Healthy Subjects and Patients With Panic Anxiety Before and After Alprazolam Treatment. *Archives of general psychiatry* **43**, 900–909 (1986).
34. Smoller, J. W., Pollack, M. H., Otto, M. W., Rosenbaum, J. F. & Kradin, R. L. Panic anxiety, dyspnea, and respiratory disease. Theoretical and clinical considerations. *Am J Resp Crit Care* **154**, 6–17 (1996).
35. Critchley, H. D. & Garfinkel, S. N. Interoception and emotion. *Current Opinion in Psychology* **17**, 7–14 (2017).
36. Petzschner, F. H. *et al.* Focus of attention modulates the heartbeat evoked potential. *Neuroimage* **186**, 595–606 (2019).
37. Allen, M., Frank, D., Schwarzkopf, D. S. & Fardo, F. Unexpected arousal modulates the influence of sensory noise on confidence. *eLife* (2016) doi:10.7554/elife.18103.001.
38. Domschke, K., Stevens, S., Pfleiderer, B. & Gerlach, A. L. Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical psychology review* **30**, 1–11 (2010).
39. Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F. & Quigley, K. S. Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology* **52**, 1432–1440 (2015).
40. Petzschner, F. H., Weber, L. A. E., Gard, T. & Stephan, K. E. Computational Psychosomatics and Computational Psychiatry: Toward a Joint Framework for Differential Diagnosis. *Biological Psychiatry* 1–10 (2017) doi:10.1016/j.biopsych.2017.05.012.
41. Wang, X. *et al.* Anterior insular cortex plays a critical role in interoceptive attention. *ELife* (2019) doi:10.7554/elife.42265.001.
42. Murphy, J., Catmur, C. & Bird, G. Classifying individual differences in interoception: Implications for the measurement of interoceptive awareness. *null* 1–5 (2019) doi:10.3758/s13423-019-01632-7.
43. Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K. & Critchley, H. D. Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology* **104**, 65–74 (2015).

44. Garfinkel, S. N. *et al.* Interoceptive dimensions across cardiac and respiratory axes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **371**, 20160014–10 (2016).
45. Garfinkel, S. N. *et al.* Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychology* **114**, 117–126 (2016).
46. Harrison, O. K. *et al.* The Filter Detection Task for measurement of breathing-related interoception and metacognition. *bioRxiv* 2020.06.29.176941 (2020) doi:10.1101/2020.06.29.176941.
47. Friston, K. A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 815–836 (2005).
48. Feldman, H. & Friston, K. J. Attention, Uncertainty, and Free-Energy. *Frontiers in human neuroscience* **4**, 1–23 (2010).
49. O'Reilly, J. X. & Jbabdi, S. How can a Bayesian approach inform neuroscience? *European Journal of ...* **35**, 1169–1179 (2012).
50. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience* **10**, 1214–1221 (2007).
51. Seth, A. K., Suzuki, K. & Critchley, H. D. An Interoceptive Predictive Coding Model of Conscious Presence. *Front Psychol* **2**, 395 (2012).
52. Gu, X., Hof, P. R., Friston, K. J. & Fan, J. Anterior insular cortex and emotional awareness. *J Comp Neurol* **521**, 3371–3388 (2013).
53. Tiller, J., Pain, M. & Biddle, N. Anxiety Disorder and Perception of Inspiratory Resistive Loads. *Chest* **91**, 547–551 (1987).
54. Mehling, W. Differentiating attention styles and regulatory aspects of self-reported interoceptive sensibility. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20160013–11 (2016).
55. Ewing, D. L. *et al.* Sleep and the heart: Interoceptive differences linked to poor experiential sleep quality in anxiety and depression. *Biological Psychology* **127**, 163–172 (2017).
56. Harrison, O. K., Marlow, L. L., Finnegan, S. L., Ainsworth, B. & Pattinson, K. T. S. Heterogeneity in asthma: Dissociating symptoms from mood and their influence on interoception and attention. *BioRXiv* (2020).
57. Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. *State-trait anxiety (STAI) manual*. (Palo Alto, 1970).
58. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* **166**, 1092–1097 (2006).
59. Taylor, S. *et al.* Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment* **19**, 176–188 (2007).
60. Radloff, L. S. The CES-D scale a self-report depression scale for research in the general population. *Applied psychological measurement* **1**, 385–401 (1977).
61. Porges, S. W. Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A Polyvagal Theory. *Psychophysiology* **32**, 301–318 (1995).
62. Mehling, W. E. *et al.* The Multidimensional Assessment of Interoceptive Awareness (MAIA). *PLOS ONE* **7**, e48230-22 (2012).
63. Sullivan, M. J. L., Bishop, S. R. & Pivik, J. The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment* **7**, 524–532 (1995).
64. McCracken, L. M. "Attention" to pain in persons with chronic pain: A behavioral approach. *Behavior Therapy* **28**, 271–284 (1997).
65. Watson, D., Clark, L. A. & Tellegen, A. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of personality and social psychology* **54**, 1063–1070 (1988).
66. Connor, K. M. & Davidson, J. R. T. Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety* **18**, 76–82 (2003).
67. Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K. & Zhang, J. X. The Assessment of Optimistic Self-beliefs: Comparison of the German, Spanish, and Chinese Versions of the General Self-efficacy Scale. *Applied Psychology* **46**, 69–88 (1997).
68. Krupp, L. B., LaRocca, N. G., Muir-Nash, J. & Steinberg, A. D. The Fatigue Severity Scale: Application to Patients With Multiple Sclerosis and Systemic Lupus Erythematosus. *Archives of Neurology* **46**, 1121–1123 (1989).
69. Rouault, M., Seow, T., Gillan, C. M. & Fleming, S. M. Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry* 1–9 (2018) doi:10.1016/j.biopsych.2017.12.017.
70. Fleming, S. M. HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness* 377–14 (2017) doi:10.1093/nc/nix007.

71. Stanislaw, H. & Todorov, N. Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* **31**, 137–149 (1999).
72. Mathys, C. D. *et al.* Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in human neuroscience* **8**, 825 (2014).
73. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience* **5**, 39 (2011).
74. Frässle, S. *et al.* TAPAS: an open-source software package for Translational Neuromodeling and Computational Psychiatry. *undefined* doi:10.1101/2021.03.12.435091.
75. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).
76. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies — Revisited. *Neuroimage* **84**, 971–985 (2014).
77. Rescorla, R. A., Wagner, A. R., Black, A. H. & Prokasy, W. F. *Classical conditioning II: current research and theory*. (1972).
78. Iglesias, S. *et al.* Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron* **80**, 519–530 (2013).
79. Daw, N. D. Trial-by-trial data analysis using computational models. in *undefined* 3–38 (2011). doi:10.1093/acprof:oso/9780199600434.003.0001.
80. Paulus, M. P., Feinstein, J. S. & Khalsa, S. S. An Active Inference Approach to Interoceptive Psychopathology. *Annual Review of Clinical Psychology* **15**, 97–122 (2019).
81. Faull, O. K., Jenkinson, M., Clare, S. & Pattinson, K. T. S. Functional subdivision of the human periaqueductal grey in respiratory control using 7 tesla fMRI. *Neuroimage* **113**, 356–364 (2015).
82. Chang, C. & Glover, G. H. Relationship between respiration, end-tidal CO2, and BOLD signals in resting-state fMRI. *Neuroimage* **47**, 1381–1393 (2009).
83. Kasper, L. *et al.* The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods* **276**, 56–72 (2017).
84. Harrison, S. J. *et al.* A Hilbert-based method for processing respiratory timeseries. *Neuroimage* **230**, 117787 (2021).
85. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
86. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* **5**, 143–156 (2001).
87. Smith, S. M. Fast robust automated brain extraction. *Human brain mapping* **17**, 143–155 (2002).
88. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *Neuroimage* **14**, 1370–1386 (2001).
89. Griffanti, L. *et al.* Hand classification of fMRI ICA noise components. *Neuroimage* 1–18 (2017) doi:10.1016/j.neuroimage.2016.12.036.
90. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage* **17**, 825–841 (2002).
91. Andersson, J. L., Jenkinson, M. & Smith, S. Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. *FMRIB Analysis Group of the University of Oxford* (2007).
92. Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
93. Walter, H. *et al.* Self-control and interoception: Linking the neural substrates of craving regulation and the prediction of aversive interoceptive states induced by inspiratory breathing restriction. *Neuroimage* **215**, 116841 (2020).
94. Roy, M. *et al.* Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience* **17**, 1607–1612 (2014).
95. Grahl, A., Onat, S. & Büchel, C. The periaqueductal gray and Bayesian integration in placebo analgesia. *eLife* **7**, 1–20 (2018).
96. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex* **26**, 3508–3526 (2016).
97. Simmons, A., Strigo, I., Matthews, S. C., Paulus, M. P. & Stein, M. B. Anticipation of Aversive Visual Stimuli Is Associated With Increased Insula Activation in Anxiety-Prone Subjects. *Biological Psychiatry* **60**, 402–409 (2006).
98. Mogg, K. *et al.* Selective attention to threat: A test of two cognitive models of anxiety. *Cognition & Emotion* **14**, 375–399 (2000).

99. Bach, D. R. Anxiety-Like Behavioural Inhibition Is Normative under Environmental Threat-Reward Correlations. *Plos Comput Biol* **11**, e1004646 (2015).
100. Paulus, M. P. & Stein, M. B. An Insular View of Anxiety. *Biological Psychiatry* **60**, 383–387 (2006).
101. Tan, Y. *et al.* The role of mid-insula in the relationship between cardiac interoceptive attention and anxiety: evidence from an fMRI study. *Scientific Reports* 1–12 (2018) doi:10.1038/s41598-018-35635-6.
102. Bossaerts, P. Risk and risk prediction error signals in anterior insula. *Brain Structure and Function* **214**, 645–653 (2010).
103. Carlson, J. M., Greenberg, T., Rubin, D. & Mujica-Parodi, L. R. Feeling anxious: anticipatory amygdalo-insular response predicts the feeling of anxious anticipation. *Social Cognitive and Affective Neuroscience* **6**, 74–81 (2010).
104. Allen, M. Unravelling the Neurobiology of Interoceptive Inference. *Trends in Cognitive Sciences* **24**, 265–266 (2020).
105. Leupoldt, A. von & Dahme, B. Differentiation between the sensory and affective dimension of dyspnea during resistive load breathing in normal subjects. *Chest* **128**, 3345–3349 (2005).
106. Leupoldt, A. von *et al.* Down-Regulation of Insular Cortex Responses to Dyspnea and Pain in Asthma. *American journal of respiratory and critical care medicine* **180**, 232–238 (2009).
107. Alius, M. G., Pané-Farré, C. A., Leupoldt, A. von & Hamm, A. O. Induction of dyspnea evokes increased anxiety and maladaptive breathing in individuals with high anxiety sensitivity and suffocation fear. *Psychophysiology* **50**, 488–497 (2013).
108. Stoeckel, M. C., Esser, R. W., Gamer, M., Büchel, C. & Leupoldt, A. von. Brain Responses during the Anticipation of Dyspnea. *null* **2016**, 1–10 (2016).
109. Singer, T., Critchley, H. D. & Preuschoff, K. A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences* **13**, 334–340 (2009).
110. Manjaly, Z.-M. & Iglesias, S. A Computational Theory of Mindfulness Based Cognitive Therapy from the "Bayesian Brain" Perspective. *Frontiers in Psychiatry* **11**, 25 (2020).
111. Kleckner, I. R. *et al.* Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour* **1**, 0069–15 (2017).
112. Owens, A. P., Allen, M., Ondobaka, S. & Friston, K. J. Interoceptive inference: From computational neuroscience to clinic. *Neuroscience & Biobehavioral Reviews* **90**, 174–183 (2018).
113. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience* **218**, nsw154-23 (2016).
114. Pezzulo, G., Rigoli, F. & Friston, K. J. Hierarchical Active Inference: A Theory of Motivated Control. *Trends in Cognitive Sciences* 1–13 (2018) doi:10.1016/j.tics.2018.01.009.
115. Smith, R., Thayer, J. F., Khalsa, S. S. & Lane, R. D. The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews* **75**, 274–296 (2017).
116. Quadt, L., Critchley, H. D. & Garfinkel, S. N. The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences* **101**, 6333–17 (2018).
117. Craig, A. D. How do you feel - now? The anterior insula and human awareness. *Nature Reviews Neuroscience* (2009).
118. Bergh, O. V. den, Witthöft, M., Petersen, S. & Brown, R. J. Symptoms and the body: Taking the inferential leap. *Neuroscience & Biobehavioral Reviews* **74**, 185–203 (2017).

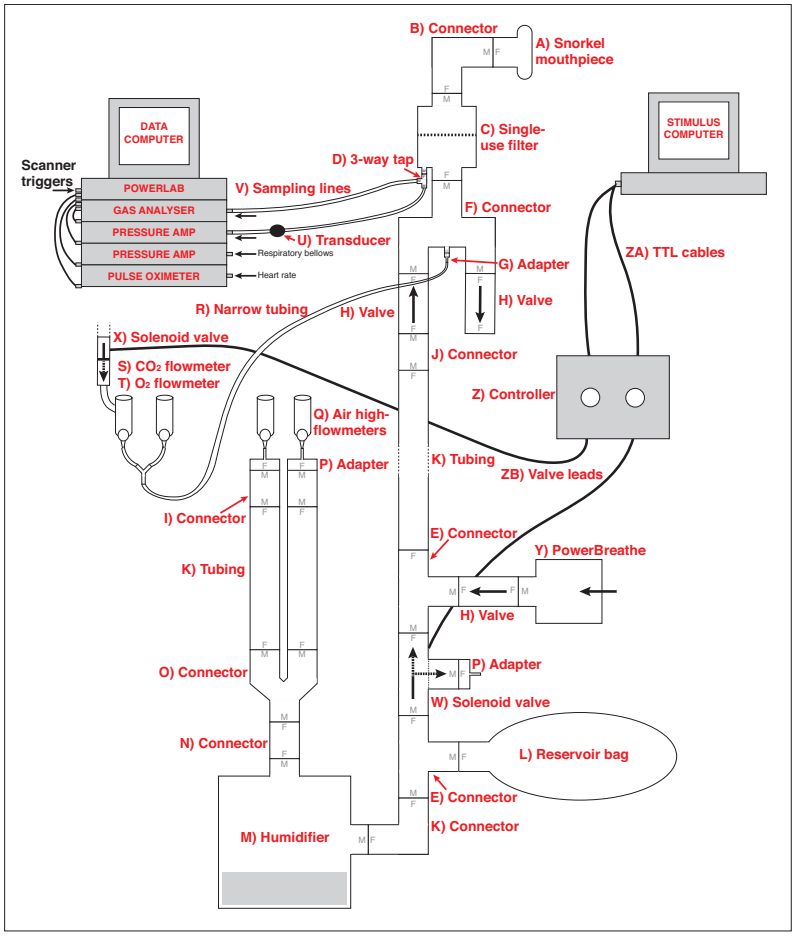
Supplementary Material: Interoception of breathing and its relationship with anxiety

Olivia K. Harrison[1,2,3*], Laura Nanz[1], Stephanie Marino[1], Roger Lüchinger[4], Franciszek Hennel[4], Alexander J. Hess[1], Stefan Frässle[1], Sandra Iglesias[1], Fabien Vinckier[1,5,6], Frederike Petzschner[1], Samuel J. Harrison[1,3], Klaas E. Stephan[1,7]

[1] Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Switzerland
[2] School of Pharmacy, University of Otago, New Zealand
[3] Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom
[4] Institute for Biomedical Engineering, University of Zürich and ETH Zürich, Switzerland
[5] Université de Paris, France
[6] Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, France
[7] Max Planck Institute for Metabolism Research, Germany

* Corresponding author:
Dr Olivia Harrison (née Faull)
Translational Neuromodeling Unit
Institute for Biomedical Engineering
University of Zurich and ETH Zurich

Supplementary Methods: Breathing Learning Task equipment



Supplementary Figure 1. Schematic of the inspiratory resistance circuit that allows remote administrations of inspiratory resistance. Medical air is supplied to the subject, with a reservoir of 2 L. Excess flow and expiration escapes through a one-way valve (labelled H), close to the mouth to minimise rebreathing. Chest movements are measured using respiratory bellows (non-metallic pneumographic belt; Lafayette Instrument Company, Lafayette, USA) connected to a pressure transducer (Blood Pressure Transducer; ADInstruments Ltd, Oxford, United Kingdom) and amplifier (Bridge Amp; ADInstruments Ltd). Heart rate is measured using a pulse oximeter (Philips Healthcare, Amsterdam, The Netherlands). A diving mouthpiece (labelled A) is connected to a bacterial and viral filter (labelled C), and sampling lines connect to a pressure transducer (labelled U) and amplifier (Pressure transducer indicator, PK Morgan Ltd, Kent, UK) for inspiratory pressure readings, and to a gas analyser (via sampling line labelled V) (Gas Analyser; ADInstruments Ltd, Oxford, United Kingdom) for respiratory gases. All physiological measurement devices were connected to a data acquisition device (Powerlab; ADInstruments Ltd) coupled to a computer with recording software (Labchart 7; ADInstruments Ltd). Inspiratory resistive loading is automatically achieved via the stimulus computer, whereby signals are sent through the parallel port to control valve 1 (labelled W) to redirect the supply of medical air to vent to the environment, forcing the subject to draw air through the POWERbreathe device (labelled Y). Periodically throughout scanning, small boluses of additional carbon dioxide (CO_2) can be administered through computer control via valve 2 (labelled X), to raise the partial pressure of end-tidal CO_2 ($P_{ET}CO_2$) to match the $P_{ET}CO_2$ rise induced by inspiratory loading periods. A final flowmeter (labelled T) is available for manual input of additional oxygen (O_2) to the system. Figure is adapted from Rieger et al. (2020).

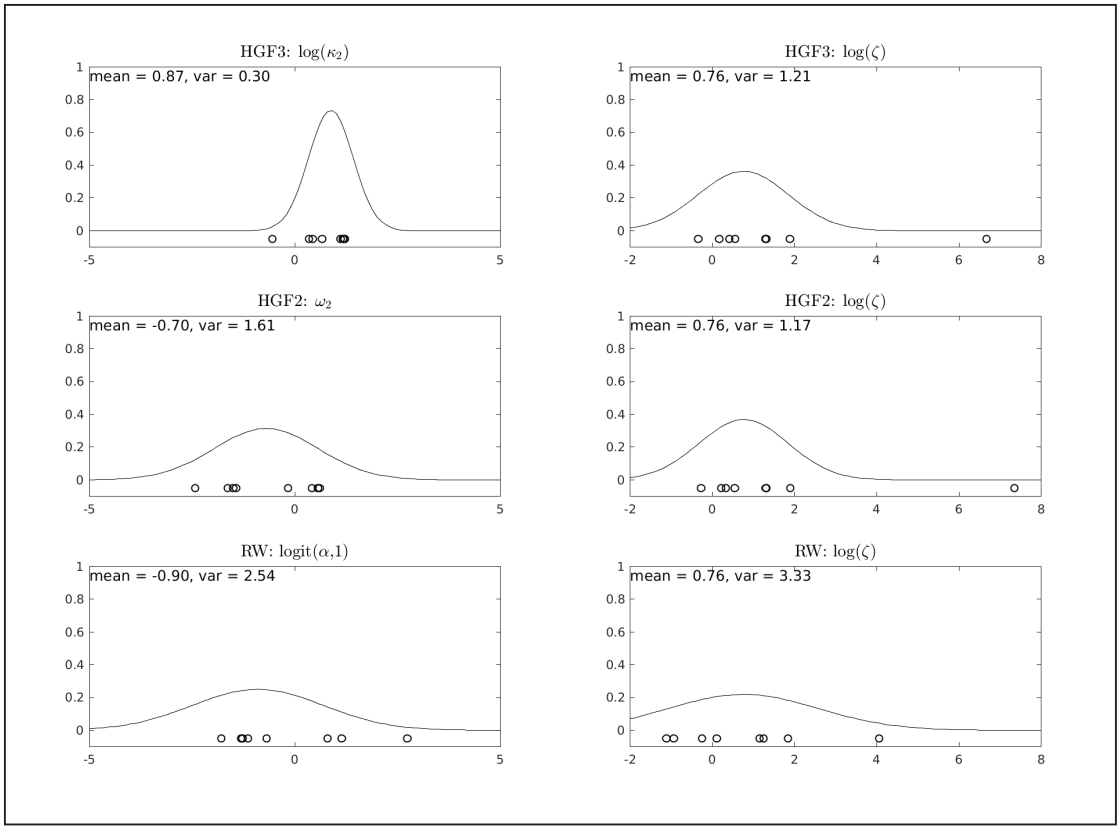
Supplementary Methods: Computational modelling of Breathing Learning Task

For the computational modelling of the BLT behavioural data, we a priori performed a set of simulation analyses as outlined by Wilson and Collins[1]. Each of the three perceptual models in our model space was used in a configuration where one parameter was kept free (α in RW, ω_2 in HGF2, κ_2 in HGF3). The perceptual model was in all three cases combined with a unit-square sigmoid response model, which had another free parameter ζ (inverse decision temperature) to be estimated. To analyse the experimental data set, maximising the likelihood on a holdout data set (consisting of 8 pilot participants) enabled us to estimate prior densities for each of the models in our model space. Individual fits and estimated prior densities of the free parameters are displayed in Supplementary Figure 2, and prior values for all parameters of the three models are listed in Supplementary Table 1. By adopting this procedure, prior densities were in a regime of the parameter space that is representative of the actual behavioural responses observed when participants performed the task. At the same time, the arbitrariness inherent to the specification of prior densities in non-hierarchical inference is reduced to a minimum.

To demonstrate face validity of the models considered in our model space, we assessed both parameter recovery and model identifiability for each of them. Data for 60 synthetic subjects were generated for each of the candidate models by randomly sampling values from the prior densities that were placed over the parameters of the perceptual model. This synthetic data was generated for different noise levels ($\zeta_{sim} = 1,5,10$). Subsequently, MAP estimation as implemented in the HGF Toolbox was used to fit the synthetic data sets. This allowed us to quantify parameter recovery and model identifiability across three different noise levels for each of the candidate models. Parameter recovery of the perceptual parameters was assessed using Pearson's correlation coefficient (PCC) and by visual inspection of simulated and recovered parameter values (Supplementary Figure 3). Mean and standard deviation of estimated ζ values (from the response model) were computed for every noise level. Model identifiability was quantified by calculating the proportion of correctly identified models using approximate LME scores and assessing whether the former was greater than the upper bound of the 90% confidence interval when assuming every model is equally likely a priori. For the resulting confusion matrices (Supplementary Figure 4), we additionally computed the mean proportion of correctly identified models (balanced accuracy scores).

The outlined procedure for assessing parameter recovery and model identifiability was repeated over 10 iterations with different seed values, to ensure robustness against any particular setting of the random number generator. The final results (PCCs for the perceptual parameters and ζ_{est} values) for every given level of noise were calculated as the average over all iterations, and are presented in Supplementary Table 2.

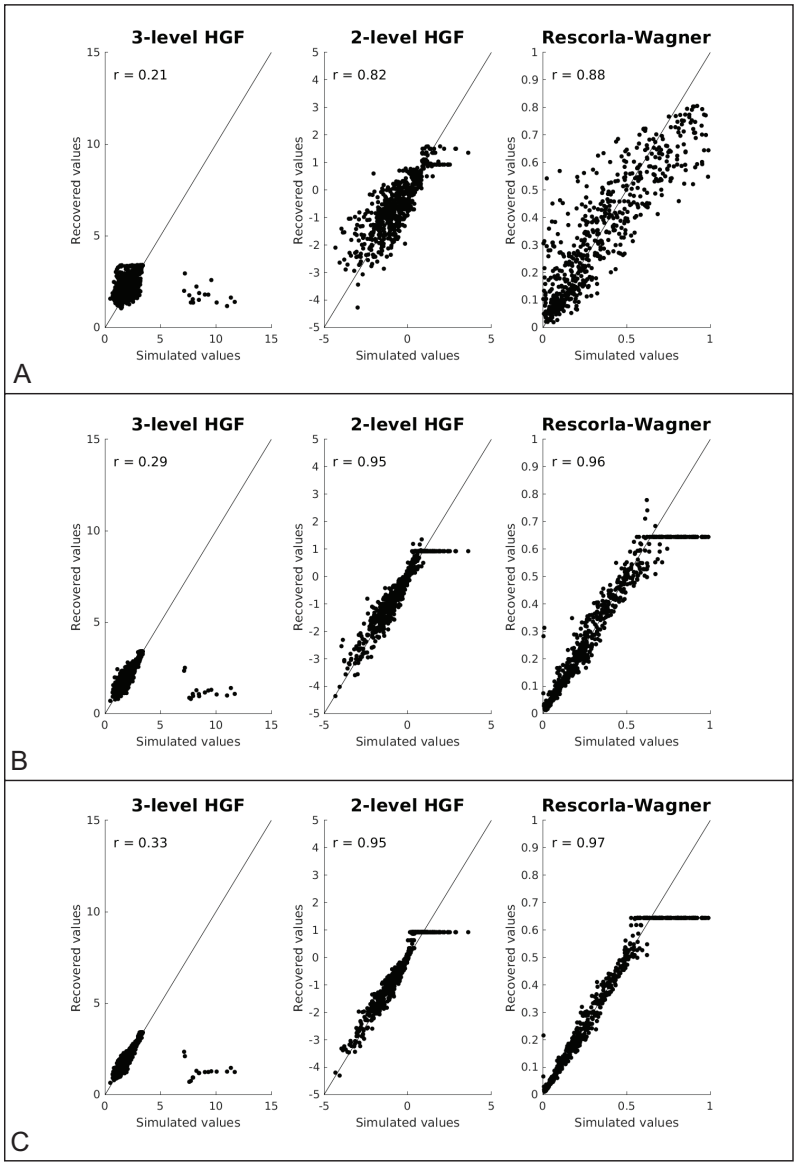
[1]Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.



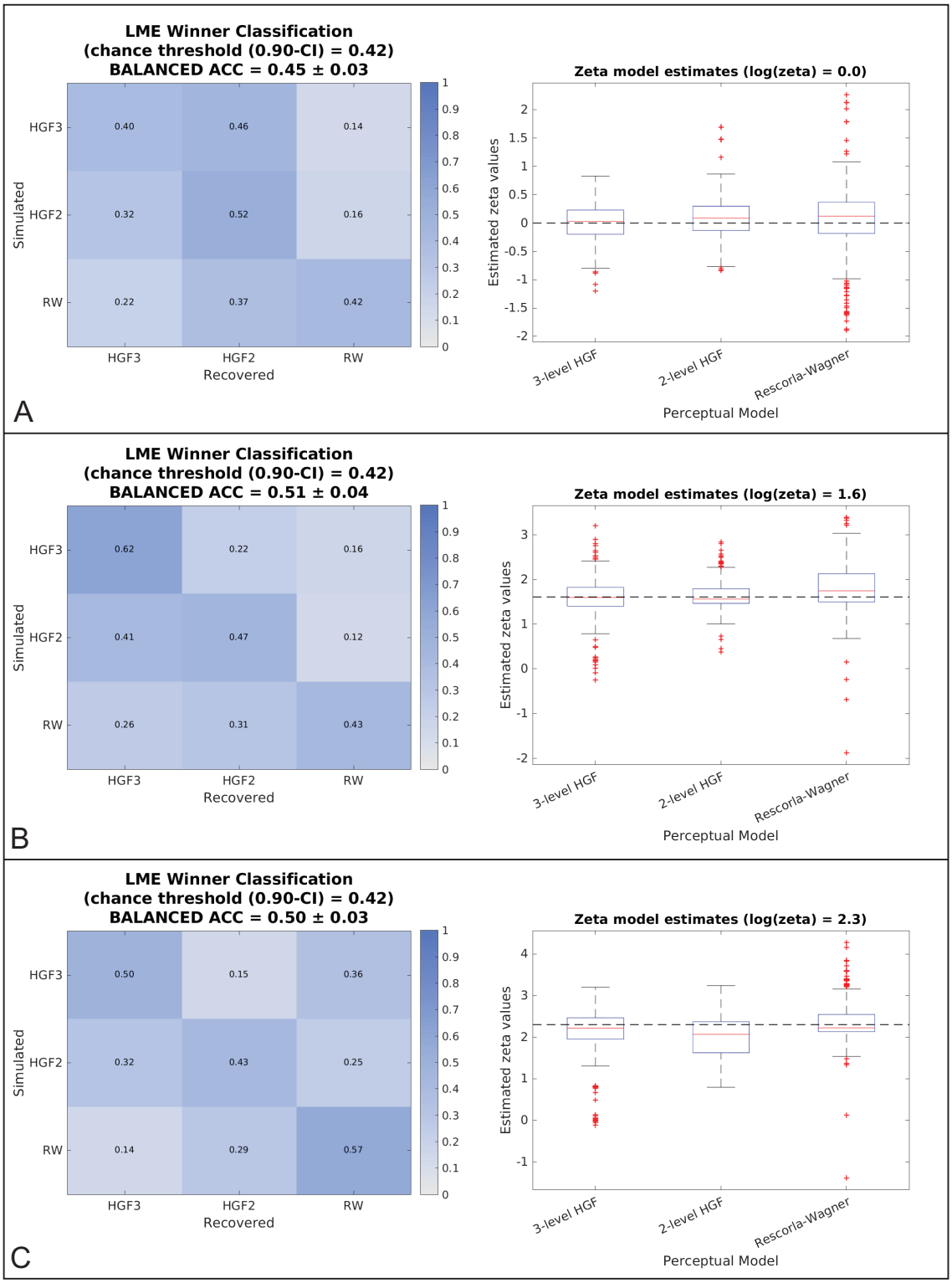
Supplementary Figure 2. Summary distributions of model parameters estimated from 8 pilot participants for each of the three candidate perceptual models (3-level Hierarchical Gaussian Filter, HGF3; 2-level Hierarchical Gaussian Filter, HGF2; and Rescorla-Wagner, RW) paired with the unit-square sigmoid response model. The fitted distributions were then used as the prior distributions for the remaining empirical data fits.

Supplementary Table 1. Parameter configurations and priors for each of the candidate models. If the prior variance is set to 0 for a parameter then it is not estimated, and the prior mean and variance for the estimated parameters (in bold) were taken from the maximum likelihood fits of the pilot participant data. Prior means are given in native space, prior variances in estimation (transformed) space.

Rescorla Wagner			
Parameter	Prior mean	Prior variance	Transformation
$v^{(0)}$	0.5	0	logit
$\boldsymbol{\alpha}$	**0.29**	**2.54**	**logit**
Observation model ζ	**2.14**	**3.33**	**log**
Hierarchical Gaussian Filter (2-level)			
Parameter	Prior mean	Prior variance	Transformation
$\mu_2^{(0)}$	0	0	none
$\mu_3^{(0)}$	1	0	none
$\sigma_2^{(0)}$	0.1	0	log
$\sigma_3^{(0)}$	1	0	log
ρ_2	0	0	none
ρ_3	0	0	none
κ_1	1	0	log
κ_2	0	0	log
$\boldsymbol{\omega_2}$	**-0.70**	**1.61**	**none**
ω_3	$-\infty$	0	none
Observation model ζ	**2.14**	**1.17**	**log**
Hierarchical Gaussian Filter (3-level)			
Parameter	Prior mean	Prior variance	Transformation
$\mu_2^{(0)}$	0	0	none
$\mu_3^{(0)}$	1	0	none
$\sigma_2^{(0)}$	0.1	0	log
$\sigma_3^{(0)}$	1	0	log
ρ_2	0	0	none
ρ_3	0	0	none
κ_1	1	0	log
$\boldsymbol{\kappa_2}$	**2.39**	**0.30**	**log**
ω_2	-3	0	none
ω_3	-6	0	none
Observation model ζ	**2.14**	**1.21**	**log**



Supplementary Figure 3. Demonstration of parameter recovery using simulated participants from the prior distributions presented in Supplementary Figure 2. Three noise levels were used for the simulations, with an inverse decision temperature (ζ) of 1 (Panel A), 5 (Panel B) and 10 (Panel C), representing very noisy ($\zeta = 1$) to very deterministic ($\zeta = 10$) settings. 60 simulated participant responses were generated using 10 different seed values, totalling n=600 simulations plotted here.

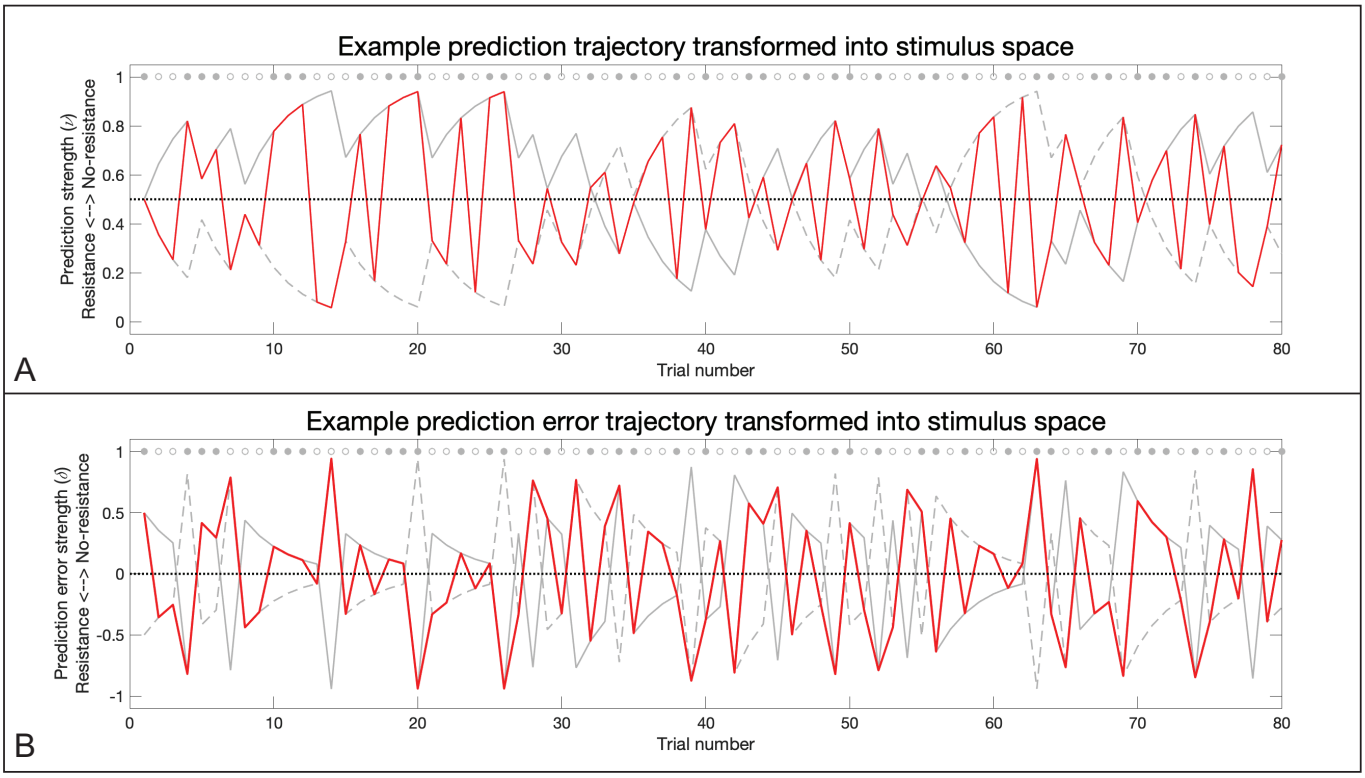


Supplementary Figure 4. Demonstration of model identifiability using simulated participants from the prior distributions presented in Supplementary Figure 2. Three noise levels were used for the simulations, with an inverse decision temperature (ζ) of 1 (Panel A), 5 (Panel B) and 10 (Panel C), representing very noisy ($\zeta = 1$) to very deterministic ($\zeta = 10$) settings. 60 simulated participant responses were generated using 10 different seed values, and the confusion matrices, balanced accuracy and zeta estimates are the average values across the 10 simulation runs.

Supplementary Table 2. Parameter configurations and priors for each of the candidate models. If the prior variance is set to 0 for a parameter then it is not estimated, and the prior mean and variance for the estimated parameters (in bold) were taken from the maximum likelihood fits of the pilot participant data. Prior means are given in native space, prior variances in estimation space. R values are Pearson Correlation Coefficients that have been Fisher's Z-transformed prior to averaging, and then back-transformed into R values.

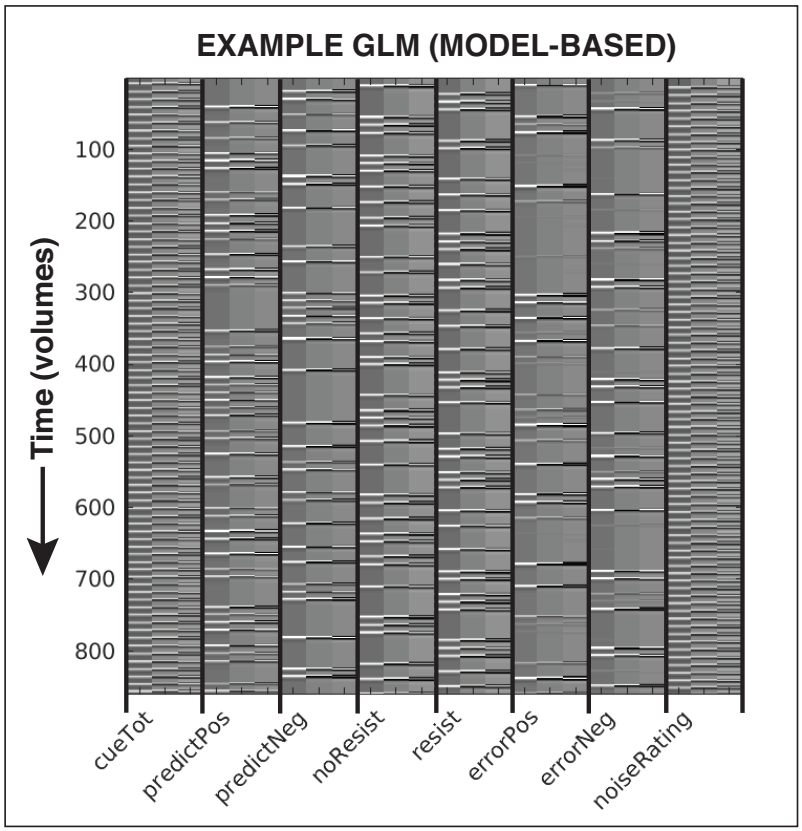
Rescorla Wagner						
	$\zeta = 1$		$\zeta = 5$		$\zeta = 10$	
	Mean	Std	Mean	Std	Mean	Std
α R	0.88	0.07	0.96	0.14	0.97	0.11
ζ_{est}	1.08	1.70	5.82	1.56	10.62	1.61
Hierarchical Gaussian Filter (2-level)						
	$\zeta = 1$		$\zeta = 5$		$\zeta = 10$	
	Mean	Std	Mean	Std	Mean	Std
ω_2 R	0.83	0.10	0.95	0.14	0.96	0.15
ζ_{est}	1.11	1.47	5.05	1.36	7.89	1.57
Hierarchical Gaussian Filter (3-level)						
	$\zeta = 1$		$\zeta = 5$		$\zeta = 10$	
	Mean	Std	Mean	Std	Mean	Std
κ_2 R	0.32	0.31	0.49	0.49	0.56	0.59
ζ_{est}	1.01	1.35	4.92	1.50	8.52	1.75

Supplementary Methods: Computational model trajectory conversion to stimulus space

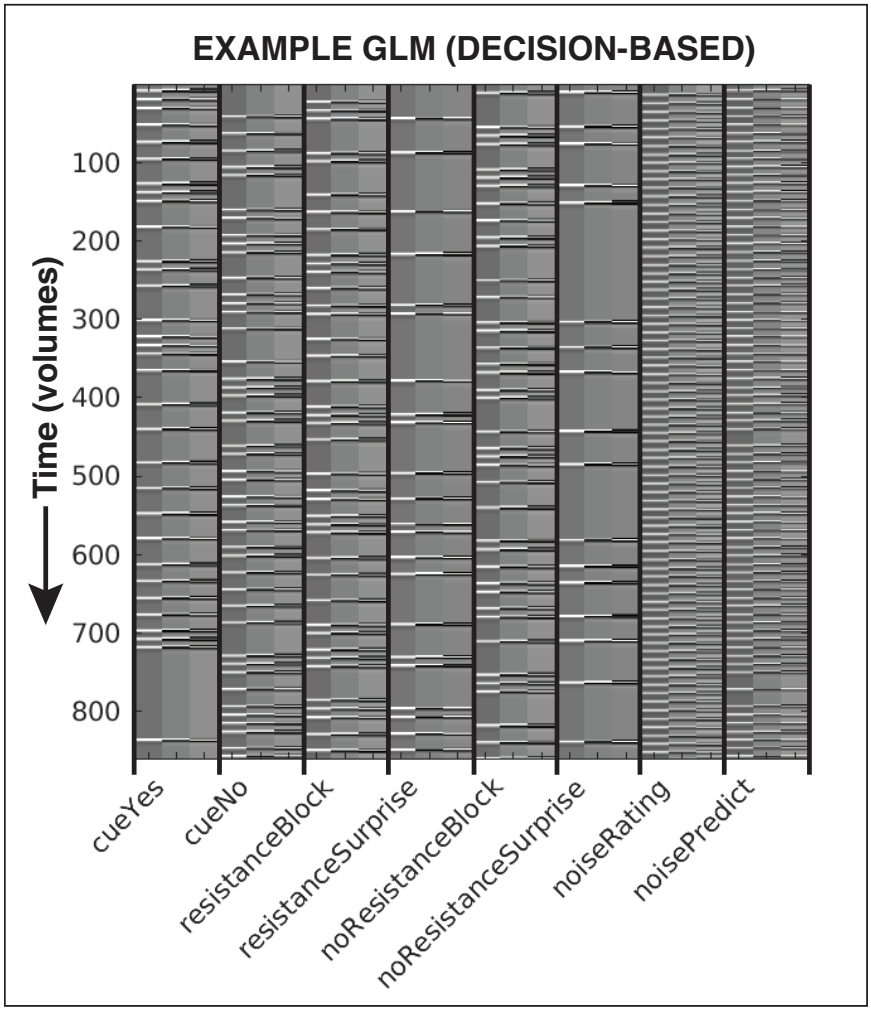


Supplementary Figure 5. Transformation of the prediction (A) and prediction error (B) trajectories from contingency space to stimulus space. In Panel A, the fitted trajectory (in contingency space) is demonstrated by the solid grey line, where the value 1 is assigned when one cue (cue 1) predicts no resistance and the opposing cue (cue 2) predicts a resistance (the value 0 is assigned for the opposing conditions). The trajectories were then transformed into stimulus space, where a value of 1 was assigned when no resistance was delivered, while a value of 0 was assigned when a resistance was delivered. For this transformation, a mirrored trajectory was firstly generated (dashed grey line) to represent the second cue, as the participants were explicitly told that the cues acted as a pair that had opposite probabilities (20% or 80%) of predicting resistance. The solid grey trajectory thus represents the cue that started with an 80% probability of being followed by no resistance in stimulus space (cue 1), while the dashed grey line represents the cue that started with a 20% probability of being followed by no resistance (cue 2). The values at each trial were taken from the trajectory of the cue that was presented at that trial: either cue 1 (trials with a closed grey circle) or cue 2 (trials with an open grey circle). The same transformation was performed on the prediction error trajectories in Panel B, where the solid grey line represents the prediction error associated with cue 1, while the dashed grey line represents the prediction error associated with cue 2. The example trajectories were taken from the participant with the closest learning rate to the mean value across all participants.

Supplementary Methods: Example fMRI general linear model

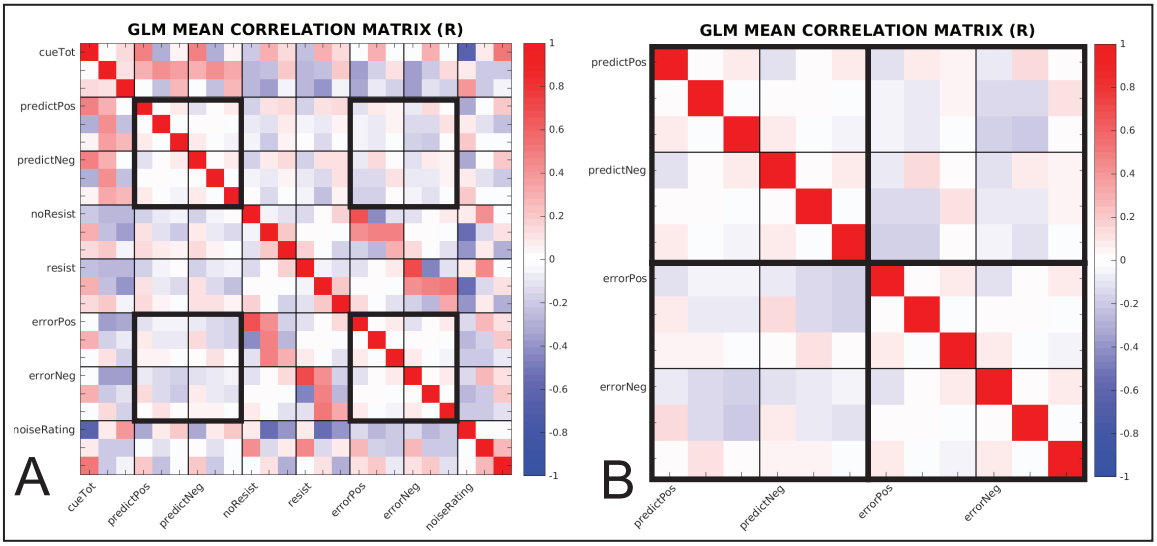


Supplementary Figure 6. An example general linear model from a single participant model-based fMRI analysis. Each of the main regressors also include a temporal and dispersion derivative, and consist of: 1) 'cueTot' – the time periods covering the presentation of all cues; 2) 'predictPos' – the value of the positive predictions (i.e. when the prediction value was closer to the no resistance condition), with an onset at the beginning of the cue and a duration of 0.5 seconds; 3) 'predictNeg' – the value of the negative predictions (i.e. when the prediction value was closer to the resistance condition), with an onset at the beginning of the cue and a duration of 0.5 seconds; 4) 'noResist' – the stimulus periods when no resistance was applied, from the onset of the first inspiration following the presentation of the circle cue to the end of the circle presentation; 5) 'resist' – the stimulus periods when resistance was applied, from the onset of the inspiration against the increased inspiratory pressure following the presentation of the circle cue to the end of the circle presentation; 6) 'errorPos' – the value of the positive prediction errors (i.e. when the prediction error value was closer to the no resistance condition), with an onset at the beginning of the corresponding stimulus period and a duration of 0.5 seconds; 7) 'errorNeg' – the value of the negative prediction errors (i.e. when the prediction error value was closer to the resistance condition), with an onset at the beginning of the corresponding stimulus period and a duration of 0.5 seconds; 8) 'noiseRating' – the periods at the end of each trial where the participant rated the intensity of the previous stimulus, with an onset at the beginning and duration that encompassed the length of the rating period. Noise regressors not shown: the convolved end-tidal carbon dioxide trace (plus temporal and dispersion derivatives), the RETROICOR and convolved respiratory volume per unit of time (RVT) regressors provided by the PhysIO toolbox, 6 motion regressors and 6 extended motion regressors (derivatives), and the timeseries of all identified noise components from the independent component analysis conducted during preprocessing were also included in the model.

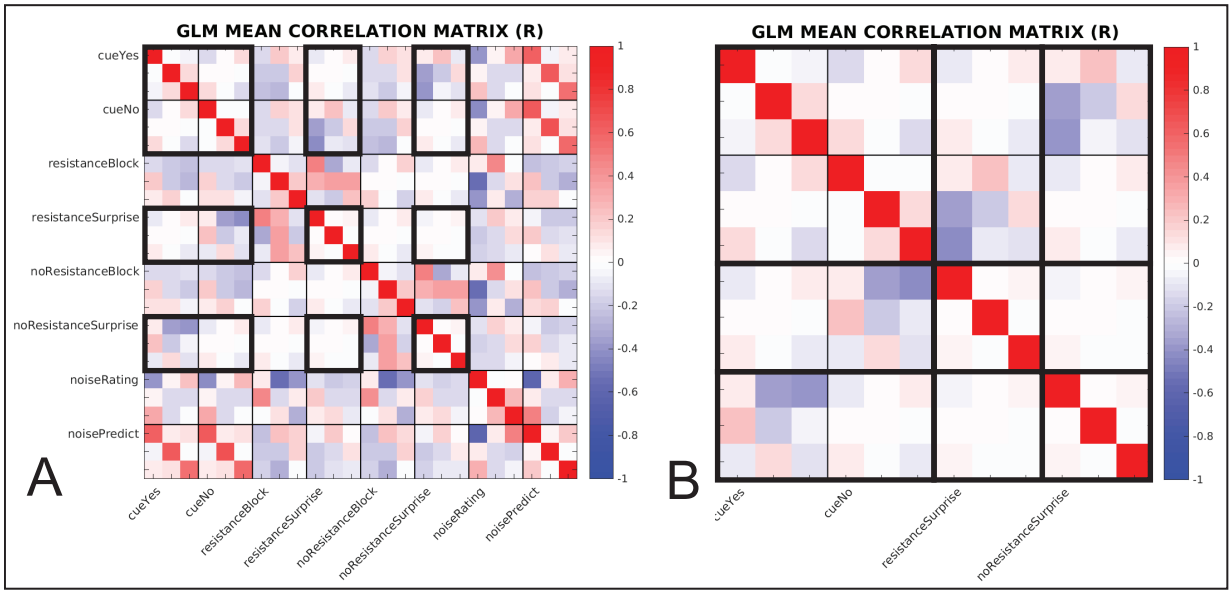


Supplementary Figure 7. An example general linear model from a single participant decision-based fMRI analysis. Each of the main regressors also include a temporal and dispersion derivative, and consist of: 1) 'cueYes' – the time periods covering the presentation of cues where the participant predicted an upcoming resistance; 2) 'cueNo' – the time periods covering the presentation of cues where the participant predicted no upcoming resistance; 3) 'resistanceBlock' – the stimulus periods when resistance was applied, from the onset of the inspiration against the increased inspiratory pressure following the presentation of the circle cue to the end of the circle presentation; 4) 'resistanceSurprise' – resistance stimuli that were surprising (i.e. when participant had predicted no resistance), with an onset at the beginning of the corresponding stimulus period and a duration of 0.5 seconds; 5) 'noResistanceBlock' – the stimulus periods when no resistance was applied, from the onset of the first inspiration following the presentation of the circle cue to the end of the circle presentation; 6) 'noResistanceSurprise' – no-resistance stimuli that were surprising (i.e. when participant had predicted resistance), with an onset at the beginning of the corresponding stimulus period and a duration of 0.5 seconds; 7) 'noiseRating' – the periods at the end of each trial where the participant rated the intensity of the previous stimulus, with an onset at the beginning and duration that encompassed the length of the rating period. 8) 'noisePredict' – the button press periods during the cue presentation, with an onset given by the response time for the button press on each trial and a duration of 0.5 seconds. Noise regressors not shown: the convolved end-tidal carbon dioxide trace (plus temporal and dispersion derivatives), the RETROICOR and convolved respiratory volume per unit of time (RVT) regressors provided by the PhysIO toolbox, 6 motion regressors and 6 extended motion regressors (derivatives), and the timeseries of all identified noise components from the independent component analysis conducted during preprocessing were also included in the model.

Supplementary Methods: Correlations within fMRI general linear models



Supplementary Figure 8. Average correlation matrix (using Fischer's R-to-Z transformation prior to averaging) from all single subject general linear models used in the fMRI analysis displayed in Supplementary Figure 6. A) Correlation matrix between all main regressors in the model (noise regressors not shown). B) Targeted correlation matrix to demonstrate the relationship between predictions and errors.



Supplementary Figure 9. Average correlation matrix (using Fischer's R-to-Z transformation prior to averaging) from all single subject general linear models used in the fMRI analysis displayed in Supplementary Figure 7. A) Correlation matrix between all main regressors in the model (noise regressors not shown). B) Targeted correlation matrix to demonstrate the relationship between predictions and errors.

Supplementary Methods: Multimodal analysis

Principal Component Analysis (PCA): PCA is an orthogonal linear transformation that transforms the $n \times m$ data matrix **X** (participants × measures) into a new matrix **P**, where the dimensions of the variance explained in the data are projected onto the new 'principal components' in descending order. Each principal component consists of a vector of coefficients or weights $\boldsymbol{w}$, corresponding to the contribution of each measure m to each component. The PCA also transforms the original $n \times m$ data matrix **X** to map each row (participant) vector $\boldsymbol{x}_i$ of **X** onto a new vector of principal component scores $\boldsymbol{t}_i$, given by:

$$t_{k(i)} = \boldsymbol{x}_i \times \boldsymbol{w}_k \qquad \text{for} \qquad i = 1, \dots, n; \qquad k = 1, \dots, m$$

where $t_{k(i)}$ is the score for each participant i within each component k.

Supplementary Results: Breathing Learning Task

Supplementary Table 3: Physiological summaries and group comparison results from the stimulus periods of the 'Breathing Learning Task' (BLT). Abbreviations: Wxn, Wilcoxon rank sum test; Ttest, students independent T-test.

	Total	Low	Moderate	P-value	Test
RESISTANCE					
Avg pressure (cmH_2O)	-4.0 (3.7)	-4.2 (3.8)	-3.8 (2.9)	0.72	Wxn
Max pressure (cmH_2O)	-7.3 (4.9)	-7.3 (6.7)	-7.3 (3.9)	0.68	Wxn
Avg breathing rate (bpm)	13.6 (6.7)	14.6 (5.6)	13.0 (10.1)	0.39	Wxn
Avg breathing depth (%)	103.9 (25.5)	97.0 (23.7)	105.9 (26.1)	0.29	Ttest
Heart rate (bpm)	66.3 (13.5)	66.9 (12.5)	66.0 (15.6)	0.69	Ttest
NO RESISTANCE					
Avg pressure (cmH_2O)	-0.1 (0.1)	-0.1 (0.1)	-0.1 (0.1)	0.05	Wxn
Max pressure (cmH_2O)	-0.8 (0.5)	-0.8 (0.8)	-0.7 (0.4)	0.59	Wxn
Avg breathing rate (bpm)	14.8 (5.1)	14.9 (5.0)	14.7 (5.8)	0.92	Ttest
Avg breathing depth (%)	106.6 (17.0)	105.0 (17.0)	108.8 (21.7)	0.35	Ttest
Heart rate (bpm)	66.7 (12.3)	67.6 (12.1)	66.6 (14.0)	0.84	Wxn

Supplementary Table 4. Behavioural group comparison results from the 'Breathing Learning Task' (BLT). Behavioural parameters include the fitted perceptual and response model parameters (learning rate, α; and inverse decision temperature ζ), with all participants included in the comparison. Abbreviations: Wxn, Wilcoxon rank sum test; Ttest, students independent T-test.

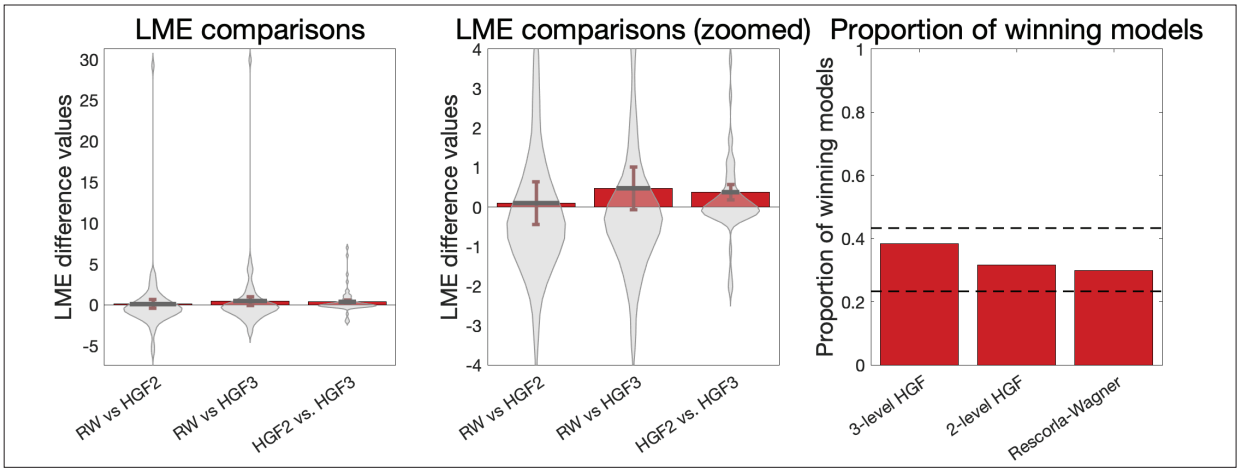
Learning rate (α)	0.25 (0.18)	0.24 (0.14)	0.25 (0.21)	0.64	Wxn
Inv. decision temp (ζ)	2.60 (3.42)	2.70 (3.12)	2.32 (3.65)	0.88	Wxn

*Supplementary Table 5. Exploratory regression analysis conducted on the fitted model learning rate parameter and the subjective ratings of breathing difficulty and anxiety. Regression parameters consisted of trait anxiety scores (from the STAI-T questionnaire), depression scores (from the CES-D questionnaire) and gender (male=1). **Significant coefficient at p<0.05 with multiple comparison correction for the three exploratory regression models.*

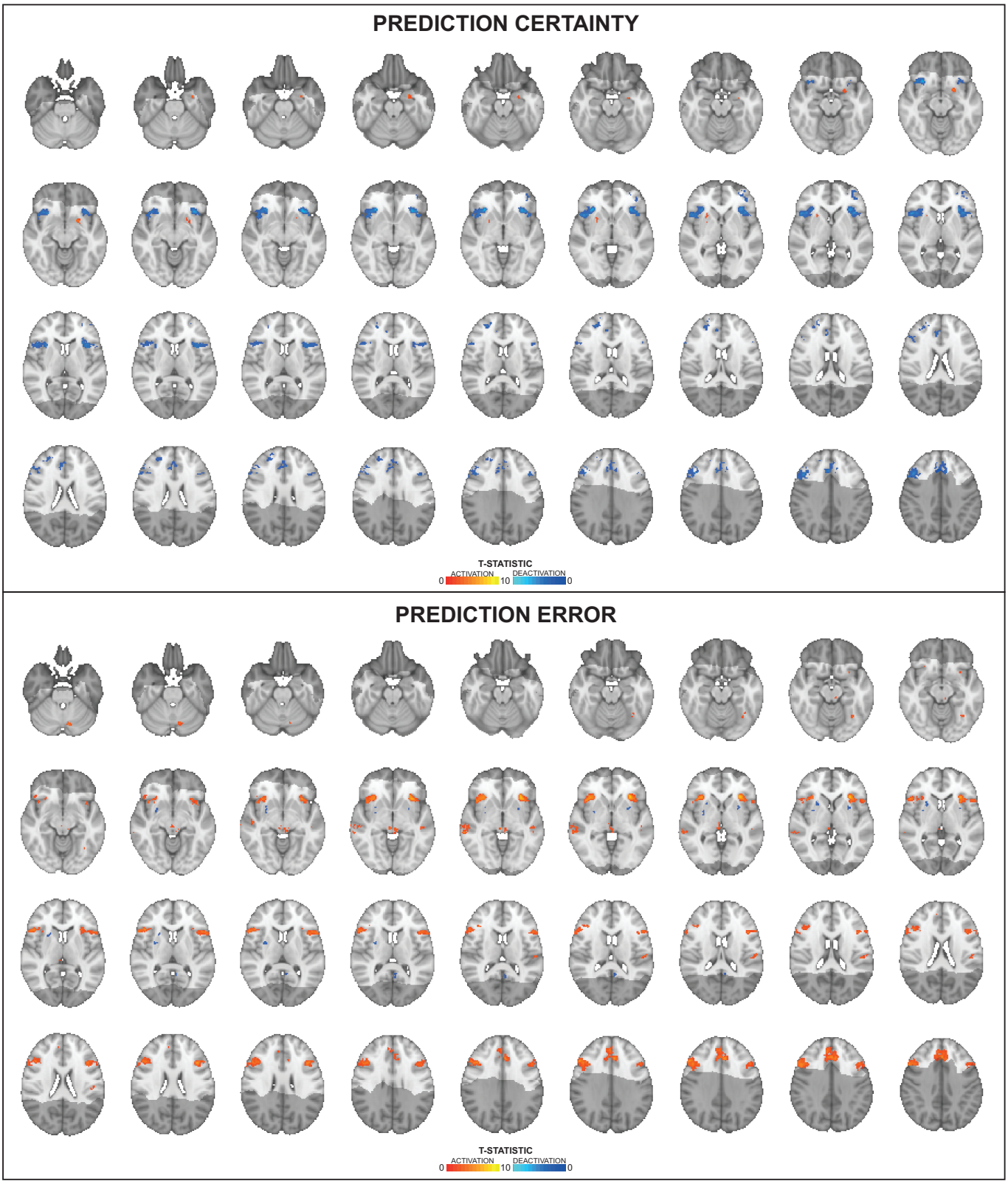
	Trait anxiety	p-value	Depression score	p-value	Gender (male)	p-value
Learning rate	< 0.01	0.99	< 0.01	0.50	**-0.14**	**< 0.01****
Breathing anxiety	1.44	0.02	0.16	0.85	-1.77	0.77
Breathing difficulty	0.59	0.05	-0.74	0.08	5.49	0.08

Supplementary Table 6. Whole and individual-group model comparison results. Abbreviations: XP, exceedance probability; PXP, protected exceedance probability.

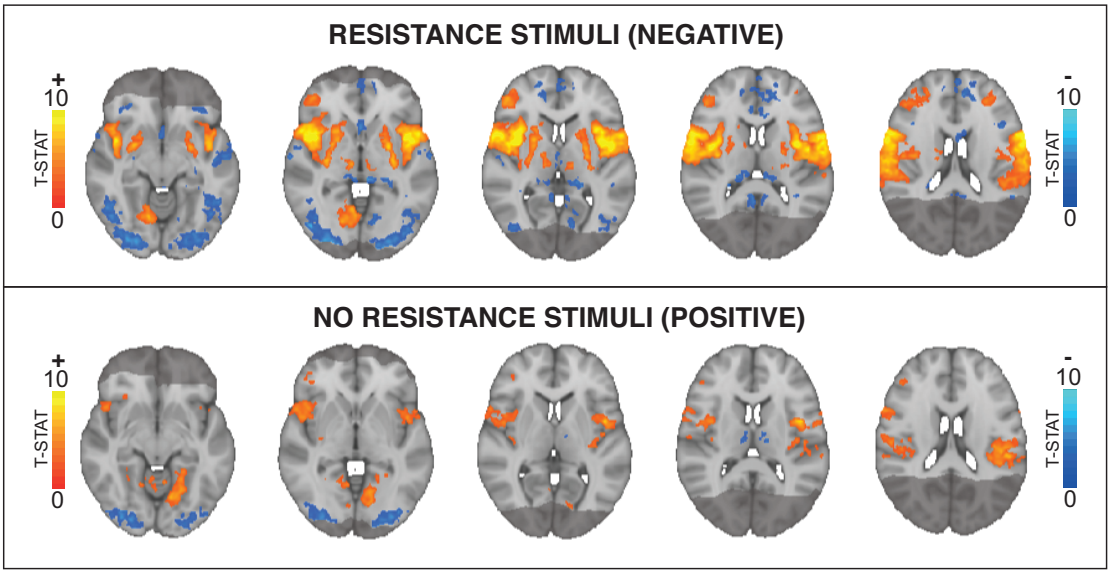
	RW	HGF2	HGF3
Whole group			
XP	0.01	0.99	0.00
PXP	0.30	0.40	0.30
Low anxiety			
XP	0.30	0.24	0.46
PXP	0.33	0.33	0.34
Moderate anxiety			
XP	0.01	0.99	0.00
PXP	0.26	0.48	0.26



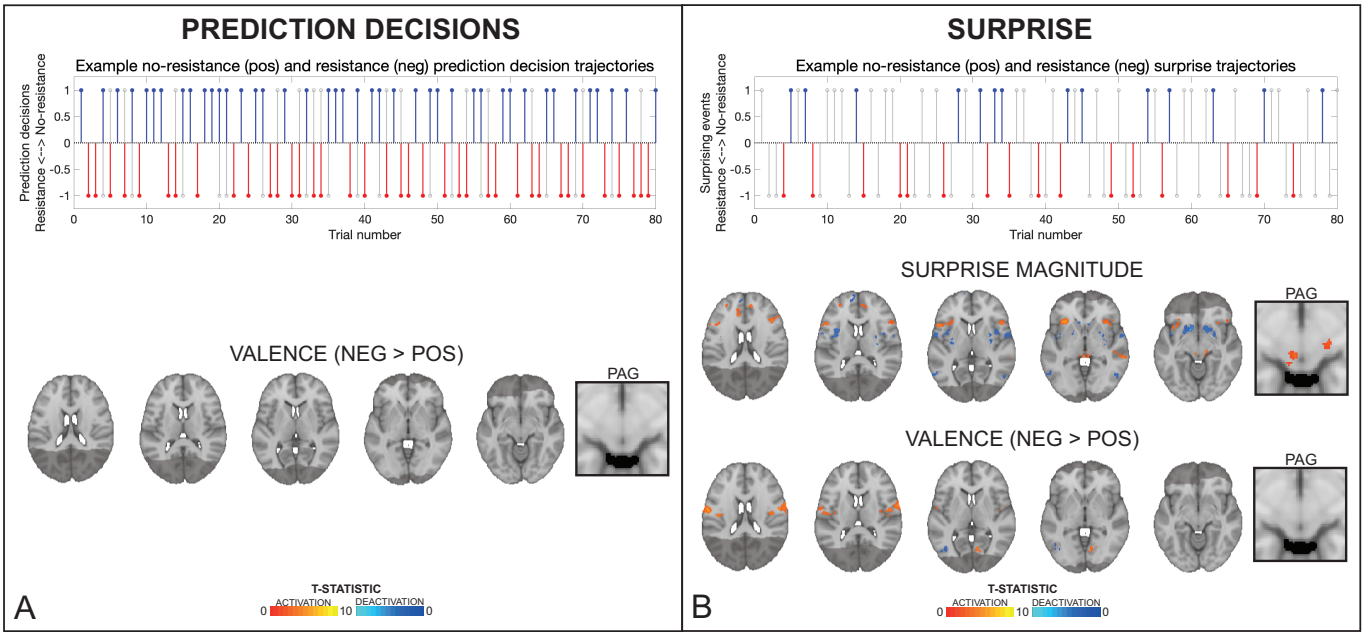
Supplementary Figure 10. Comparisons of fitted log model evidence (LME) values across all participants. Left and middle: Comparisons between each model pair, where bar plots represent mean±standard error values of the specified differences in LME, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab). Right: Proportion of highest LME values across all 60 participants. Dotted lines represent upper and lower 95% confidence intervals for chance, and all proportions of winner classifications lie within the chance range.



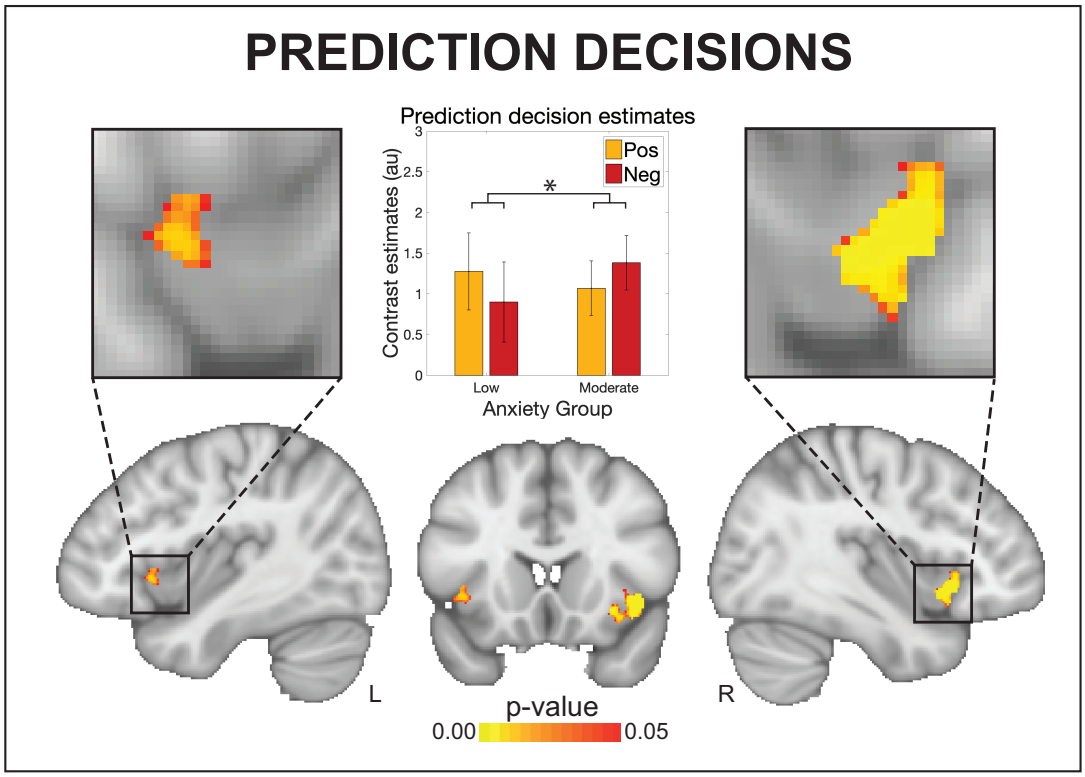
Supplementary Figure 11. A) Significant BOLD activity associated with prediction certainty, averaged over trials with positive and negative prediction certainty. B) Significant BOLD activity associated with prediction error magnitude, averaged over trials with positive and negative prediction errors. The images consist of a colour-rendered statistical map superimposed on a standard (MNI 1x1x1mm) brain. The bright grey region represents the coverage of the coronal-oblique functional scan. Significant regions are displayed with a cluster threshold of $p<0.05$, FWE corrected for multiple comparisons across all voxels included in the functional volume. Images are an expanded view of those presented in Figure 4.



Supplementary Figure 12. Whole group results from inspiratory resistance (top panel) and no inspiratory resistance (bottom panel) stimulus periods. The images consist of a colour-rendered statistical map superimposed on a standard (MNI 1x1x1mm) brain. The bright grey region represents the coverage of the coronal-oblique functional scan. Significant regions are displayed with a cluster threshold of p<0.05, FWE corrected for multiple comparisons across all voxels included in the functional volume.



Supplementary Figure 13. Overall results from the non-computational decision-based analyses of the 'Breathing Learning Task' (BLT). The plots in both (A) and (B) demonstrate how prediction decisions (in A) and surprise (in B) trajectories are encoded into positive (i.e. towards no resistance) and negative (i.e. towards resistance, red) values. The grey lines in both plots represent the stimulus at each trial, while the blue (positive) and red (negative) lines in (A) denote the prediction decisions (prior to the stimulus) and in (B) the surprising events (where the prediction decision was incorrect). In both trajectories the dotted black line denotes the boundaries between positive and negative valence, and the distance from the dotted line is taken as the final value (i.e. absolute values). The brain images in (A) represent the influence of valence on prediction decisions (difference between negative and positive decisions), while there is no equivalent representation of overall predictions in a binary decision model compared to the computational model design (the average over positive and negative prediction decisions simply represents the cue presentation). The brain images in (B) represent the activity associated with average surprise (average over positive and negative surprise trajectories) and the influence of valence on surprise (difference between negative and positive surprise trajectories). The images consist of a colour-rendered statistical map superimposed on a standard (MNI 1x1x1mm) brain. The bright grey region represents the coverage of the coronal-oblique functional scan. Significant regions are displayed with a cluster threshold of $p<0.05$, FWE corrected for multiple comparisons across all voxels included in the functional volume. Abbreviations: PAG, periaqueductal gray.

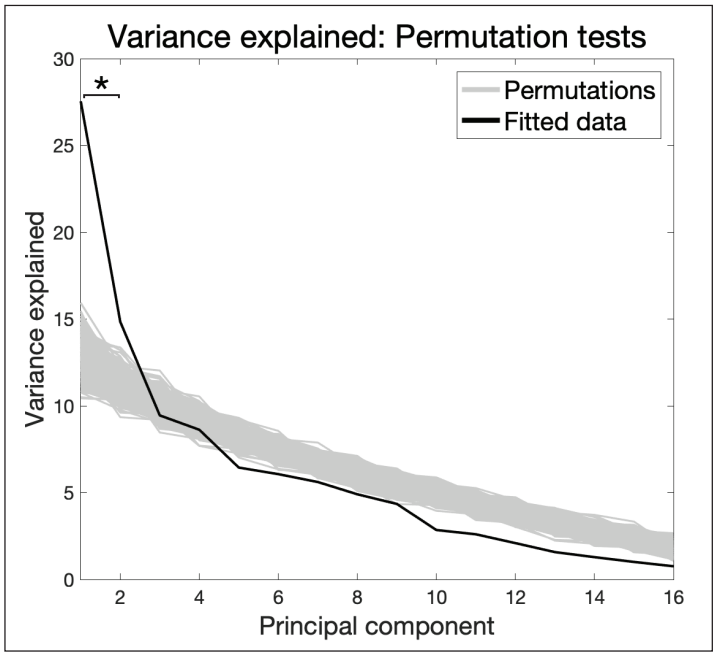


Supplementary Figure 14. Brain activity group comparison results from the binary decision model for the 'Breathing Learning Task' (BLT). An interaction effect was observed between valence (i.e. positive vs. negative) and anxiety group (low vs. moderate) for the anterior insula activity related to the valence of the prediction decisions of interoceptive breathing stimuli. The images consist of a colour-rendered statistical map superimposed on a standard (MNI 1x1x1mm) brain. Voxel-wise statistics were performed using non-parametric permutation testing within a mask of the anterior insula and periaqueductal gray, with significant results determined by p<0.05 (corrected for multiple comparisons within the mask).

Supplementary Results: Multi-modal analysis

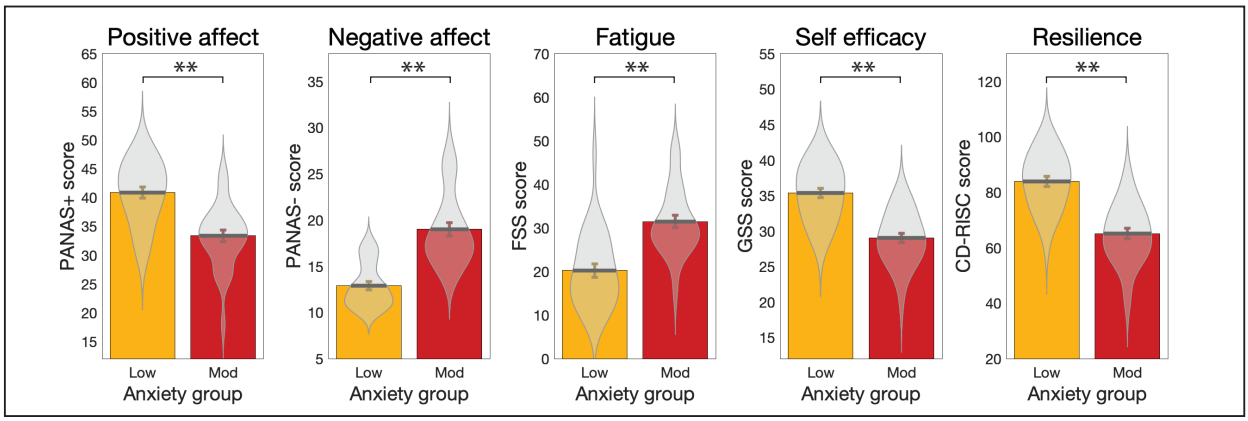
Supplementary Table 7. Correlation matrix across task modalities, with Pearson's correlation coefficients given above the diagonal and p values below the diagonal. Correlations with a p value<0.05 are represented in bold text, and those p<0.01 are shaded grey. Variables: 1) State anxiety (STAI-S); 2) Anxiety disorder (GAD-7); 3) Anxiety sensitivity (ASI); 4) Depression (CES-D); 5) Body perception (BPQ); 6) Interoceptive awareness (MAIA); 7) Breathing-related catastrophising (PCS-B); 8) Breathing-related vigilance (PVQ-B); 9) Perceptual threshold (from the FDT); 10) Decision bias (from the FDT); 11) Metacognitive bias (average confidence, from the FDT); 12) Metacognitive performance (from the FDT); 13) BOLD activity associated with positive predictions (from the BLT); 14) BOLD activity associated with negative predictions (from the BLT); 15) BOLD activity associated with positive prediction errors(from the BLT); 63) BOLD activity associated with negative prediction errors (from the BLT).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		**0.70**	**0.57**	**0.72**	0.22	**-0.57**	**0.42**	0.06	**0.26**	**-0.28**	**-0.30**	-0.23	-0.04	0.10	0.01	-0.10
2	**<0.01**		**0.55**	**0.72**	**0.35**	**-0.51**	**0.39**	0.12	0.18	-0.16	-0.25	-0.17	-0.07	0.04	0.12	-0.15
3	**<0.01**	**<0.01**		**0.68**	**0.40**	**-0.40**	**0.62**	0.20	0.20	-0.18	**-0.30**	-0.21	0.01	0.13	0.16	-0.03
4	**<0.01**	**<0.01**	**<0.01**		**0.31**	**-0.43**	**0.61**	**0.27**	**0.28**	-0.14	-0.24	-0.17	-0.01	0.13	0.12	-0.10
5	0.10	**0.01**	**<0.01**	**0.02**		-0.16	0.25	-0.02	0.04	0.19	-0.03	-0.12	0.06	0.06	0.06	-0.08
6	**<0.01**	**<0.01**	**<0.01**	**<0.01**	0.24		**-0.34**	0.02	-0.13	0.07	0.23	0.16	0.14	-0.01	-0.12	0.07
7	**<0.01**	**<0.01**	**<0.01**	**<0.01**	0.05	**0.01**		**0.49**	-0.01	0.01	-0.21	**-0.30**	0.17	0.24	-0.05	-0.10
8	0.66	0.36	0.13	**0.04**	0.89	0.87	**<0.01**		-0.10	-0.05	0.10	0.06	0.16	0.13	0.09	-0.08
9	**0.05**	0.17	0.13	**0.03**	0.78	0.34	0.97	0.48		0.20	0.05	0.02	-0.12	-0.03	0.03	-0.10
10	**0.04**	0.23	0.18	0.29	0.16	0.62	0.95	0.69	0.13		**0.35**	0.02	-0.06	0.01	0.14	0.06
11	**0.02**	0.06	**0.02**	0.07	0.85	0.08	0.12	0.46	0.70	**0.01**		**0.35**	-0.21	**-0.28**	-0.00	0.07
12	0.09	0.20	0.11	0.20	0.38	0.22	**0.02**	0.65	0.87	0.87	**0.01**		-0.23	-0.23	0.13	**0.42**
13	0.75	0.60	0.91	0.96	0.64	0.29	0.19	0.22	0.36	0.68	0.12	0.08		**0.79**	-0.22	**-0.30**
14	0.44	0.76	0.31	0.34	0.68	0.97	0.07	0.32	0.83	0.94	**0.03**	0.08	**<0.01**		-0.10	**-0.46**
15	0.92	0.39	0.23	0.37	0.64	0.36	0.72	0.52	0.80	0.28	0.98	0.33	0.10	0.45		0.08
16	0.45	0.26	0.84	0.46	0.53	0.58	0.46	0.57	0.45	0.63	0.60	**<0.01**	**0.02**	**<0.01**	0.57	

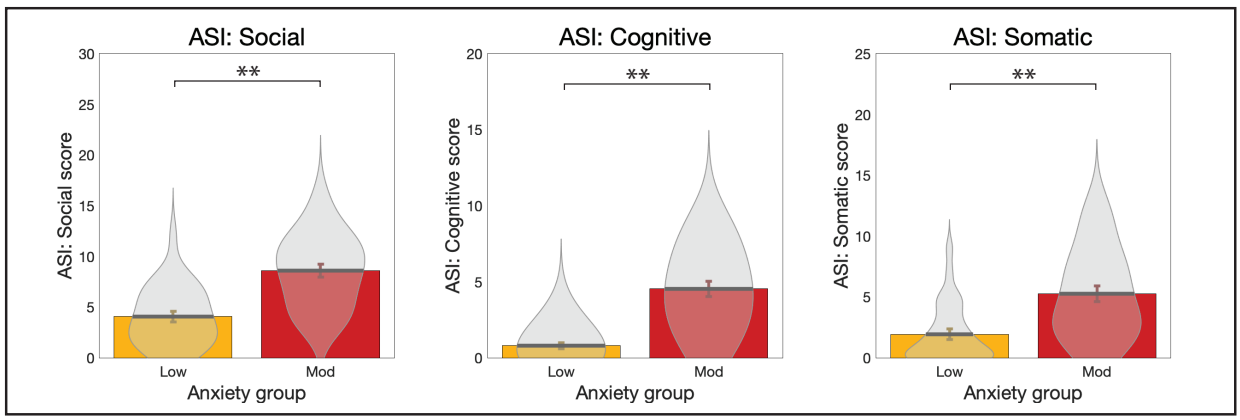


*Supplementary Figure 15. Results of the permutation tests conducted to determine the number of significant principal components within the data. The variance explained for principal components 1 and 2 are considered significant (denoted by *; p<0.05), as the values lie above the null distribution (grey lines), created by shuffling the participant scores within each measure.*

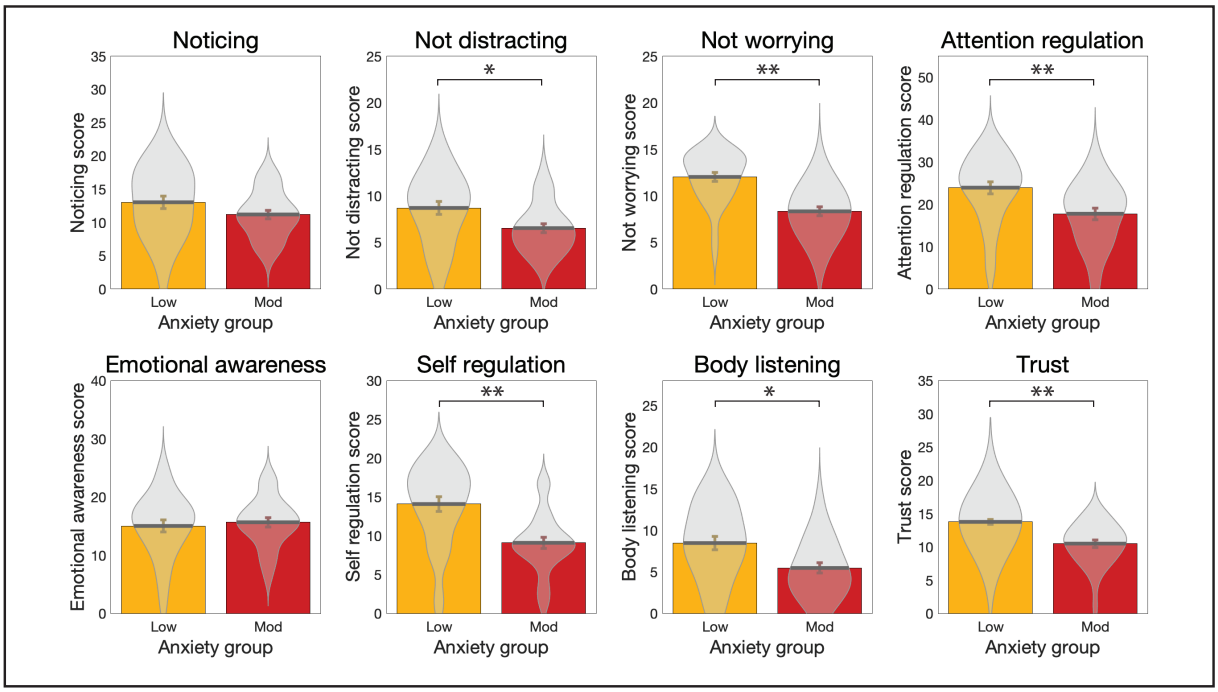
Supplementary Results: Questionnaires



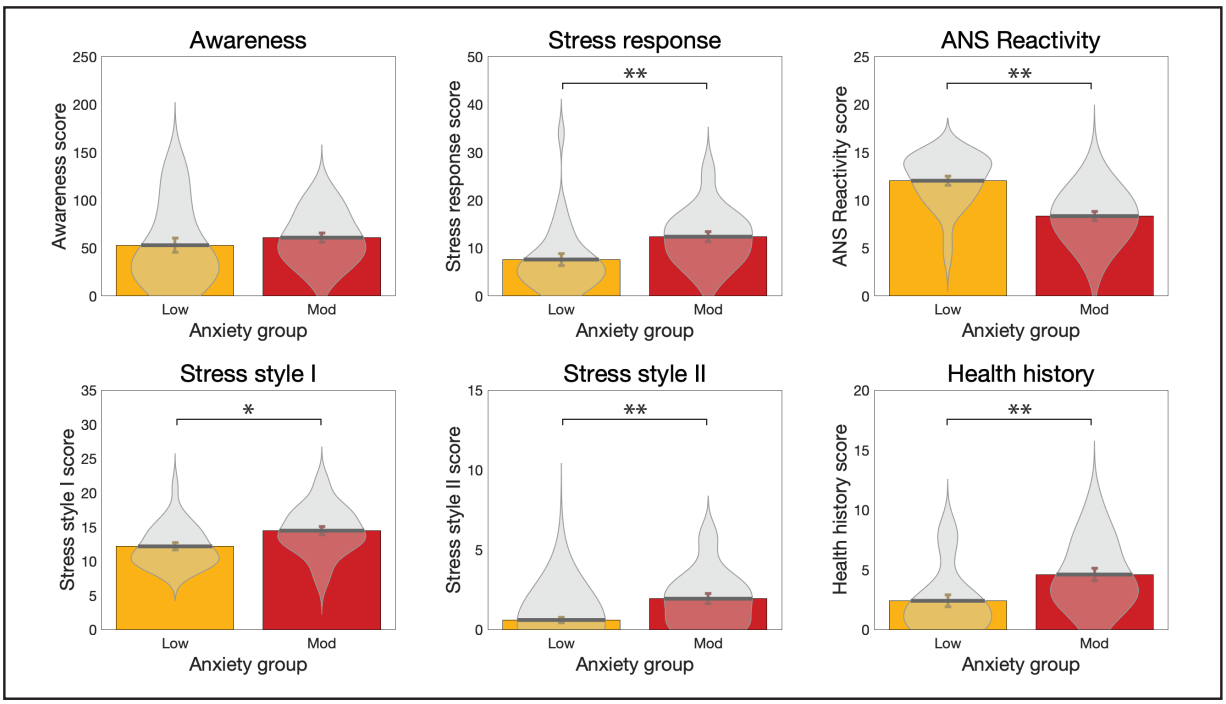
*Supplementary Figure 16. Results from the additional questionnaires measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Questionnaires: 'Positive affect' and 'Negative affect' from the Positive Affect Negative Affect Schedule (PANAS-T), 'Fatigue' from the Fatigue Severity Scale (FSS), 'Self efficacy' from the General Self-Efficacy Scale, and 'Resilience' from the Connor-Davidson Resilience Scale. Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. *Significant at p<0.05; **Significant following correction for multiple comparisons across all questionnaire measures. Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*



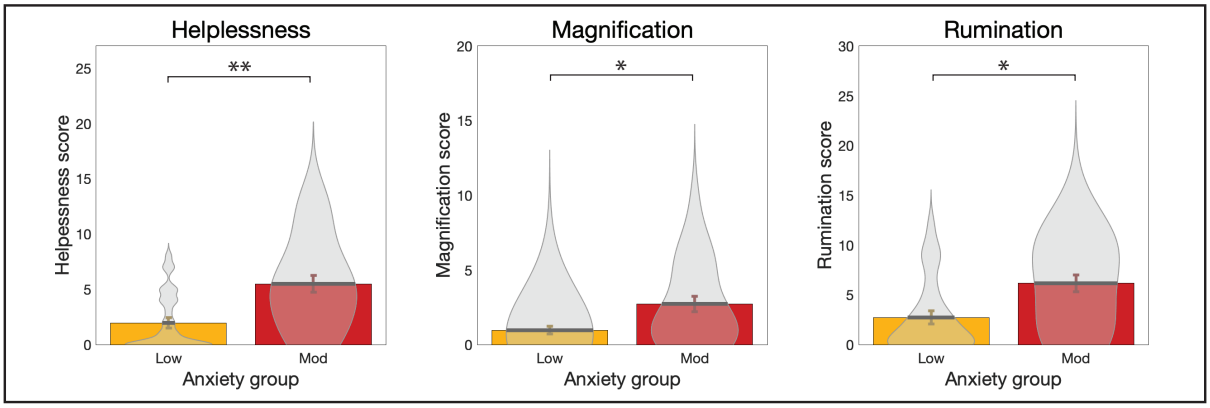
*Supplementary Figure 17. Results from the sub-scores of the Anxiety Sensitivity Index (ASI-3) questionnaire, measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. *Significant at p<0.05; **Significant at p<0.01, with no correction for multiple comparisons (exploratory results). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*



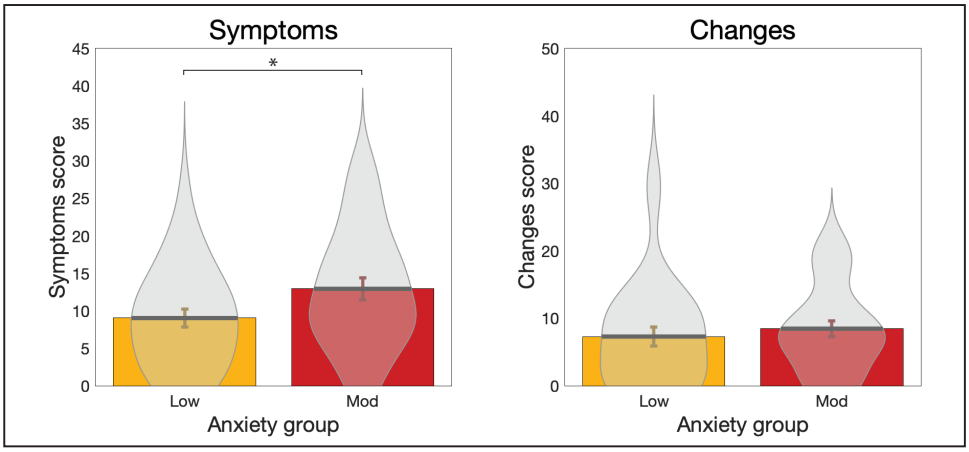
*Supplementary Figure 18. Results from the sub-scores of the Multidimensional Assessment of Interoceptive Awareness Questionnaire (MAIA), measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. **Significant at p<0.01, with no correction for multiple comparisons (exploratory results). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*



*Supplementary Figure 19. Results from the sub-scores of the Body Perception Questionnaire (BPQ), measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. **Significant at p<0.01, with no correction for multiple comparisons (exploratory results). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*



*Supplementary Figure 20. Results from the sub-scores of the Pain Catastrophising Scale (with the word 'pain' substituted for 'breathing', PCS-B), measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. **Significant at p<0.01, with no correction for multiple comparisons (exploratory results). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*



*Supplementary Figure 21. Results from the sub-scores of the Pain Vigilance Awareness Questionnaire (with the word 'pain' substituted for 'breathing', PVQ-B), measured in groups of healthy individuals with either low levels of anxiety (score of 20-25 on the Spielberger Trait Anxiety Inventory, STAI-T) or moderate anxiety (score of 35+ on the STAI-T). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. **Significant at p<0.01, with no correction for multiple comparisons (exploratory results). Bar plots represent mean±standard error values, with the distribution of values overlaid in grey. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).*