Neuron

Interoception of breathing and its relationship with anxiety

Highlights

- A novel trial-by-trial breathing-related interoception task during fMRI at 7 T
- Anxiety relates to differences in breathing perception, metacognition, and learning
- Anterior insula activity reflects respiratory predictions and prediction errors
- Anterior insula activity during interoceptive predictions differs with anxiety

Authors

Olivia K. Harrison, Laura Köchli, Stephanie Marino, ..., Frederike H. Petzschner, Samuel J. Harrison, Klaas E. Stephan

Correspondence

faull@biomed.ee.ethz.ch

In brief

Measuring brain activity while manipulating breathing resistance, Harrison et al. demonstrate that activity in the anterior insula reflects breathing-related predictions and prediction errors. Furthermore, prediction-related brain activity is altered with anxiety, along with sensitivity of and insight into breathing perceptions. These findings elucidate the link between breathing perception and anxiety.

Harrison et al., 2021, Neuron *109*, 1–14
December 15, 2021
https://doi.org/10.1016/j.neuron.2021.09.045

CellPress

Neuron

CellPress
OPEN ACCESS

Article

Interoception of breathing and its relationship with anxiety

Olivia K. Harrison,[1,2,3,8,*] Laura Köchli,[1] Stephanie Marino,[1] Roger Luechinger,[4] Franciszek Hennel,[4] Katja Brand,[2] Alexander J. Hess,[1] Stefan Frässle,[1] Sandra Iglesias,[1] Fabien Vinckier,[1,5,6] Frederike H. Petzschner,[1] Samuel J. Harrison,[1,3] and Klaas E. Stephan[1,7]

[1]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland
[2]Department of Psychology, University of Otago, Dunedin, New Zealand
[3]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[4]Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland
[5]Université de Paris, Paris, France
[6]Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, Paris, France
[7]Max Planck Institute for Metabolism Research, Cologne, Germany
[8]Lead contact
*Correspondence: faull@biomed.ee.ethz.ch
https://doi.org/10.1016/j.neuron.2021.09.045

SUMMARY

Interoception, the perception of internal bodily states, is thought to be inextricably linked to affective qualities such as anxiety. Although 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 modeling and high-field (7 T) fMRI to assess brain activity relating to dynamic changes in inspiratory resistance of varying predictability. Notably, the anterior insula was associated with both breathing-related prediction certainty and prediction errors, suggesting an important role in representing and updating models of the body. Individuals with low versus moderate anxiety traits showed differential anterior insula activity for prediction certainty. Multi-modal analyses of data from fMRI, computational assessments of breathing-related metacognition, and questionnaires demonstrated that anxiety-interoception links span all levels from perceptual sensitivity to metacognition, with strong effects seen at higher levels of interoceptive processes.

INTRODUCTION

We perceive the world through our body. Although 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 (Barrett and Simmons, 2015; Craig, 2002; Seth, 2013; Tsakiris and Critchley, 2016). ''Interoception,'' the perception of our body and inner physiological condition (Seth, 2013), constitutes a fundamental component of cerebral processes for maintaining bodily homeostasis (Berntson and Khalsa, 2021; Chen et al., 2021; Pezzulo et al., 2015; Quigley et al., 2021; Stephan et al., 2016). However, it has also been suggested to play a wider role within the systems governing emotion, social cognition, and decision making (Adolfi et al., 2017; Tsakiris and Critchley, 2016). An impaired ability to monitor bodily signals has also been hypothesized to exist across a host of psychiatric illnesses (Bonaz et al., 2021; Khalsa et al., 2018), in particular anxiety (Paulus, 2013; Paulus and Stein, 2010). 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 (Paulus, 2013). 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 (Paulus and Stein, 2010), which can initiate and perpetuate symptoms of anxiety.

Studying interoception is not without significant challenges, as bodily signals are both noisy and difficult to safely manipulate (Khalsa et al., 2018). Controlled manipulations of respiratory processes represent a promising way to address these challenges: suitable experimental setups allow dynamic yet safe changes in visceral aspects of respiration as one interoceptive modality (Berner et al., 2018; DeVille et al., 2018; Faull and Pattinson, 2017; Faull et al., 2016, 2018; Hayen et al., 2017; Paulus et al., 2012; Rieger et al., 2020). Furthermore, given the vitally important role of breathing for survival, respiratory changes are highly salient. Indeed, labored, unsatisfied, unexpected, or uncontrolled breathing can itself be perceived as a dangerous and

debilitating interoceptive threat (Hayen et al., 2013; Herigstad et al., 2011; Schwartzstein et al., 1990). Beyond respiratory diseases (Carrieri-Kohlman et al., 2010; Hayen et al., 2013; Herigstad et al., 2011; Janssens et al., 2011; Marlow et al., 2019; Parshall et al., 2012), aversive breathing symptoms have been noted to be particularly prevalent in individuals suffering from psychiatric conditions such as anxiety and panic disorder (Giardino et al., 2010; Mallorquí-Bagué et al., 2016; McNally and Eke, 1996; Paulus, 2013; Smoller et al., 1996; Woods et al., 1986).

Work toward conceptualizing 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 (Critchley and Garfinkel, 2017). These layers encompass multiple levels, ranging from metrics of afferent signal strength at "lower" levels (using techniques such as heartbeat evoked potentials; Allen et al., 2016; Petzschner et al., 2019) and psychophysical properties (such as measuring perceptual sensitivity; Domschke et al., 2010; Klecker et al., 2015; Petzschner et al., 2017) to psychological and cognitive components at "higher" levels (Critchley and Garfinkel, 2017). Notable domains within these higher levels include attention toward bodily signals (Berner et al., 2018; Murphy et al., 2019; Wang et al., 2019), static and dynamic beliefs and models of body state (Critchley and Garfinkel, 2017; Seth, 2013; Tsakiris and Critchley, 2016), and insight into both our interoceptive abilities (Garfinkel et al., 2015, 2016a, 2016b; Harrison et al., 2021a) and the accuracy of our interoceptive beliefs ("metacognition") (Petzschner et al., 2017; Stephan et al., 2016). 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, such as predictive coding (Behrens et al., 2007; Feldman and Friston, 2010; Friston, 2005; O'Reilly et al., 2012), have been extended to interoception and used to propose how the brain may build models of the changing internal environment (Barrett and Simmons, 2015; Gu et al., 2013; Seth, 2013; Seth et al., 2012; Stephan et al., 2016). Although theoretical models have been proposed, realistic synthetic data have been produced (Allen et al., 2019; Tschantz et al., 2021), and initial learning models have now been fit to empirical cardiac data (Smith et al., 2020), concurrent measures of dynamic brain processes during interoceptive learning have not yet been demonstrated.

Here, we build on these conceptual advances and assess the relationship between anxiety and breathing-related interoception across the multiple hierarchical levels of processing. Importantly, although many theoretical proposals have been put forward as to how anxiety may interrupt the brain's processing of dynamic (trial-by-trial) interoceptive predictions and/or prediction errors (Allen, 2020; Barrett and Simmons, 2015; Brewer et al., 2021; Paulus, 2013; Paulus and Stein, 2006, 2010; Paulus et al., 2019), these are as yet untested. Therefore, within a rigorous assessment profile, we included neuroimaging of a novel breathing-related interoceptive learning paradigm, providing the first empirical insight into the brain activity associated with interoceptive predictions and prediction errors. Furthermore, it is not yet known how alterations in interoceptive learning may relate to previously identified relationships between anxiety and lower level breathing sensitivity (Garfinkel et al., 2016a; Tiller et al., 1987), higher level beliefs (Ewing et al., 2017; Garfinkel et al., 2016b; Mehling, 2016; Paulus and Stein, 2010), or metacognition (Harrison et al., 2021c). Therefore, we aimed to both assess how anxiety is related to dynamic interoceptive learning and additionally provide a unifying perspective on anxiety and breathing-related interoception across the hierarchical levels of interoceptive processing. We adopted a multimodal experimental approach to investigate multiple levels of breathing-related interoceptive processing, including low-level perceptual sensitivity and related higher level metacognition via the filter detection task (FDT) (Harrison et al., 2021a), subjective interoceptive beliefs via questionnaires, and trial-by-trial interoceptive learning and related brain activity in a novel breathing learning task (BLT). Both the FDT and trial-by-trial behavioral and functional magnetic resonance imaging (fMRI) data from the BLT were analyzed 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 breathing-related interoceptive processing across the hierarchy, from sensitivity to metacognition.

RESULTS

Results overview

Below we present the results from each of our task modalities: questionnaires, a breathing perception and metacognition task (the FDT), and a novel interoceptive learning task (the BLT), in which group-wise comparisons between each of the measures of interest were conducted. The results from the questionnaires and FDT are contextualized by previous findings related to anxiety, while the results from the BLT were validated against an additional unseen dataset and the relationship with anxiety was assessed. We then present the results of a combined multi-modal analysis, in which we compared our measures both within and across task modalities. The principal components (PCs) of these measures were identified to formalize and asses any shared variance between measures, which spanned multiple dimensions of breathing-related interoceptive processing.

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 1. The group summary values and statistics presented in text are either mean ± SE 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 used were found to be highly significantly different between groups with low and moderate trait anxiety: individuals with moderate levels of trait anxiety demonstrated higher state anxiety (Spielberger State Anxiety Inventory [STAI-S] mean ± SE: low anxiety, 25.7 ± 0.7; moderate anxiety, 34.1 ± 1.2; $t = -6.1$, $p < 0.01$), higher levels of anxiety disorder symptoms (Generalized Anxiety Disorder Questionnaire

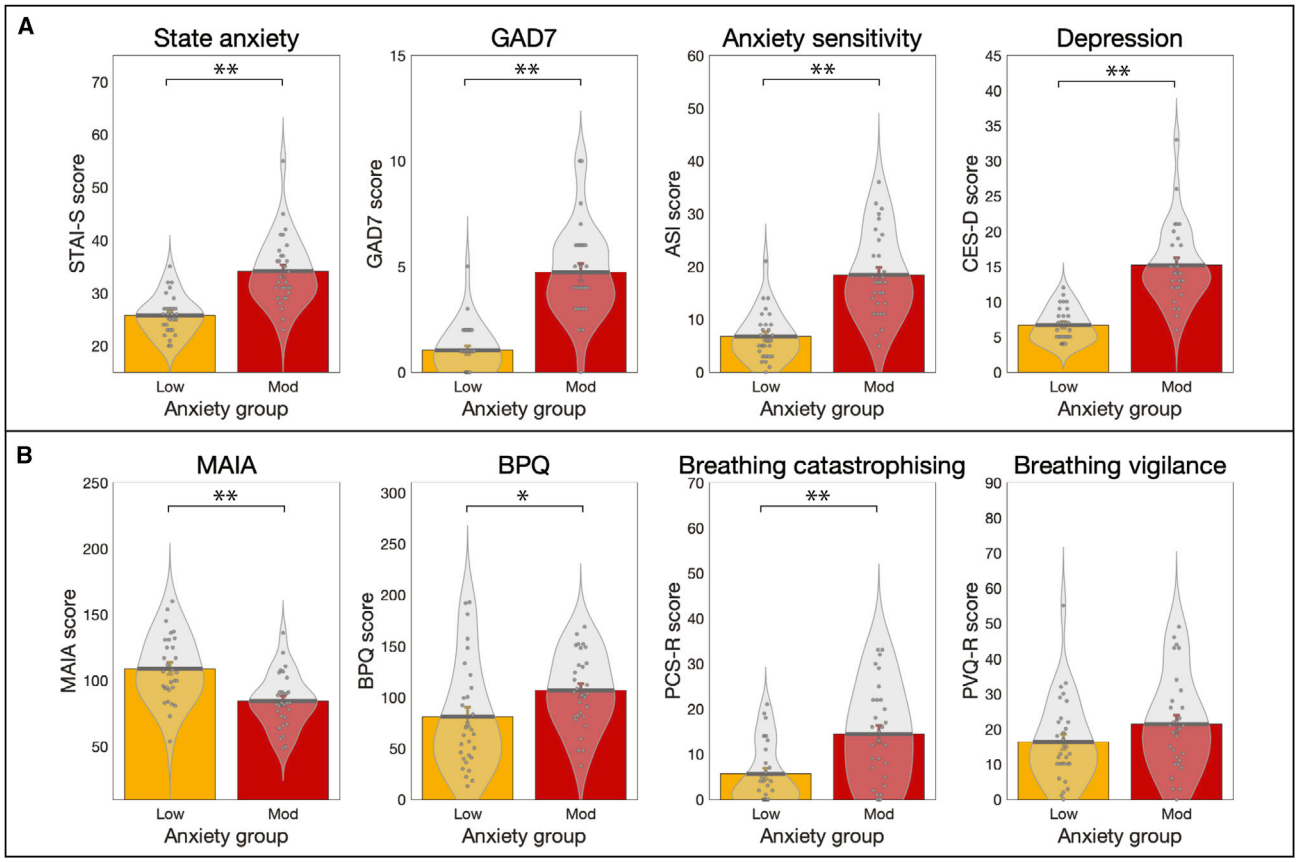

Figure 1. Results from the affective and interoceptive questionnaires measured in groups of healthy individuals with either low or moderate levels of anxiety

Participants with low anxiety scored 20–25 on the Spielberger Trait Anxiety Inventory (STAI-T), and those with moderate anxiety scored ≥35 on the STAI-T.

(A) Affective questionnaires: state anxiety, Spielberger State Anxiety Inventory; GAD-7, Generalized Anxiety Disorder Questionnaire; anxiety sensitivity, Anxiety Sensitivity Index; depression, Center for Epidemiologic Studies Depression Scale.

(B) Interoceptive questionnaires: MAIA, Multidimensional Assessment of Interoceptive Awareness Questionnaire; BPQ, Body Perception Questionnaire; breathing catastrophizing, Pain Catastrophizing 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 eight questionnaires. Bar plots represent mean ± SE, with the distribution of values overlaid in gray. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab). See also Figure S1.

[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 (Anxiety Sensitivity Index [ASI] mean ± SE: low anxiety, 6.8 ± 0.8; moderate anxiety, 18.4 ± 1.5; $t = -6.9$, $p < 0.01$), and higher levels of depression symptoms (Center for Epidemiologic Studies Depression Scale [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 catastrophizing, and breathing symptom vigilance. All except breathing-related vigilance were also found to be significantly different between groups. Individuals with moderate levels of trait anxiety demonstrated reduced "positively minded" interoceptive awareness (Multidimensional Assessment of Interoceptive Awareness [MAIA] mean ± SE: 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 (Body Perception Questionnaire [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 (Ewing et al., 2017; Garfinkel et al., 2016b; Mehling, 2016; Paulus and Stein, 2010). Additionally, elevated levels of breathing-related catastrophizing were observed in the moderate anxiety group (Pain Catastrophizing Scale [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 (Pain Vigilance Awareness Questionnaire [PVQ-B] mean ± SE: 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 Figure S1.

FDT results

The group summaries and comparisons for each of the FDT measures are displayed in Figure 2. The FDT output includes

Please cite this article in press as: Harrison et al., Interoception of breathing and its relationship with anxiety, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.09.045

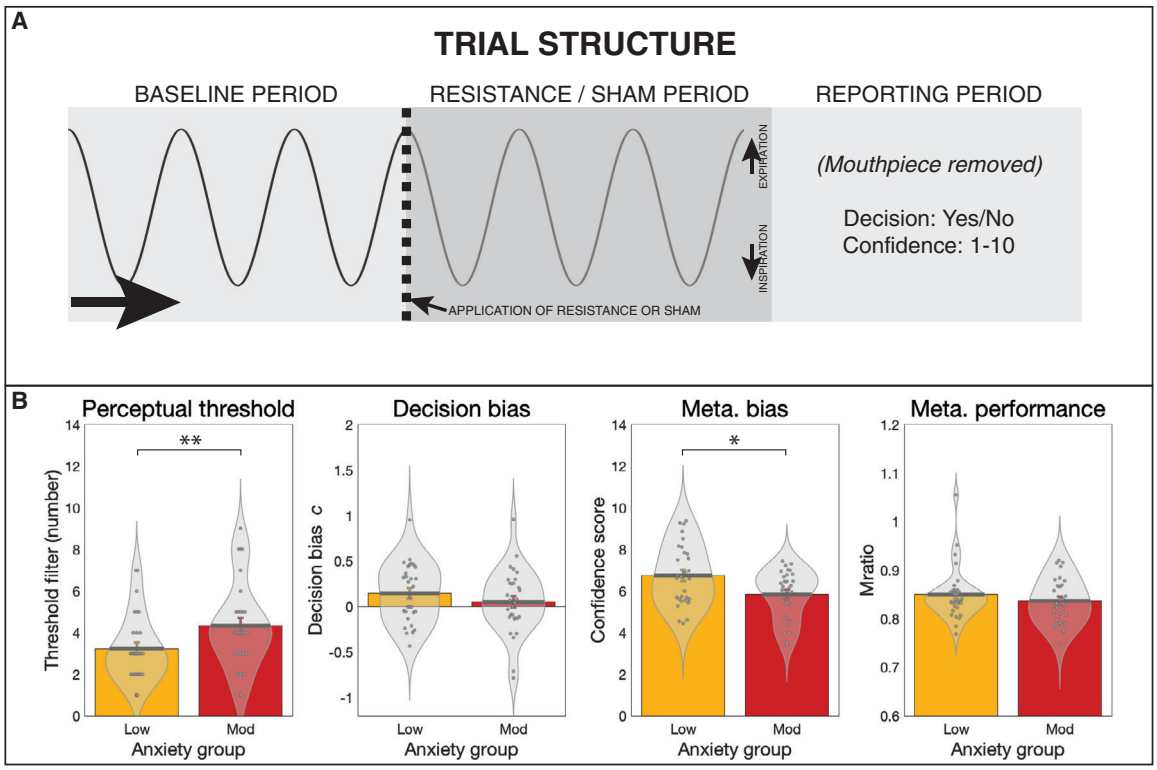

Figure 2. Trial structure and results from the filter detection task (FDT)

(A) 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 or no), and their confidence in their decision (1–10, where 1 = not at all confident/guessing and 10 = maximally confident). Adapted from Harrison et al. (2021a) under Creative Commons license.

(B) Results from the FDT: individuals with moderate anxiety (scores of ≥35 on the Spielberger Trait Anxiety Inventory [STAI-T]) demonstrated a higher (less sensitive) perceptual threshold and lower metacognitive bias (lower average confidence) compared with individuals with low levels of anxiety (scores of 20–25 on the STAI-T). No difference was found between groups for decision bias (where *c* values below zero indicate a tendency to report the presence of resistance) or 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 ± SE, with the distribution of values overlaid in gray. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab).

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 resistance), 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; Garfinkel et al., 2016a; Tiller et al., 1987) (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 ± SE; 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 (Rouault et al., 2018).

BLT results

Behavioral data modeling

A visual depiction of the BLT and example fitted trajectories for prediction certainty and prediction error magnitude are provided in Figure 3. When comparing the plausibility of the three alternative models (a Rescorla-Wagner [RW] model; a two-level hierarchical Gaussian filter [HGF2], and a three-level hierarchical Gaussian filter [HGF3]) using random-effects Bayesian model selection (Rigoux et al., 2014; Stephan et al., 2009), 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; Table S4). 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. Importantly, the finding that none of the models demonstrated a PXP greater than 90% does not provide

any absolute statement about the quality of these models. Rather, this finding indicates that none of the chosen models is conclusively superior to the others in explaining the data. To ensure that the chosen model (RW) provided an adequate explanation of the data, we compared it to a "null model" (i.e., where the choices were due to chance and not related to any associative learning mechanism) using a likelihood ratio test. We found that in 58 of the 60 participants, the RW model fit the behavior significantly better than the null model, demonstrating that the chosen model captured important aspects of their behavior. The two participants (one from each anxiety group) who did not show model fits above chance were excluded from any further model-based analyses and comparisons.

To further establish the adequacy of our chosen model to explain learning behavior in this novel interoceptive learning task, we completed a model validation on 15 additional held-out datasets for the BLT. These participants were not preselected for any particular level of anxiety (see STAR Methods for details). A logistic regression was conducted to assess whether the model prediction trajectory from the original data was able to significantly explain the prediction decisions made by the 15 unseen participants in the validation sample. A representative prediction trajectory from the original 60 participant model fits (the trajectory from the participant with the closest learning rate to the mean) was used in this regression, as well as an intercept term. The beta estimate for the original prediction trajectory was 3.1 ± 0.3 ($t = 12.1$, $p = 1.0 \times 10^{-33}$), denoting a highly significant ability of the trajectory to predict unseen data. The beta estimate for the intercept term was -0.2 ± 0.1 ($t = -3.1$, $p = 1.7 \times 10^{-3}$). For a qualitative representation of model fits for both the original and validation data, see Figure S3.

Both model-based and behavioral parameter comparisons are presented in Table 1. For the estimated model parameters, no difference was observed between the groups for either learning rate (α) or inverse decision temperature (ζ) (Table 1). Results from parameter comparisons between groups including the excluded participants can be found in Table S5. 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 (Table S1), nor in relative head motion during the task (average root-mean-square displacement ± SD: low anxiety, 0.17 ± 0.10 mm; moderate anxiety, 0.18 ± 0.07 mm; Wilcoxon rank-sum $p = 0.91$).

Computational modeling 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 associated with breathing-related prediction certainty and prediction errors across all participants are shown in Figure 4. Dorsolateral prefrontal cortex (dlPFC), anterior insula (aIns), anterior cingulate cortex (ACC), and middle frontal gyrus (MFG) all 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 the periaqueductal gray (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). Although no main effect of anxiety group was observed, an interaction effect was found using the region-of-interest (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 Figure S6 for comparison with previously published results (Berner et al., 2018; DeVille et al., 2018; Faull and Pattinson, 2017; Faull et al., 2016, 2018; Hayen et al., 2017; Paulus et al., 2012).

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 as well as across levels of breathing-related interoceptive processing. The full correlation matrix of all 16 included measures is displayed in Figure 6A and Table S7. To briefly summarize, the strongest across-task modality correlations were found between all affective and interoceptive questionnaires (Figure 6A). Concerning affective questionnaires 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 catastrophizing was weakly related to metacognitive performance on the FDT. Last, between the FDT and aIns activity, metacognitive performance was strongly related to the peak aIns activity associated with negative (i.e., resistance-related) prediction errors, while metacognitive bias was weakly related to aIns activity associated with negative (i.e., resistance-related) prediction certainty. Non-parametric correlations (using Spearman's rho values) produced highly consistent results and are presented in Table S7.

Principal-component analysis

Finally, to assess the extent of shared variance across interoceptive measures, the multi-modal data matrix was then subjected to a principal-component analysis (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 anxiety. Two PCs were found to be significant, where the variance explained with each component was above the 95% confidence interval of its null distribution. The properties of each of these significant components are displayed in Figures 6B and 6C. The first PC demonstrated a highly significant ($p < 1 \times 10^{-11}$) 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 catastrophizing 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

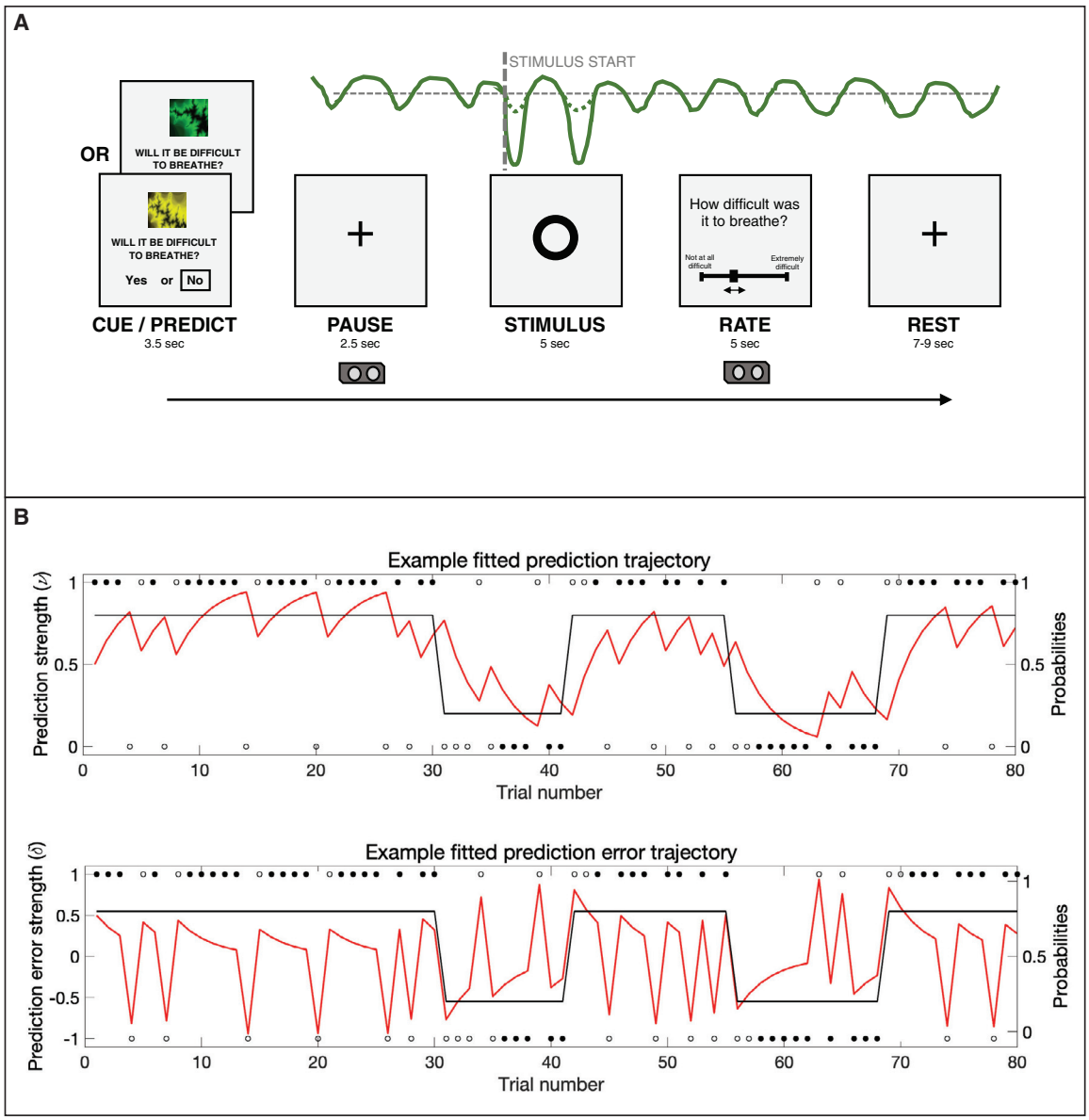

Figure 3. The breathing learning task (BLT), used to measure dynamic learning of breathing-related stimuli

(A) An overview of the single trial structure, in which one of two cues was presented and participants were asked to predict (on the basis of 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 s, 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 (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.

See also Figures S2 and S3 and Tables S1–S4 and S6.

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.

DISCUSSION

Main findings

Interoceptive abilities are thought to be tightly linked to affective properties such as anxiety. Here we have provided a unifying analysis by characterizing this relationship across multiple

Please cite this article in press as: Harrison et al., Interoception of breathing and its relationship with anxiety, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.09.045

CellPress
OPEN ACCESS

Table 1. Behavioral and model-based group comparison results from the breathing learning task (BLT)

	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 temperature (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[a]	Wxn
Response time (s)	1.28 (0.33)	1.23 (0.36)	1.29 (0.30)	0.73	Wxn

All data are expressed 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. Wxn, Wilcoxon rank-sum test. If a Wilcoxon rank-sum test was used, reported values are median (inter-quartile range).

[a]Significant difference between groups at $p < 0.05$ with multiple comparison correction for the number of behavioral parameters. See also Table S5.

interoceptive levels in the breathing domain, including novel findings of altered brain activity within the aIns when processing dynamic breathing predictions. This study is the first to demonstrate brain activity related to dynamic interoceptive learning and, specifically, activity in the insula that is related to breathing-related prediction error and prediction certainty in a trial-by-trial fashion. Notably, aIns activity related to the certainty of predictions about breathing resistance was found to differ between trait anxiety levels. Furthermore, this study is also among the first to simultaneously tackle multiple levels of interoceptive processing, using breathing as a salient and accessible channel of body perception. The tasks used reflect the broad range of targeted processes; not only were questionnaires used that spanned affect and body perceptions, but behavioral data from two different tasks were assessed by separate computational models. These analyses allowed formal assessments of both breathing-related 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. Our multi-modal approach revealed that not only is the relationship between breathing-related interoception and trait anxiety broad, it is most strongly detected (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"; Garfinkel et al., 2015) followed by metacognitive aspects of breathing perceptions. Notably, the peak aIns brain activity associated with breathing-related interoceptive learning appeared to be largely independent of other interoceptive measures, with the exception of negative prediction error-related brain activity and metacognitive performance.

Affect and levels of breathing-related 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. Using 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 that 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 catastrophizing. General body awareness and the breathing-related metacognitive measures (bias and performance) were the next largest contributors to this shared variance, followed by the perceptual sensitivity and decision bias parameters, and last peak aIns activity from the BLT. These results suggest that the relationship with anxiety is particularly prominent at the level of subjective interoceptive beliefs in the breathing domain, which are thought to exist at the higher levels of interoceptive space (Critchley and Garfinkel, 2017), followed by metacognitive insight into breathing perception. In comparison, the relationship of trait anxiety to lower level properties such as interoceptive sensitivity (Critchley and Garfinkel, 2017; Garfinkel et al., 2015, 2016a, 2016b) appear to be present but less prominent in the breathing domain. However, it must be noted that quantifying these higher interoceptive levels may be less noisy in comparison with measuring psychophysical properties such as sensitivity, and thus the relationship with anxiety might be most easily detected rather than being inherently stronger.

Although strong relationships were observed between affective qualities and many of our interoceptive breathing 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 breathing-related interoception, as proposed (Critchley and Garfinkel, 2017). The only notably strong cross-modal relationship was found between metacognitive performance and aIns activity, for which greater insight into breathing sensitivity correlated with greater aIns activity during negative prediction errors. This relationship may reflect a previously proposed contribution of error processing toward metacognitive awareness, whereby deviations between actual and predicted bodily inputs are propagated to metacognitive areas via interoceptive brain structures such as the aIns (Stephan et al., 2016).

Novel measures of dynamic interoceptive predictions and prediction errors

Many theories surrounding anxiety have hypothesized an important role of altered predictions regarding upcoming threat (Bach, 2015; Mogg et al., 2000; Simmons et al., 2006) and in particular interoceptive threat (Paulus and Stein, 2010; Paulus et al., 2019) in the aIns (Allen, 2020; Bossaerts, 2010; Carlson et al., 2011; Paulus and Stein, 2006; Tan et al., 2018). Although numerous studies have used inspiratory resistive loads to evoke threatening interoceptive stimuli (Alius et al., 2013; Berner et al., 2018; Faull and Pattinson, 2017; Faull et al., 2016, 2018; Hayen

Please cite this article in press as: Harrison et al., Interoception of breathing and its relationship with anxiety, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.09.045

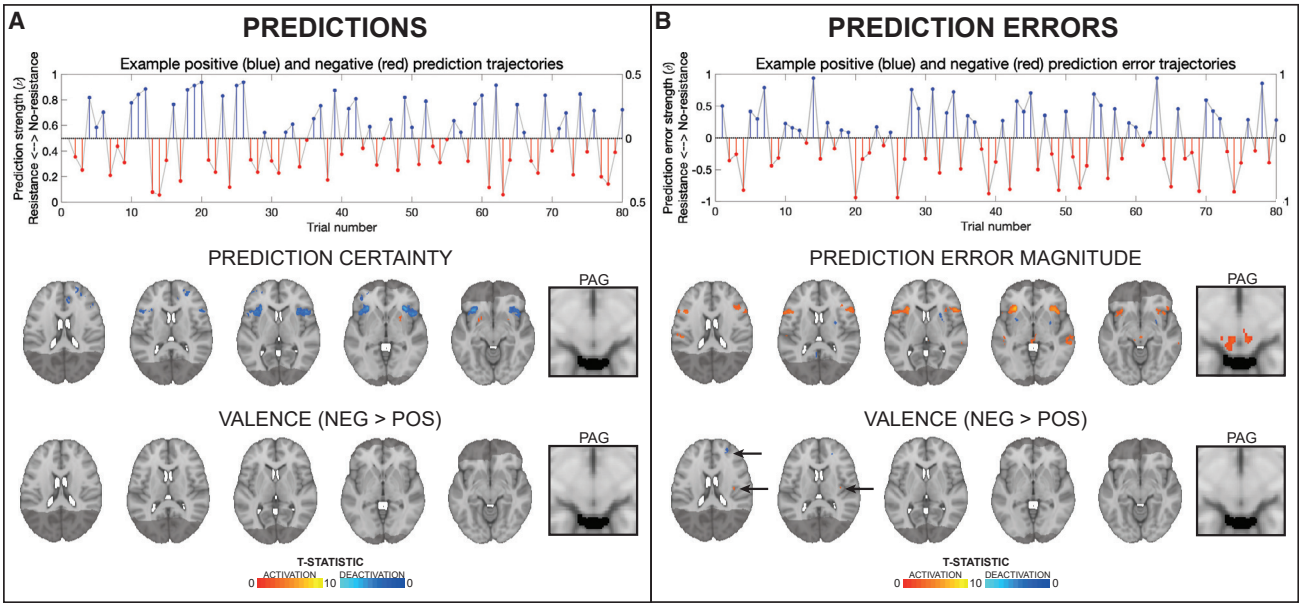

Figure 4. Prediction and prediction-error-related trajectories and brain activity

(A and B) Demonstration of how estimated prediction (A) and prediction error (B) trajectories are encoded as positive (i.e., toward no resistance) and negative (i.e., toward 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 gray 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 zero (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 color-rendered statistical map superimposed on a standard (MNI 1 × 1 × 1 mm) brain. The bright gray region represents the coverage of the coronal-oblique functional scan. Significant regions are displayed with a cluster threshold of $p < 0.05$, family-wise error (FWE) corrected for multiple comparisons across all voxels included in the functional volume. PAG, periaqueductal gray. See also Figures S4–S7.

et al., 2017; von Leupoldt and Dahme, 2005; von Leupoldt et al., 2009; Paulus et al., 2012; Stoeckel et al., 2016; Walter et al., 2020), the BLT approach presented here is, to our knowledge, the first investigation of dynamic (trial-by-trial) brain activity associated with model-based interoceptive predictions and prediction errors for respiration. By fitting an associative learning model to each participant's behavioral 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 (Allen, 2020; Grahl et al., 2018; Paulus and Stein, 2006; Roy et al., 2014; Singer et al., 2009) (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 with 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 hypothesized differences in brain prediction processing (Paulus and Stein, 2006, 2010; Paulus et al., 2019). In comparison, no anxiety group differences or interactions were found in the breathing prediction error domain, contrasting with some previously proposed hypotheses (Barrett and Simmons, 2015; Brewer et al., 2021; Paulus and Stein, 2006, 2010).

Beyond the aIns and independent of anxiety, multiple (and largely consistent) proposals have been made regarding which brain networks might be involved in processing interoceptive predictions and prediction errors (Allen, 2020; Barrett, 2017; Barrett and Simmons, 2015; Craig, 2009; Khalsa et al., 2018; Kleckner et al., 2017; Manjaly and Iglesias, 2020; Marlow et al., 2019; Owens et al., 2018; Paulus et al., 2019; Pezzulo et al., 2015, 2018; Quadt et al., 2018; Seth, 2013; Smith et al., 2017; Stephan et al., 2016). These proposed networks are loosely hierarchical in structure and typically assign metacognitive processes to higher cortical areas (e.g., prefrontal cortex [PFC]), while interoceptive predictions are thought to originate from regions that may serve as an interface between interoceptive and visceromotor function (e.g., aIns and ACC). In these concepts, prediction errors have two different roles: on one hand, they are thought to be sent up the cortical hierarchy of interoceptive regions in order to update predictions in aIns and ACC (Barrett and Simmons, 2015; Pezzulo et al., 2015; Seth et al., 2012); on the other hand, they are thought to determine regulatory signals, sent from visceromotor regions and brainstem structures such as the PAG to the autonomic nervous system and bodily organs (Petzschner et al., 2017; Stephan et al., 2016).

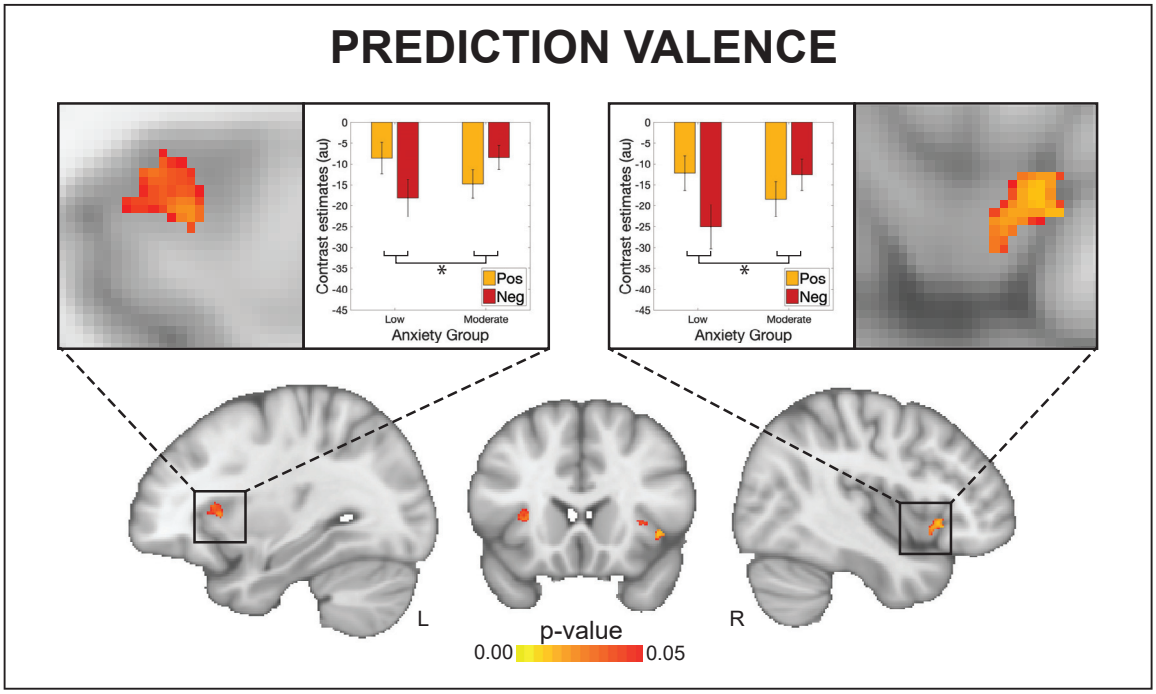

Figure 5. An interaction effect observed between valence (i.e., trials with positive versus negative predictions) and anxiety group (low versus moderate) for activity in the anterior insula related to prediction certainty

The images consist of a color-rendered statistical map superimposed on a standard (MNI 1 × 1 × 1 mm) 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). Bar plots represent mean ± SE for the individual contrast estimates within the significant voxels, plotted separately for each side of the anterior insula. See also Figures S4–S7.

Although 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 breathing network. Although 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 related mainly to prediction uncertainty, where blood-oxygen-level-dependent (BOLD) activity was decreased in proportion to increases in the certainty of predictions (Feldman and Friston, 2010; Friston, 2005). Furthermore, it is notable that the aIns 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 (Allen, 2020; Van den Bergh et al., 2017; Manjaly and Iglesias, 2020; Paulus and Stein, 2006; Paulus et al., 2019; Seth, 2013; Stephan et al., 2016; Walter et al., 2020), given that greater certainty (precision of beliefs) reduces and greater prediction errors increase belief (model) updating (Petzschner et al., 2017).

Our PAG findings are of particular interest. Although the PAG has been previously noted during anticipation of certain breathing resistance stimuli (Faull and Pattinson, 2017; Faull et al., 2016) and has been related to the precision of prior beliefs about placebo-induced reductions in pain intensity (Grahl et al., 2018), here we observed that PAG activity did not appear to be related to the extent of prediction certainty toward upcoming breathing stimuli (Figure 4). Concerning prediction error activity in the PAG, this has previously been demonstrated in relation to pain (Roy et al., 2014); 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 (Stephan et al., 2016).

Finally, overall prediction and prediction error-related activity did not appear to be dissociated between anterior and posterior insula cortices (respectively), as has been previously hypothesized (Allen, 2020; Barrett and Simmons, 2015; Stephan et al., 2016). However, a small valence difference in prediction errors was observed, with negative prediction errors (the unexpected presence of inspiratory resistance stimuli) producing greater activity in the right posterior insula than positive prediction errors (the unexpected absence of inspiratory resistance stimuli; Figure 4). It is therefore possible that the representation of homeostatically relevant inputs in the posterior insula is enhanced for events that may negatively affect homeostasis. However, these results are the first demonstration of brain activity related to dynamic interoceptive prediction and prediction errors; furthermore, the functional images from this study do not have the resolution required for layer-specific identification

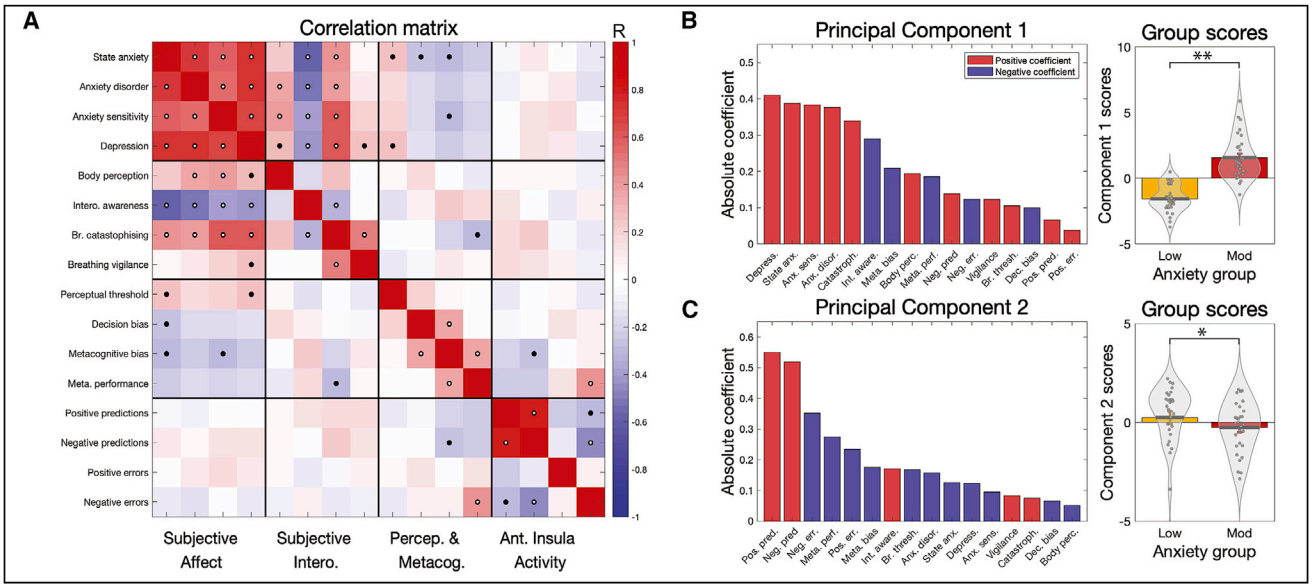

Figure 6. Results from the multi-modal analysis incorporating questionnaires, breathing task data, and peak brain activity in the anterior insula

(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 ± SE, with the distribution of values overlaid in gray. Bar plot code adapted from the CANLAB Toolbox (https://github.com/canlab). See also Table S7.

of prediction and prediction error processing in the insula. Additionally, the represented prediction error-related activity may be specific to the breathing domain within interoceptive processing. This latter caveat of course also applies to the wider results presented here; as only one interoceptive channel (i.e., inspiratory resistances within the breathing domain) was tested, we cannot assume these results would translate to other interoceptive processes (e.g., related to cardiac or gastric states).

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 catastrophizing about breathing symptoms, followed by metacognitive measures related to breathing perception. Furthermore, a novel interaction between trait anxiety and valence was found within the aIns, associated with dynamic prediction certainty (but not prediction errors) of breathing-related interoceptive processing. More generally, this study provides the first empirical demonstration of brain activity associated with dynamic (trial-by-trial) interoceptive learning. These results provide new and comprehensive insights into how anxiety is related to levels of interoceptive processing in the breathing domain and provide evidence of brain activity associated with trial-wise predictions and prediction errors about bodily states in interoceptive breathing networks.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants
- METHOD DETAILS
  - Procedural overview
  - Questionnaires
  - Filter detection task
  - Breathing learning task
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis overview
  - Questionnaire analysis
  - FDT analysis
  - BLT analysis
  - Multi-modal analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2021.09.045.

ACKNOWLEDGMENTS

We would like to thank Professor Klaas Pruessmann and Dr. Lars Kasper for their work establishing and supporting the MRI protocol. O.K.H. (née Faull) was supported by a Marie Skłodowska-Curie Postdoctoral Fellowship from the European Union's Horizon 2020 research and innovation program under the grant agreement 793580. S.F. was supported by UZH Forschungskredit Postdoc (FK-18-046), as well as the ETH Zurich Postdoctoral Fellowship Program and the Marie Skłodowska-Curie Actions for People COFUND program (FEL-49 15-2). F.V. was supported by Fondation Deniker, Fondation pour la Recherche Médicale, and Fondation Bettencourt Schueller. S.J.H. was supported by grant 2017-403 of the Strategic Focus Area "Personalized Health and Related Technologies (PHRT)" of the ETH Domain. K.E.S. was supported by the René and Susanne Braginsky Foundation and the University of Zurich.

AUTHOR CONTRIBUTIONS

O.K.H. and K.E.S. conceptualized and designed the study. O.K.H., F.H.P., S.J.H., and K.E.S. developed the methodology. O.K.H., L.K., S.M., R.L., and F.H. acquired the data. O.K.H. and S.J.H. analyzed the data with input from K.E.S., A.J.H., S.F., S.I., and F.V. K.B. and S.J.H. validated the data and analysis methods. O.K.H. and K.E.S. wrote the manuscript, with edits from all remaining authors.

DECLARATION OF INTERESTS

F.V. has been invited to scientific meetings, consulted and/or served as speaker, and received compensation from Lundbeck, Servier, Recordati, Janssen, Otsuka, LivaNova, and Chiesi. None of these links are related to this work. The authors declare no other competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way.

Received: May 13, 2021
Revised: September 1, 2021
Accepted: September 23, 2021
Published: October 20, 2021

REFERENCES

Adolfi, F., Couto, B., Richter, F., Decety, J., Lopez, J., Sigman, M., Manes, F., and Ibáñez, A. (2017). Convergence of interoception, emotion, and social cognition: a twofold fMRI meta-analysis and lesion approach. Cortex *88*, 124–142.

Alius, M.G., Pané-Farré, C.A., Von Leupoldt, A., and Hamm, A.O. (2013). Induction of dyspnea evokes increased anxiety and maladaptive breathing in individuals with high anxiety sensitivity and suffocation fear. Psychophysiology *50*, 488–497.

Allen, M. (2020). Unravelling the neurobiology of interoceptive inference. Trends Cogn. Sci. *24*, 265–266.

Allen, M., Frank, D., Schwarzkopf, D.S., Fardo, F., Winston, J.S., Hauser, T.U., and Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. eLife *5*, e18103.

Allen, M., Levy, A., Parr, T., and Friston, K.J. (2019). In the body's eye: the computational anatomy of interoceptive inference. bioRxiv. https://doi.org/10.1101/603928.

Andersson, J.L., Jenkinson, M., and Smith, S. (2007). Non-linear registration, aka spatial normalisation (FMRIB technical report TR07JA2) (FMRIB Analysis Group of the University of Oxford).

Bach, D.R. (2015). Anxiety-like behavioural inhibition is normative under environmental threat-reward correlations. PLoS Comput. Biol. *11*, e1004646.

Barrett, L.F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Soc. Cogn. Affect. Neurosci. *12*, 1–23.

Barrett, L.F., and Simmons, W.K. (2015). Interoceptive predictions in the brain. Nat. Rev. Neurosci. *16*, 419–429.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

Berner, L.A., Simmons, A.N., Wierenga, C.E., Bischoff-Grethe, A., Paulus, M.P., Bailer, U.F., Ely, A.V., and Kaye, W.H. (2018). Altered interoceptive activation before, during, and after aversive breathing load in women remitted from anorexia nervosa. Psychol. Med. *48*, 142–154.

Berntson, G.G., and Khalsa, S.S. (2021). Neural circuits of interoception. Trends Neurosci. *44*, 17–28.

Bonaz, B., Lane, R.D., Oshinsky, M.L., Kenny, P.J., Sinha, R., Mayer, E.A., and Critchley, H.D. (2021). Diseases, disorders, and comorbidities of interoception. Trends Neurosci. *44*, 39–51.

Bossaerts, P. (2010). Risk and risk prediction error signals in anterior insula. Brain Struct. Funct. *214*, 645–653.

Brewer, R., Murphy, J., and Bird, G. (2021). Atypical interoception as a common risk factor for psychopathology: a review. Neurosci. Biobehav. Rev. *130*, 470–508.

Carlson, J.M., Greenberg, T., Rubin, D., and Mujica-Parodi, L.R. (2011). Feeling anxious: anticipatory amygdalo-insular response predicts the feeling of anxious anticipation. Soc. Cogn. Affect. Neurosci. *6*, 74–81.

Carrieri-Kohlman, V., Donesky-Cuenco, D., Park, S.K., Mackin, L., Nguyen, H.Q., and Paul, S.M. (2010). Additional evidence for the affective dimension of dyspnea in patients with COPD. Res. Nurs. Health *33*, 4–19.

Chang, C., and Glover, G.H. (2009). Relationship between respiration, end-tidal CO_2, and BOLD signals in resting-state fMRI. Neuroimage *47*, 1381–1393.

Chen, W.G., Schloesser, D., Arensdorf, A.M., Simmons, J.M., Cui, C., Valentino, R., Gnadt, J.W., Nielsen, L., Hillaire-Clarke, C.S., Spruance, V., et al. (2021). The emerging science of interoception: sensing, integrating, interpreting, and regulating signals within the self. Trends Neurosci. *44*, 3–16.

Connor, K.M., and Davidson, J.R.T. (2003). Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress. Anxiety *18*, 76–82.

Craig, A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nat. Rev. Neurosci. *3*, 655–666.

Craig, A.D. (2009). How do you feel—now? The anterior insula and human awareness. Nat. Rev. Neurosci. *10*, 59–70.

Critchley, H.D., and Garfinkel, S.N. (2017). Interoception and emotion. Curr. Opin. Psychol. *17*, 7–14.

Daw, N.D. (2011). Trial-by-trial data analysis using computational models. In Decision Making, Affect, and Learning: Attention and Performance XXIII, M.R. Delgado, E.A. Phelps, and T.W. Robbins, eds. (New York: Oxford University Press), pp. 3–38.

DeVille, D.C., Kerr, K.L., Avery, J.A., Burrows, K., Bodurka, J., Feinstein, J., Khalsa, S.S., Paulus, M.P., and Simmons, W.K. (2018). The neural bases of interoceptive encoding and recall in healthy and depressed adults. Biol. Psychiatry Cogn. Neurosci. Neuroimaging *3*, 546–554.

Domschke, K., Stevens, S., Pfleiderer, B., and Gerlach, A.L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: an overview and integration of neurobiological findings. Clin. Psychol. Rev. *30*, 1–11.

Ewing, D.L., Manassei, M., Gould van Praag, C., Philippides, A.O., Critchley, H.D., and Garfinkel, S.N. (2017). Sleep and the heart: interoceptive differences

linked to poor experiential sleep quality in anxiety and depression. Biol. Psychol. *127*, 163–172.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. Cereb. Cortex *26*, 3508–3526.

Faull, O.K., and Pattinson, K.T. (2017). The cortical connectivity of the periaqueductal gray and the conditioned response to the threat of breathlessness. eLife *6*, e21749–e21767.

Faull, O.K., Jenkinson, M., Clare, S., and Pattinson, K.T.S. (2015). Functional subdivision of the human periaqueductal grey in respiratory control using 7 tesla fMRI. Neuroimage *113*, 356–364.

Faull, O.K., Jenkinson, M., Ezra, M., and Pattinson, K.T.s. (2016). Conditioned respiratory threat in the subdivisions of the human periaqueductal gray. eLife *5*, e12047–e12066.

Faull, O.K., Cox, P.J., and Pattinson, K.T.S. (2018). Cortical processing of breathing perceptions in the athletic brain. Neuroimage *179*, 92–101.

Feldman, H., and Friston, K.J. (2010). Attention, uncertainty, and free-energy. Front. Hum. Neurosci. *4*, 215.

Fleming, S.M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. Neurosci. Conscious. *2017*, nix007.

Frässle, S., Aponte, E.A., Bollmann, S., Brodersen, K.H., Do, C.T., Harrison, O.K., Harrison, S.J., Heinzle, J., Iglesias, S., Kasper, L., et al. (2021). TAPAS: an open-source software package for translational neuromodeling and computational psychiatry. Front. Psychiatry *12*, 680811.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., and Penny, W.D. (2011). Statistical Parametric Mapping: The Analysis of Functional Brain Images (London: Elsevier).

Garfinkel, S.N., Seth, A.K., Barrett, A.B., Suzuki, K., and Critchley, H.D. (2015). Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. Biol. Psychol. *104*, 65–74.

Garfinkel, S.N., Manassei, M.F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H.D., and Engels, M. (2016a). Interoceptive dimensions across cardiac and respiratory axes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20160014-10.

Garfinkel, S.N., Tiley, C., O'Keeffe, S., Harrison, N.A., Seth, A.K., and Critchley, H.D. (2016b). Discrepancies between dimensions of interoception in autism: implications for emotion and anxiety. Biol. Psychol. *114*, 117–126.

Giardino, N.D., Curtis, J.L., Abelson, J.L., King, A.P., Pamp, B., Liberzon, I., and Martinez, F.J. (2010). The impact of panic disorder on interoception and dyspnea reports in chronic obstructive pulmonary disease. Biol. Psychol. *84*, 142–146.

Grahl, A., Onat, S., and Büchel, C. (2018). The periaqueductal gray and Bayesian integration in placebo analgesia. eLife *7*, 1–20.

Griffanti, L., Douaud, G., Bijsterbosch, J., Evangelisti, S., Alfaro-Almagro, F., Glasser, M.F., Duff, E.P., Fitzgibbon, S., Westphal, R., Carone, D., et al. (2017). Hand classification of fMRI ICA noise components. Neuroimage *154*, 188–205.

Gu, X., Hof, P.R., Friston, K.J., and Fan, J. (2013). Anterior insular cortex and emotional awareness. J. Comp. Neurol. *521*, 3371–3388.

Harrison, O.K., Garfinkel, S.N., Marlow, L., Finnegan, S.L., Marino, S., Köchli, L., Allen, M., Finnemann, J., Keur-Huizinga, L., Harrison, S.J., et al. (2021a). The filter detection task for measurement of breathing-related interoception and metacognition. Biol. Psychol. *165*, 108185.

Harrison, S.J., Bianchi, S., Heinzle, J., Stephan, K.E., Iglesias, S., and Kasper, L. (2021b). A Hilbert-based method for processing respiratory timeseries. Neuroimage *230*, 117787.

Harrison, O.K., Marlow, L.L., Finnegan, S.L., Ainsworth, B., and Pattinson, K.T.S. (2021c). Heterogeneity in asthma: dissociating symptoms from mood and their influence on interoception and attention. Biological Psychology *165* (108193). https://doi.org/10.1016/j.biopsycho.2021.108193.

Hayen, A., Herigstad, M., and Pattinson, K.T.S. (2013). Understanding dyspnea as a complex individual experience. Maturitas *76*, 45–50.

Hayen, A., Wanigasekera, V., Faull, O.K., Campbell, S.F., Garry, P.S., Raby, S.J.M., Robertson, J., Webster, R., Wise, R.G., Herigstad, M., and Pattinson, K.T.S. (2017). Opioid suppression of conditioned anticipatory brain responses to breathlessness. Neuroimage *150*, 383–394.

Herigstad, M., Hayen, A., Wiech, K., and Pattinson, K.T.S. (2011). Dyspnoea and the brain. Respir. Med. *105*, 809–817.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E.M., and Stephan, K.E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron *80*, 519–530.

Janssens, T., De Peuter, S., Stans, L., Verleden, G., Troosters, T., Decramer, M., and Van den Bergh, O. (2011). Dyspnea perception in COPD: association between anxiety, dyspnea-related fear, and dyspnea in a pulmonary rehabilitation program. Chest *140*, 618–625.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Med. Image Anal. *5*, 143–156.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825–841.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., and Smith, S.M. (2012). FSL. Neuroimage *62*, 782–790.

Kasper, L., Bollmann, S., Diaconescu, A.O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T.U., Sebold, M., Manjaly, Z.-M., Pruessmann, K.P., and Stephan, K.E. (2017). The PhysIO Toolbox for modeling physiological noise in fMRI data. J. Neurosci. Methods *276*, 56–72.

Khalsa, S.S., Adolphs, R., Cameron, O.G., Critchley, H.D., Davenport, P.W., Feinstein, J.S., Feusner, J.D., Garfinkel, S.N., Lane, R.D., Mehling, W.E., et al. (2018). Interoception and mental health: a roadmap. Biol. Psychiatry Cogn. Neurosci. Neuroimaging *3*, 501–513.

Kleckner, I.R., Wormwood, J.B., Simmons, W.K., Barrett, L.F., and Quigley, K.S. (2015). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. Psychophysiology *52*, 1432–1440.

Kleckner, I.R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W.K., Quigley, K.S., Dickerson, B.C., and Barrett, L.F. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. Nat. Hum. Behav. *1*, 0069.

Krupp, L.B., LaRocca, N.G., Muir-Nash, J., and Steinberg, A.D. (1989). The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch. Neurol. *46*, 1121–1123.

Mallorquí-Bagué, N., Bulbena, A., Pailhez, G., Garfinkel, S.N., and Critchley, H.D. (2016). Mind-body interactions in anxiety and somatic symptoms. Harv. Rev. Psychiatry *24*, 53–60.

Manjaly, Z.-M., and Iglesias, S. (2020). A computational theory of mindfulness based cognitive therapy from the "Bayesian brain" perspective. Front. Psychiatry *11*, 404.

Marlow, L.L., Faull, O.K., Finnegan, S.L., and Pattinson, K.T.S. (2019). Breathlessness and the brain: the role of expectation. Curr. Opin. Support. Palliat. Care *13*, 200–210.

Mathys, C., Daunizeau, J., Friston, K.J., and Stephan, K.E. (2011). A Bayesian foundation for individual learning under uncertainty. Front. Hum. Neurosci. *5*, 39.

Mathys, C.D., Lomakina, E.I., Daunizeau, J., Iglesias, S., Brodersen, K.H., Friston, K.J., and Stephan, K.E. (2014). Uncertainty in perception and the hierarchical Gaussian filter. Front. Hum. Neurosci. *8*, 825.

McCracken, L.M. (1997). "Attention" to pain in persons with chronic pain: A behavioral approach. Behav. Ther. *28*, 271–284.

McNally, R.J., and Eke, M. (1996). Anxiety sensitivity, suffocation fear, and breath-holding duration as predictors of response to carbon dioxide challenge. J. Abnorm. Psychol. *105*, 146–149.

Mehling, W. (2016). Differentiating attention styles and regulatory aspects of self-reported interoceptive sensibility. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20160013-11.

Mehling, W.E., Price, C., Daubenmier, J.J., Acree, M., Bartmess, E., and Stewart, A. (2012). The Multidimensional Assessment of Interoceptive Awareness (MAIA). PLoS ONE *7*, e48230.

Mogg, K., McNamara, J., Powys, M., Rawlinson, H., Seiffer, A., and Bradley, B.P. (2000). Selective attention to threat: a test of two cognitive models of anxiety. Cogn. Emotion *14*, 375–399.

Murphy, J., Catmur, C., and Bird, G. (2019). Classifying individual differences in interoception: implications for the measurement of interoceptive awareness. Psychon. Bull. Rev. *26*, 1467–1471.

O'Reilly, J.X., Jbabdi, S., and Behrens, T.E. (2012). How can a Bayesian approach inform neuroscience? Eur. J. Neurosci. *35*, 1169–1179.

Owens, A.P., Allen, M., Ondobaka, S., and Friston, K.J. (2018). Interoceptive inference: from computational neuroscience to clinic. Neurosci. Biobehav. Rev. *90*, 174–183.

Parshall, M.B., Schwartzstein, R.M., Adams, L., Banzett, R.B., Manning, H.L., Bourbeau, J., Calverley, P.M., Gift, A.G., Harver, A., Lareau, S.C., et al.; American Thoracic Society Committee on Dyspnea (2012). An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. Am. J. Respir. Crit. Care Med. *185*, 435–452.

Paulus, M.P. (2013). The breathing conundrum-interoceptive sensitivity and anxiety. Depress. Anxiety *30*, 315–320.

Paulus, M.P., and Stein, M.B. (2006). An insular view of anxiety. Biol. Psychiatry *60*, 383–387.

Paulus, M.P., and Stein, M.B. (2010). Interoception in anxiety and depression. Brain Struct. Funct. *214*, 451–463.

Paulus, M.P., Flagan, T., Simmons, A.N., Gillis, K., Kotturi, S., Thom, N., Johnson, D.C., Van Orden, K.F., Davenport, P.W., and Swain, J.L. (2012). Subjecting elite athletes to inspiratory breathing load reveals behavioral and neural signatures of optimal performers in extreme environments. PLoS ONE *7*, e29394.

Paulus, M.P., Feinstein, J.S., and Khalsa, S.S. (2019). An active inference approach to interoceptive psychopathology. Annu. Rev. Clin. Psychol. *15*, 97–122.

Petzschner, F.H., Weber, L.A.E., Gard, T., and Stephan, K.E. (2017). Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis. Biol. Psychiatry *82*, 421–430.

Petzschner, F.H., Weber, L.A., Wellstein, K.V., Paolini, G., Do, C.T., and Stephan, K.E. (2019). Focus of attention modulates the heartbeat evoked potential. Neuroimage *186*, 595–606.

Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. Prog. Neurobiol. *134*, 17–35.

Pezzulo, G., Rigoli, F., and Friston, K.J. (2018). Hierarchical active inference: a theory of motivated control. Trends Cogn. Sci. *22*, 294–306.

Porges, S.W. (1995). Orienting in a defensive world: mammalian modifications of our evolutionary heritage. A polyvagal theory. Psychophysiology *32*, 301–318.

Quadt, L., Critchley, H.D., and Garfinkel, S.N. (2018). The neurobiology of interoception in health and disease. Ann. N Y Acad. Sci. *1428*, 112–128.

Quigley, K.S., Kanoski, S., Grill, W.M., Barrett, L.F., and Tsakiris, M. (2021). Functions of interoception: from energy regulation to experience of the self. Trends Neurosci. *44*, 29–38.

Radloff, L.S. (1977). The CES-D scale a self-report depression scale for research in the general population. Appl. Psychol. Meas. *1*, 385–401.

Rescorla, R.A., Wagner, A.R., Black, A.H., and Prokasy, W.F. (1972). Classical Conditioning II: Current Research and Theory (New York: Appleton-Century-Crofts).

Rieger, S.W., Stephan, K.E., and Harrison, O.K. (2020). Remote, automated, and MRI-compatible administration of interoceptive inspiratory resistive loading. Front. Hum. Neurosci. *14*, 161.

Rigoux, L., Stephan, K.E., Friston, K.J., and Daunizeau, J. (2014). Bayesian model selection for group studies—revisited. Neuroimage *84*, 971–985.

Rouault, M., Seow, T., Gillan, C.M., and Fleming, S.M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biol. Psychiatry *84*, 443–451.

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G.E., and Wager, T.D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. Nat. Neurosci. *17*, 1607–1612.

Schwartzstein, R.M., Manning, H.L., Weiss, J.W., and Weinberger, S.E. (1990). Dyspnea: a sensory experience. Lung *168*, 185–199.

Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K., and Zhang, J.X. (1997). The assessment of optimistic self-beliefs: comparison of the German, Spanish, and Chinese versions of the General Self-Efficacy Scale. Appl. Psychol. *46*, 69–88.

Seth, A.K. (2013). Interoceptive inference, emotion, and the embodied self. Trends Cogn. Sci. *17*, 565–573.

Seth, A.K., Suzuki, K., and Critchley, H.D. (2012). An interoceptive predictive coding model of conscious presence. Front. Psychol. *2*, 395.

Simmons, A., Strigo, I., Matthews, S.C., Paulus, M.P., and Stein, M.B. (2006). Anticipation of aversive visual stimuli is associated with increased insula activation in anxiety-prone subjects. Biol. Psychiatry *60*, 402–409.

Singer, T., Critchley, H.D., and Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. Trends Cogn. Sci. *13*, 334–340.

Smith, S.M. (2002). Fast robust automated brain extraction. Hum. Brain Mapp. *17*, 143–155.

Smith, S.M., and Nichols, T.E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage *44*, 83–98.

Smith, R., Thayer, J.F., Khalsa, S.S., and Lane, R.D. (2017). The hierarchical basis of neurovisceral integration. Neurosci. Biobehav. Rev. *75*, 274–296.

Smith, R., Kuplicki, R., Feinstein, J., Forthman, K.L., Stewart, J.L., Paulus, M.P., and Khalsa, S.S.; Tulsa 1000 investigators (2020). A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. PLoS Comput. Biol. *16*, e1008484.

Smoller, J.W., Pollack, M.H., Otto, M.W., Rosenbaum, J.F., and Kradin, R.L. (1996). Panic anxiety, dyspnea, and respiratory disease. Theoretical and clinical considerations. Am. J. Respir. Crit. Care Med. *154*, 6–17.

Spielberger, C.D., Gorsuch, R.L., and Lushene, R.E. (1970). State-Trait Anxiety (STAI) Manual (Palo Alto, CA: Consulting Psychologists Press).

Spitzer, R.L., Kroenke, K., Williams, J.B.W., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. *166*, 1092–1097.

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. Behav. Res. Methods Instrum. Comput. *31*, 137–149.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. Neuroimage *46*, 1004–1017.

Stephan, K.E., Manjaly, Z.M., Mathys, C.D., Weber, L.A.E., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S.M., Haker, H., Seth, A.K., and Petzschner, F.H. (2016). Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. Front. Hum. Neurosci. *10*, 550.

Stoeckel, M.C., Esser, R.W., Gamer, M., Büchel, C., and von Leupoldt, A. (2016). Brain responses during the anticipation of dyspnea. Neural Plast. *2016*, 6434987.

Sullivan, M.J.L., Bishop, S.R., and Pivik, J. (1995). The Pain Catastrophizing Scale: development and validation. Psychol. Assess. *7*, 524–532.

Tan, Y., Wei, D., Zhang, M., Yang, J., Jelinčić, V., and Qiu, J. (2018). The role of mid-insula in the relationship between cardiac interoceptive attention and anxiety: evidence from an fMRI study. Sci. Rep. *8*, 17280.

Taylor, S., Zvolensky, M.J., Cox, B.J., Deacon, B., Heimberg, R.G., Ledley, D.R., Abramowitz, J.S., Holaway, R.M., Sandin, B., Stewart, S.H., et al. (2007). Robust dimensions of anxiety sensitivity: development and initial validation of the Anxiety Sensitivity Index-3. Psychol. Assess. *19*, 176–188.

Tiller, J., Pain, M., and Biddle, N. (1987). Anxiety disorder and perception of inspiratory resistive loads. Chest *91*, 547–551.

Tsakiris, M., and Critchley, H. (2016). Interoception beyond homeostasis: affect, cognition and mental health. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20160002–20160006.

Tschantz, A., Barca, L., Maisto, D., Buckley, C.L., Seth, A.K., and Pezzulo, G. (2021). Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference. bioRxiv. https://doi.org/10.1101/2021.02.16.431365.

Van den Bergh, O., Witthöft, M., Petersen, S., and Brown, R.J. (2017). Symptoms and the body: taking the inferential leap. Neurosci. Biobehav. Rev. *74* (Pt A), 185–203.

von Leupoldt, A., and Dahme, B. (2005). Differentiation between the sensory and affective dimension of dyspnea during resistive load breathing in normal subjects. Chest *128*, 3345–3349.

von Leupoldt, A., Sommer, T., Kegat, S., Eippert, F., Baumann, H.J., Klose, H., Dahme, B., and Büchel, C. (2009). Down-regulation of insular cortex responses to dyspnea and pain in asthma. Am. J. Respir. Crit. Care Med. *180*, 232–238.

Walter, H., Kausch, A., Dorfschmidt, L., Waller, L., Chinichian, N., Veer, I., Hilbert, K., Lüken, U., Paulus, M.P., Goschke, T., and Kruschwitz, J.D. (2020). 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.

Wang, X., Wu, Q., Egan, L., Gu, X., Liu, P., Gu, H., Yang, Y., Luo, J., Wu, Y., Gao, Z., and Fan, J. (2019). Anterior insular cortex plays a critical role in interoceptive attention. eLife *8*, e42265.

Watson, D., Clark, L.A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. *54*, 1063–1070.

Wilson, R.C., and Collins, A.G. (2019). Ten simple rules for the computational modeling of behavioral data. eLife *8*, e49547.

Woods, S.W., Charney, D.S., Loke, J., Goodman, W.K., Redmond, D.E., Jr., and Heninger, G.R. (1986). 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. Arch. Gen. Psychiatry *43*, 900–909.

Woolrich, M.W., Ripley, B.D., Brady, M., and Smith, S.M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage *14*, 1370–1386.

Please cite this article in press as: Harrison et al., Interoception of breathing and its relationship with anxiety, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.09.045

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Summary maps	ETH Research Collection	https://doi.org/10.3929/ethz-b-000503396
Data	TNU data sharing	email: tnu-datasharing@biomed.ee.ethz.ch
Software and algorithms		
MATLAB v2017b	Mathworks	https://www.mathworks.com
FSL v6.0.1	Jenkinson et al., 2012	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
SPM12	Friston et al., 2011	https://www.fil.ion.ucl.ac.uk/spm/software/
Psychtoolbox3	Psychtoolbox	https://psychtoolbox.org/
Analysis code	Gitlab	https://gitlab.ethz.ch/tnu/code/harrison_breathing_anxiety_code; https://doi.org/10.5281/zenodo.5523258
Other		
Analysis plan	Gitlab	https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Olivia Harrison (née Faull; faull@biomed.ee.ethz.ch).

Materials availability

This study neither used any reagent nor generated new materials.

Data and code availability

The raw data reported in this study cannot be deposited in a public repository due to GDPR requirements. To request access, contact tnu-datasharing@biomed.ee.ethz.ch with a description of the request (n = 58 participants with full datasets are available with permission for data sharing). In addition, summary fMRI statistics (thresholded and unthresholded group maps) derived from these data – and a description of the data that can be made available upon request – have been deposited in the ETH Research Collection (https://www.research-collection.ethz.ch), and are publicly available as of the date of publication (https://doi.org/10.3929/ethz-b-000503396).

All original code has been deposited at https://gitlab.ethz.ch/tnu/code/harrison_breathing_anxiety_code, and is publicly available as of the date of publication (https://doi.org/10.5281/zenodo.5523258).

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

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 (Spielberger et al., 1970); 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. Study numbers were based on a power calculation for a two-sided two-sample t test with an α-level of 5% (moderate effect size d = 0.5), where a power of 90% is achieved with 30 participants in each group. Behavioral data (not used in any other analyses) from an additional 8 participants served to determine model priors, with four participants (two females and two males) from each of the low and moderate anxiety groups. In this way, prior values could be drawn from a comparable group of participants as the

main study sample. All participants signed a written, informed consent, and the study was approved by the Cantonal Ethics Committee Zurich (Ethics approval BASEC-No. 2017-02330).

Behavioral data for an additional group of 15 individuals (12 female) were collected for model validation purposes. These participants were not pre-selected based on anxiety values (mean ± std trait anxiety = 38.9 ± 12.5), but were screened for non-smoking status and no history of major somatic or psychological conditions. Participants were aged (mean ± std) 23.1 ± 5.6 years. All participants signed a written, informed consent, and the study was approved by the New Zealand Health and Disability Ethics Committee (HDEC) (Ethics approval 20/CEN/168).

METHOD DETAILS

Procedural overview

Each participant completed three tasks over two testing sessions: a behavioral 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). The length of time between testing sessions (mean ± std) was 4 ± 3 days for all participants. Participants in the validation group completed the BLT in a behavioral session only.

Questionnaires

The main questionnaire set employed was designed to first capture subjective affective measures, and second both general and breathing-specific subjective interoceptive beliefs. The assignment of participants to groups was based on the Spielberger Trait Anxiety Inventory (STAI-T) (Spielberger et al., 1970). Affective qualities that were additionally assessed included state anxiety (Spielberger State Anxiety Inventory; STAI-S (Spielberger et al., 1970)), symptoms that are part of anxiety disorder (Generalized Anxiety Disorder Questionnaire; GAD-7 (Spitzer et al., 2006)), anxiety sensitivity (anxiety regarding the symptoms of anxiety; Anxiety Sensitivity Index; ASI-3 (Taylor et al., 2007)), and symptoms of depression (Centre for Epidemiologic Studies Depression Scale; CES-D (Radloff, 1977)). To obtain self-reports of body awareness we used the Body Perception Questionnaire (BPQ) (Porges, 1995), while the Multidimensional Assessment of Interoceptive Awareness Questionnaire (MAIA) (Mehling et al., 2012) was used to measure positive and 'mindful' attention toward body symptoms. We also measured breathing-related catastrophising using the Pain Catastrophising Scale (PCS-B) (Sullivan et al., 1995), and breathing-related vigilance using the Pain Vigilance Awareness Questionnaire (PVQ-B) (McCracken, 1997) (in both questionnaires, the word 'breathless' or 'breathlessness' 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) (Watson et al., 1988), Connor-Davidson Resilience Scale (Connor and Davidson, 2003), General Self-Efficacy Scale (Schwarzer et al., 1997), Fatigue Severity Scale (FSS) (Krupp et al., 1989). The STAI-T and CES-D were completed online as part of the pre-screening process; all other questionnaires were completed in the behavioral session at the laboratory.

Filter detection task

To systematically test properties of breathing perception and related metacognition, we utilized a perceptual threshold breathing task (the Filter Detection Task; FDT), described in detail elsewhere (Harrison et al., 2021a). 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 utilized to alter task difficulty until participants were between 60%–85% accuracy (Harrison et al., 2021a), and 60 trials were completed at the corresponding level of filter load once the threshold level had been identified (using a 'constant' staircase procedure, as described by (Harrison et al., [2021a]). Respiratory threshold detection (Garfinkel et al., 2016b), metacognitive bias (Rouault et al., 2018) and interoceptive metacognitive performance (Harrison et al., 2021c) have previously been linked to anxiety symptomology.

Breathing learning task

To measure behavior 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 an 80% or 20% chance of a subsequent inspiratory resistive load. Participants were explicitly told the probabilities that were being used, as well as that the cues were paired together – if one cue indicated an 80% chance of resistance then the other must indicate a 20% chance. Participants were also told that the cues could only ***swap*** their contingencies throughout the task, and could not act independently of each other.

The number of trials was limited to 80 to ensure feasibility for participants. The number and structure of the cue contingency swaps (n = 4) within the 80 trials was chosen as simulated data demonstrated both parameter recovery and model identifiability of the three candidate learning models (see Figure S2 and 'Quantification and Statistical Analysis' details below).

The visual information for the task was presented through the VisualStim system (Resonance Technology, Northridge, CA, US). As outlined in Figure 3, participants were required to predict (via button press) whether they would experience a breathing resistance following the presentation of one of the cues. The visual cues were counter-balanced such that each was first matched with an 80% chance of resistance for half the participants, as well as counter-balancing of whether the answer 'yes' to the prediction question was presented on the left or right of the screen. Following this prediction and a short (2.5 s) pause, a circle appeared on the screen to indicate the stimulus period (5 s), 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 7-9 s were pseudo-randomized 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 previous work (Rieger et al., 2020)). 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 3). 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 analog 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 Supplemental information (Figures S4–S7).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis overview

All analyses and hypotheses were pre-specified in an analysis plan (https://gitlab.ethz.ch/tnu/analysis-plans/harrison_breathing_anxiety). Each measure within each task modality was first compared between groups, and multiple comparison correction was applied within modalities. Pre-selection and comparisons across low or moderate anxiety scores allowed us to control for important factors such as age and sex, with equal numbers of men and women recruited into each group. Finally, a cross-modal analysis was performed on the key measures from each task. As the trait anxiety measure that was used to recruit participants into the separate groups was not included in this final analysis, we employed a correlation-based approach as the full spectrum of scores for each task were available.

Questionnaire 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 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 Figure S1.

FDT 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 analyzed using the hierarchical HMeta-d statistical model (Fleming, 2017), as previously described (Harrison et al., 2021a). This model first utilizes signal detection theory (Stanislaw and Todorov, 1999) to provide single subject parameter estimates for task difficulty (d'; not analyzed 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 normalizing by single subject values for d'). Finally, metacognitive bias was calculated as the average confidence scores across all analyzed trials.

The four FDT parameters that were used for group comparison analyses were: sensitivity (filter number at perceptual threshold), decision bias (*c*), metacognitive bias (average confidence scores over threshold trials) and metacognitive performance (Mratio). These data 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. Importantly, as the single subject Mratio parameters were fit using a whole-group hierarchical model (with equal group numbers), standard statistical comparisons between groups for these parameters can also be performed. Hypothesized group differences were based on previous findings, where respiratory threshold detection level was hypothesized to be higher (Garfinkel et al., 2016a; Tiller et al., 1987), perceptual decisions biased toward 'yes' (or deciding the resistance was present; denoted by more negative values for *c*), metacognitive bias to be lower (Rouault et al., 2018) and interoceptive metacognitive performance to be lower (Harrison et al., 2021c) with greater anxiety. Bonferroni correction for the four tests was applied, requiring $p < 0.013$ for a corrected significant group difference.

BLT analysis

Model space

For the trial-by-trial analysis of behavioral 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 HGF 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 (Mathys et al., 2011, 2014) (version 5.3) from the open-source TAPAS software (Frässle et al., 2021) (https://www.translationalneuromodeling.org/tapas/).

Prior selection and simulation analyses

Prior means and variances were determined using the distribution of maximum likelihood estimates fit across a holdout dataset (consisting of 8 participants who were distinct from the participants of our study). Individual fits and estimated prior densities of the free parameters are given in Table S2, as well as values for all remaining parameters of the three models. By adopting this procedure, prior densities were in a regime of the parameter space that is representative of the actual behavioral 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 three models considered in our model space (each with one parameter free: α in RW, ω_2 in HGF2, κ_2 in HGF3), we assessed both parameter recovery and model identifiability for each (Wilson and Collins, 2019). 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, maximum *a posteriori* (MAP) parameter estimates were obtained using the Brayden-Fletcher-Goldfarb-Shanno algorithm, as implemented in the HGF toolbox, to fit the synthetic datasets. 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 (Figure S2). 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 log model evidence (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 (Figure S2), 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 Table S3.

Model comparison and selection

Following MAP estimates of the 60 empirical datasets using each of the candidate models, the models were formally compared using random effects Bayesian model selection (BMS) as implemented in SPM12 (Friston et al., 2011; Rigoux et al., 2014; Stephan et al., 2009). BMS utilizes the LME to determine the most likely among a set of competing hypotheses (i.e., models) that may have generated observed data, and is robust to outliers (Stephan et al., 2009). 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 Figure S2 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 (Rescorla et al., 1972):

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

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 (Mathys et al., 2014). 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:

$$p(y_{(k)}|v_{(k)},\zeta) = \left(\frac{v_{(k)}^{\zeta}}{v_{(k)}^{\zeta}+(1-v_{(k)})^{\zeta}}\right)^{y_{(k)}}\left(\frac{(1-v_{(k)})^{\zeta}}{v_{(k)}^{\zeta}+(1-v_{(k)})^{\zeta}}\right)^{1-y_{(k)}} \quad \text{(Equation 2)}$$

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 (Mathys et al., 2014), given the trajectories of trial-wise cues and responses (see Figure 3).

In our paradigm, trial-wise outcomes are categorical (resistance versus 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 (Iglesias et al., 2013). 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).

Comparison of fitted model parameters

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 compared following tests for data normality. 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 Table S6.

Model validation

Following random effects Bayesian model selection (BMS (Rigoux et al., 2014; Stephan et al., 2009)), 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 a 'null model' (i.e., due to chance) (Daw, 2011) using the likelihood ratio test (*lratiotest* function) provided in MATLAB. The final behavioral and brain imaging analyses presented here were run without two subjects in which non-significant ($p > 0.05$) differences to randomness were encountered. To demonstrate the extent to which the chosen model (the Rescorla Wagner) captured important aspects of participant performance, the proportion of incorrect responses across participants at each trial was compared to the mean prediction error trajectory (Figure S3B).

To further validate the application of the chosen model to the BLT, we compared the fitted model trajectory to unseen behavioral data from an additional 15 participants. For qualitative assessment, the mean prediction error trajectory from the original dataset was compared to the proportion of incorrect responses across these held-out participants at each trial (Figure S3C). For quantitative assessment, a logistic regression was conducted to assess whether the model prediction trajectory from the original data was able to significantly explain the prediction decisions made by the 15 participants in the validation sample.

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:

$$v_{(k)}^{stim} \underline{\underline{\text{def}}} \begin{cases} v_{(k)}, |\textit{if cue type} = 1 \\ 1 - v_{(k)}, |\textit{if cue type} = 2 \end{cases} \quad \text{(Equation 3)}$$

$$\delta_{(k)}^{stim} \underline{\underline{\text{def}}} \begin{cases} \delta_{(k)}, |\textit{if cue type} = 1 \\ -\delta_{(k)}, |\textit{if cue type} = 2 \end{cases} \quad \text{(Equation 4)}$$

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 Figure S4 for details).

Second, 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 (Paulus and Stein, 2010; Paulus et al., 2019) 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):

$$\text{If } v^{stim}_{(k)} > 0.5\ v^{pos}_{(k)} \underline{\underline{\text{def}}}\ v^{stim}_{(k)} - 0.5 \quad \text{(Equation 5)}$$

$$\text{If } v^{stim}_{(k)} < 0.5\ v^{neg}_{(k)} \underline{\underline{\text{def}}} 0.5 - v^{stim}_{(k)} \quad \text{(Equation 6)}$$

Here, both $v^{pos}_{(k)}$ and $v^{neg}_{(k)}$ 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):

$$\text{If } \delta_{(k)} > 0\ \delta^{pos}_{(k)} \underline{\underline{\text{def}}} \delta_{(k)} \quad \text{(Equation 7)}$$

$$\text{If } \delta_{(k)} < 0\ \delta^{neg}_{(k)} \underline{\underline{\text{def}}} - \delta_{(k)} \quad \text{(Equation 8)}$$

Here, both $\delta^{pos}_{(k)}$ and $\delta^{neg}_{(k)}$ 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 modeling two separate trajectories for resistance and no resistance outcomes (see above and (Iglesias et al., 2013)).

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 (Rieger et al., 2020)). 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 (Faull and Pattinson, 2017; Faull et al., 2015, 2016, 2018).

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 10 s and 20 s for the overshoot and undershoot, respectively (Chang and Glover, 2009). 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 (Kasper et al., 2017). Four respiratory volume per unit time (RVT) regressors (delays: −5, 0, 5, 10) were created using the Hilbert-transform estimator in PhysIO (Harrison et al., 2021b), and convolution with a respiratory response function (Kasper et al., 2017).

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 centered over the insula and midbrain structures. The FOV comprised 32 slices (sequence parameters: TE 30ms; TR 2.3 s; 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 9 s). 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) (Jenkinson et al., 2012) and SPM12 (Statistical Parametric Mapping software, London, UK) (Friston et al., 2011) 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 (Jenkinson and Smith, 2001)), removal of non-brain structures (BET (Smith, 2002)), and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting; 100 s cut-off period)(Woolrich et al., 2001). Independent component analysis (ICA) was used to identify noise due to motion, scanner and cerebrospinal fluid artifacts (Griffanti et al., 2017), 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 (Jenkinson et al., 2002); 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 (Andersson et al., 2007) 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 Supplemental information (Figures S4–S6). 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 3A and Figure S5):

1) A 'Cue' regressor (80 repeats), with onsets and durations (2.5 s) 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 0.5 s 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 0.5 s 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 3), 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 3), 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 versus negative) for both predictions and prediction errors.

Group fMRI analysis

First, 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 $p < 0.001$. Second, for our region of interest (ROI) analysis, we used FSL's non-parametric threshold-free cluster enhancement (Smith and Nichols, 2009) within a combined mask of the anterior insula and periaqueductal gray (PAG), as pre-specified in our analysis plan (section 7.5.5). 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 (Berner et al., 2018; Faull and Pattinson, 2017; Faull et al., 2016, 2018; Paulus et al., 2012; Walter et al., 2020) as well as prediction errors (Roy et al., 2014) and precision (Grahl et al., 2018) toward 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 (Fan et al., 2016) (bilateral ventral and dorsal anterior insula regions), and the PAG incorporated an anatomically-defined mask that has been used in previous fMRI publications (Faull and Pattinson, 2017; Faull et al., 2016).

Multi-modal analysis

Multi-modal data

The different task modalities were then combined into a multi-modal analysis to assess both the relationships between and shared variance among 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 (Spielberger et al., 1970), GAD-7 (Spitzer et al., 2006), ASI-3 (Taylor et al., 2007) and CES-D (Radloff, 1977));
2) The four interoceptive questionnaires (BPQ (Porges, 1995), MAIA (Mehling et al., 2012), PCS-B (Sullivan et al., 1995) and PVQ-B (McCracken, 1997));
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, centered on each participant's maximal contrast estimate within a Brainnetome atlas mask of the anterior insula (Fan et al., 2016), using the first eigenvariate of the data.

Multi-modal correlations and shared variance

A Pearson's correlation matrix of all 16 included measures was calculated in order to visualize 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). A supplementary non-parametric correlation matrix was additionally calculated using Spearman's rho values, and these results are presented in Table S7.

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 normalization using z-scoring within each variable. PCA is an orthogonal linear transformation that transforms the $n \times m$ data matrix $\mathbf{X}$ (participants × measures) into a new matrix $\mathbf{P}$, 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 $\mathbf{X}$ to map each row (participant) vector $\boldsymbol{x}_i$ of $\mathbf{X}$ onto a new vector of principal component scores $\boldsymbol{t}_i$, given by:

$$t_{k(i)} = \boldsymbol{x}_i \times \boldsymbol{w}_k \text{ for } i = 1, \ldots, n; k = 1, \ldots, m \quad \text{(Equation 9)}$$

where $t_{k(i)}$ is the score for each participant i within each component k. The number of significant components were then 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; https://doi.org/10.5281/zenodo.5523258).