An active inference model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders

Ryan Smith[1], Rayus Kuplicki[1], Justin Feinstein[1,2], Katherine L. Forthman, Jennifer L. Stewart[1,2], Martin P. Paulus[1,2], Tulsa 1000 investigators[†], and Sahib S. Khalsa[1,2]

[1]Laureate Institute for Brain Research, Tulsa, OK, USA

[2]Oxley College of Health Sciences, The University of Tulsa, Tulsa, OK.

[†]The Tulsa 1000 Investigators include the following contributors: Robin L. Aupperle, Ph.D.; Jerzy Bodurka, Ph.D.; Jonathan B. Savitz, Ph.D.; and Teresa A. Victor, Ph.D.

Corresponding author:

Ryan Smith
Laureate Institute for Brain Research
6655 S Yale Ave, Tulsa, OK 74136, USA
Email: rsmith@laureateinstitute.org

Abstract

Recent neurocomputational theories have hypothesized that abnormalities in prior expectations and/or the precision-weighting of afferent interoceptive signals (i.e. the degree to which afferent bodily signals contribute to interoceptive perceptual inference), may facilitate the transdiagnostic emergence of psychopathology. Specifically, it has been suggested that, in certain psychiatric disorders, interoceptive processing mechanisms either over-weight prior expectations or under-weight signals from the viscera (or both), leading to a failure to accurately update beliefs about the body (potentially resulting in visceral dysregulation among other maladaptive phenomena). However, this has not been directly tested empirically. To evaluate the potential roles of prior expectations and interoceptive precision in this context, we fitted behavior in a transdiagnostic patient population on an interoceptive awareness (heartbeat tapping) task to a Bayesian computational model based on active inference (i.e., in which interoceptive inference can be cast as prediction-error minimization). Modeling revealed that, during an interoceptive perturbation condition (inspiratory breath-holding during heartbeat tapping), healthy individuals (N=52) assigned greater precision to ascending cardiac signals than individuals with symptoms of anxiety (N=15), depression (N=69), co-morbid depression/anxiety (N=153), substance use disorders (N=131), and eating disorders (N=14) – who failed to increase their precision estimates from resting levels. In contrast, no differences were found in prior expectations. These results provide the first empirical computational modeling evidence of a selective dysfunction in adaptive interoceptive processing in psychiatric conditions, and lay the groundwork for future studies examining how reduced interoceptive precision influences body regulation and interoceptively-guided decision-making.

Author Summary

Interoception is the process by which the nervous system senses the internal state of the body. It provides the brain with important information to adaptively guide the regulation of both internal body states and behavior. Interoceptive dysfunction is thought to play a role in multiple psychiatric disorders. Theoretical models propose that the computational mechanisms of interoceptive dysfunction are caused by overly precise prior expectations about body states ("hyperprecise priors") or underestimates of the reliability of the information carried by ascending signals from the body ("low sensory precision"). Our empirical approach tested for evidence of these mechanisms across several psychiatric disorders, using a computational model of perception during performance of a heartbeat perception task. We found evidence of low sensory precision within individuals with anxiety, depression, eating disorders, and/or substance use disorders, relative to healthy individuals. This difference occurred only during a breath-holding condition designed to enhance heartbeat signals. We did not find evidence for hyperprecise priors in the patient groups. The data from this study support the argument for computational mechanisms of interoceptive dysfunction across several psychiatric disorders, and suggests that these conditions may be characterized by an inability to adjust sensory precision when signals from the body change.

Keywords: Interoception; Active Inference; Depression; Anxiety; Substance Use; Eating Disorders; Precision; Prior expectations; Bayesian Perception; Computational Modeling

Introduction

Interoception plays an important role in a number of psychiatric disorders. Interoceptive dysfunction has been observed in depression, anxiety, eating, and substance use disorders, among others (reviewed in (1)). For example, depressed patients exhibit reduced accuracy when asked to count their own heartbeats (2-4), and counting accuracy is negatively correlated with depressive symptoms (5). Several studies have also shown heightened interoceptive sensations in panic disorder under high arousal states (reviewed in (6)). Other studies have reported evidence of blunted neural responses during interoceptive processing in substance users (7), and that patients with eating disorders show stronger expectation effects on interoception during modulations of arousal (8). However, a well-established, principled approach for identifying a generative explanatory model of interoceptive dysfunctions in psychiatric disorders has not yet emerged.

In recent years, a growing body of theoretical work within neuroscience and psychiatry has begun to highlight plausible neurocomputational accounts of interoceptive processing (9-15), as well as accounts of exteroceptive (16, 17), cognitive (18-26), emotional (27, 28), and motor control (29-31) functions that plausibly interact with interoception. In some cases, empirical studies have also found support for computational models and computational abnormalities in relation to specific psychiatric symptoms/disorders (26, 32-34). With respect to sensory processing, which has focused to a large degree on neural models of approximate probabilistic (Bayesian) inference, empirical results in this area have come almost exclusively from the exteroceptive domain. And while recent work has begun to test qualitative predictions of Bayesian models of interoception (35), no study to our knowledge has yet explicitly fit

computational models to brain or behavioral responses during experimental modulations of interoception in individuals with psychiatric disorders.

Drawing on this theoretical literature, we and others (9, 13-15, 36-38) have previously proposed that symptoms of multiple psychiatric conditions could be explained by faulty interoceptive computational processes (and associated visceromotor control processes), due to an inappropriate weighting of prior expectations and sensory evidence. Although these models of computational dysfunction in interoception could have broad explanatory power, the putative mechanisms they propose have yet to be empirically tested using formal computational models. In this paper, we use a formal Bayesian computational model of perception to examine whether there is evidence for the transdiagnostic dysfunction of interoceptive processing suggested in these previous proposals within the cardiovascular system.

Historically, the most common empirical paradigms for studying interoception have focused on the perception of cardiac signals. Common heartbeat perception tasks include those involving heartbeat counting (39), heartbeat tapping (40), and other means of assessing cardiac interoception (41-45). While each approach has certain limitations (46-50), one consistent finding is that cardiac interoceptive awareness is quite poor in the majority of participants tested during resting conditions – where only roughly 35% of individuals appear to accurately perceive their own heartbeats (6). In contrast, when visceral states are perturbed, cardiac perception becomes more accurate. This is particularly true for interoceptive accuracy under conditions of heightened cardiorespiratory arousal (51-53).

Active inference models, and Bayesian predictive processing models more generally, have been the main computational framework within which interoceptive dysfunction has been discussed (13, 38, 54-56). Several authors have suggested that one important transdiagnostic factor within

mental disorders may be an inability of the brain to update its model of the body in the face of interoceptive prediction errors (i.e., mismatches between expected and received afferent interoceptive signals from the body). In predictive coding and active inference models, this kind of aberrant belief updating is thought to come about through a dysfunctional "precision weighting" mechanism, which governs the relative influence of prior expectations and afferent bodily signals in determining perception (and informing visceral regulation). Simply put, it is suggested that, across multiple mental health conditions, the brain may treat afferent bodily signals (and associated prediction errors) as though they are not reliable indicators of bodily states during interoceptive inference – leading perception to be insufficiently constrained by true visceral states and primarily determined by (in many cases maladaptive) prior expectations. Misestimating the state of the body could in turn promote a number of transdiagnostic symptoms. For example, interoceptive feelings are intimately tied to emotions (57-59); poor body perception (e.g., high uncertainty about internal physiological conditions and a resulting inability to efficiently regulate them) could thus maintain unpleasant emotional states. Chronic underestimates of available metabolic resources may contribute to apathy and anhedonia (56), and overestimates of the evidence uncomfortable bodily sensations provide for physical threat (e.g., a heart attack) may contribute to anxiety and panic (60, 61).

It's important to emphasize, however, that computational models often include several additional parameters. Aside from the precision-weighting of sensory signals, individuals can also have distinct differences in (for example) prior expectations about what they will perceive (assumed in any Bayesian model of perception; e.g., predictive coding (62)), and differences in how quickly they update those prior expectations over repeated observations (i.e., "learning rate"). Formal computational models are often necessary to distinguish which parameters show differences

between individuals and best explain differences in perception. Thus, there are multiple computational mechanisms that could account for individual differences in interoception in clinical populations. One major goal in computational psychiatry is to "computationally phenotype" patients by identifying which sets of a parameter values best account for their neural and behavioral responses (including self-reported perceptual experience) and use this information to guide treatment (20, 22, 25).

In the present study, we apply a novel computational phenotyping approach, using a Bayesian (active inference) model of perception, to identify the computational parameters that best explain behavior on a cardiac perception (heartbeat tapping) task performed by a transdiagnostic clinical sample of individuals with psychiatric disorders as well as a healthy comparison (HC) sample. Given the notably poor cardiac perception of most human beings at rest (6), we chose to assess individual differences in cardiac interoceptive precision in the context of normal physiological baseline states as well as during a non-invasive interoceptive perturbation (a breath-hold) condition that was expected to improve cardiac perception above floor values in a greater number of individuals (i.e., we expected that cardiac perception would be generally poor during resting conditions, and that the breath-hold condition would result in improved performance on average). Our primary aims were to 1) demonstrate the sensitivity of our novel computational approach in measuring the precision weighting of interoceptive signals and prior expectations across a transdiagnostic sample of individuals with depression, anxiety, substance use disorders, and/or eating disorders, 2) test the hypothesis (as previously proposed; e.g., (10, 38, 56)) that these patient groups would show lower interoceptive precision weightings than HCs, more precise prior expectations than HCs, or both, and 3) explore whether prior expectations and/or

interoceptive precision is abnormal in general or selectively within resting or interoceptive perturbation conditions.

Methods

Participants

Data were collected from 500 participants (153 male) as a part of the Tulsa 1000 (T1000) project (63), a naturalistic longitudinal study that recruited subjects based on the dimensional NIMH Research Domain Criteria framework. Individuals aged 18-55 years were screened on the basis of dimensional psychopathology scores. Inclusion was based on the following measures: Patient Health Questionnaire (PHQ-9; (64)) ≥ 10, Overall Anxiety Severity and Impairment Scale (OASIS; (65)) ≥ 8, Drug Abuse Screening Test (DAST-10; (66)) score > 2, and/or Eating Disorder Screen (SCOFF; (67)) score ≥ 2 (for screening measure scores, see **Table 1**). HCs who did not show elevated symptoms or psychiatric diagnoses were also included. Participants were excluded if they (i) tested positive for drugs of abuse, (ii) met criteria for psychotic, bipolar, or obsessive-compulsive disorders, or reported (iii) history of moderate-to-severe traumatic brain injury, neurological disorders, or severe or unstable medical conditions, (iv) active suicidal intent or plan, or (v) change in psychotropic medication status within 6 weeks. Full inclusion/exclusion criteria are described in (63). The study was approved by the Western Institutional Review Board. All participants provided written informed consent prior to completion of the study protocol, in accordance with the Declaration of Helsinki, and were compensated for participation. ClinicalTrials.gov identifier: #NCT02450240.

Similar to previous studies of the T1000 cohort (68), participants were grouped based on DSM-IV or DSM-5 diagnosis using the Mini International Neuropsychiatric Inventory 6 or 7 (MINI;

(69)), and included those with major depressive disorder only (DEP), an anxiety disorder only (ANX; social anxiety, generalized anxiety, panic, or posttraumatic stress disorder), comorbid MDD and anxiety disorder (DEP+ANX), substance use disorders (SUDs; recreational drugs excluding alcohol and nicotine; with or without comorbid anxiety and mood disorders), eating disorders (EDs; with or without comorbid anxiety and mood disorders), and HCs with no mental health diagnoses. In this study we examined all groups from the first 500 participants of the T1000 project (recruited from January 5, 2015 to February 22, 2017).

***Table 1**. Mean (and standard deviation) for clinical and demographic variables*

Individual difference variable*	**Healthy comparisons (N = 52)**	**Anxiety (N = 15)**	**Depression (N = 69)**	**Depression + Anxiety (N = 153)**	**Eating Disorder (N = 14)**	**Substance use disorder (N = 131)**	***p***
Age	32.04 (11.08) `	36.42 (10.01)	37.12 (11.83)	35.28 (11.23)	27.40 (9.83)	34.03 (8.88)	0.012
Sex	50% male	33% male	30% male	26% male	14% male	45% male	0.001
PHQ-9	0.83 (1.29)	7.47 (5.80)	13.48 (4.73)	13.10 (5.21)	12.93 (7.98)	6.68 (5.91)	<0.001
OASIS	1.37 (1.89)	10.60 (2.16)	7.55 (3.34)	10.51 (3.12)	9.50 (4.93)	5.94 (4.77)	<0.001
DAST-10	0.10 (0.30)	0.33 (0.49)	0.72 (1.66)	0.58 (1.13)	1.23 (2.62)	7.55 (2.05)	<0.001
PTT	0.20 (0.02)	0.20 (0.01)	0.19 (0.01)	0.20 (0.02)	0.20 (0.02)	0.20 (0.02)	0.261
BMI	27.59 (5.54)	27.05 (6.02)	28.51 (5.47)	28.77 (5.49)	22.25 (4.43)	28.23 (4.56)	0.001

***PHQ-9 = Patient Health Questionnaire 9; OASIS = Overall Anxiety Sensitivity and Impairment Scale; DAST-10 = Drug Abuse Screening Test; PTT = median pulse transit time; BMI = Body Mass Index**

Heartbeat perception task

As part of the T1000 project, participants completed a large number of assessments, self-report measures, and behavioral tasks (detailed in (63)). Here we focus on data from a cardiac perception task on which we have previously reported (i.e., on a subset of the participants reported here, with analyses unrelated to computational modeling (70, 71)), wherein participants were asked to behaviorally indicate the times at which they felt their heartbeat. The utilization of the heartbeat tapping measure as an index of perception was based on a previously developed heartbeat tapping task (40); for a more recent example, see (72)). The task was repeated under multiple conditions designed to assess the influence of cognitive strategy and physiological perturbation on performance. In the initial task condition, participants were simply instructed to close their eyes and press down on a key when they felt their heartbeat, to try to mirror their heartbeat as closely as possible, and even if they weren't sure they should take their best guess (the “guessing” condition). Participants completed this (and each other) task condition over a period of 60 seconds. In the second task condition, all instructions were identical except that they were told to only press the key when they actually feel their heartbeat, and if they do not feel their heartbeat then they should not press the key (the “no-guessing” condition). In other words, unlike the first time they completed the task, they were specifically instructed not to guess if they didn't feel anything. Finally, in the perturbation condition, participants were again instructed not to guess but were also asked to first empty their lungs of all air and then inhale as deeply as possible and hold it for as long as they could tolerate (up to the length of the one-minute trial) while reporting their perceived heartbeat sensations. This third condition (the “breath-hold” condition) was used in an attempt to putatively increase the strength of the afferent cardiac signal by increasing physiological arousal. We expected that cardiac perception would be poor in the

guessing condition, that tapping would be more conservative in the no-guessing condition, and that the breath-hold condition would result in improved performance on average. As a control condition, we also included an identical task where participants were instructed to tap every time they heard a 1000Hz auditory tone presented for 100ms (78 tones, randomly jittered by +/- 10% and presented in a pattern following a sine curve with a frequency of 13 cycles/minute, mimicking the range of respiratory sinus arrhythmia during a normal breathing range of 13 breaths per minute). This was completed between the first (guessing) and second (no-guessing) heartbeat tapping conditions.

Directly after completing each task condition, individuals were asked the following using a visual analogue scale:

“How accurate was your performance?”

“How difficult was the previous task?”

“How intensely did you feel your heartbeat?”

Each scale had anchors of “not at all” and “extremely” on the two ends. Numerical scores could range from 0 to 100.

Computational model

To model behavior on the heartbeat tapping task, we first divided each task time series into intervals corresponding to the periods of time directly before and after each heartbeat. Potentially perceivable heartbeats were based on the timing of the peak of the electrocardiogram (EKG) R-wave (signaling electrical depolarization of the atrioventricular neurons of the heart) + 200 milliseconds (ms). This 200 ms interval was considered a reasonable estimate of participants’

pulse transit time (PTT) according previous estimates for the ear PTT (73). We also measured the average PTT of each participant, defined as the distance between the peak of the EKG R-wave and the onset of the peak of the PPG waveform (signaling mechanical transmission of the systolic pressure wave to the earlobe). The length of each heartbeat interval (i.e., the "before-beat interval" and "after-beat interval") depended on the heart rate. For example, if two heartbeats were 1 second apart, the "after-beat interval" would include the first 500 ms after the initial beat and the "before-beat interval" would correspond to the 2nd 500 ms. The after-beat intervals were considered the time periods in which the systole (heart muscle contraction) signal was present and in which a tap should be chosen if it was felt. The before-beat intervals were treated as the time periods where the diastole (heart muscle relaxation) signal was present and in which tapping should not occur (i.e., assuming taps are chosen in response to detecting a systole; e.g., as supported by (74)). This allowed us to formulate each interval as a "trial" in which either a tap or no tap could be chosen and whether a systole or diastole signal was present (see **Figure 1**).

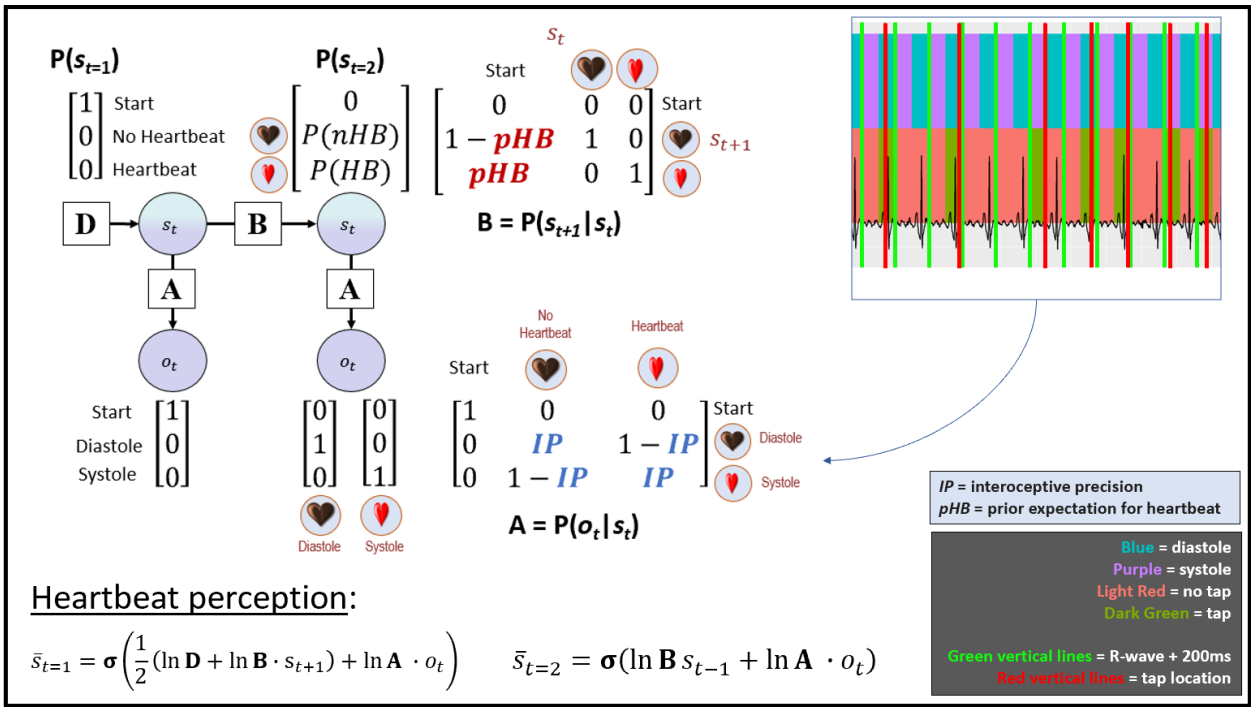


Figure 1. Bayesian approach used to model interoceptive awareness on the heartbeat tapping task. The generative model is here depicted graphically, such that arrows indicate dependencies between variables. Associated vectors/matrices are also shown. At each time point (t), observations (o) depend on hidden states (s), where this relationship is specified by the **A** matrix, and those states depend on previous states (as specified by the **B** matrix, or the initial states specified by the **D** vector). This model represents a simplified version of a commonly used active inference formulation of partially observable Markov decision processes, but which does not explicitly model action selection (for more details regarding the structure and mathematics describing these models, see (75-77). In our model, the observations were systole/diastole, and the hidden states included beliefs about the presence or absence of a heartbeat. For simplicity, the probability of choosing to tap is here assumed to correspond to the posterior distribution over states ($\bar{s}$) – that is, the relative confidence in the presence vs. absence of a heartbeat: *P*(*HB*) and *P*(*nHB*), respectively. The model parameters we estimated corresponded to: 1) interoceptive precision (IP) – the precision of the mapping from systole/diastole to beliefs about heartbeat/no heartbeat in the **A** matrix, which can be associated with the weight assigned to sensory prediction errors; and 2) prior expectations for the presence of a heartbeat (pHB). Because minimal precision corresponds to an IP value of .5, and both higher and lower values indicate that taps will more reliably track systoles (albeit in an anticipatory or reactive manner), our ultimate measure of precision subtracted 0.5 from raw IP values and then took their absolute value. The raw IP values were then used to assess for group differences in the tendency to tap before vs. after each systole. We also compared this model to an analogous model that included learning (see main text). On each trial, beliefs about the probability of a heartbeat (corresponding to the probability of choosing to tap) relied on Bayesian inference as implemented in the "heartbeat perception" equations shown at the bottom of the figure. Note that, by convention in active inference, the dot product (·) applied to matrices here indicates transposed matrix multiplication. Observed diastoles and systoles were taken from ECG traces, after dividing the time series into periods before and after each systole (exemplified in the right portion of the figure; see text for details).

To model behavior, we used a Bayesian model of perception (see **Figure 1**) derived from a Markov decision process (MDP) formulation of active inference that has been used in previous work; for more details about the structure and mathematics of this class of models, see (76, 78, 79). Unlike the full MDP model, however, which includes both a perception model and an action model, here we only explicitly included a generative model of perception. This model specified the random variables (see **Table 2**), and their dynamics over time, associated with each time period ("trial") in the task in which a systole either did or did not occur. Observations (o) in the model formally included systole, diastole, and a "start" observation. Perceptual states (s) included either feeling one's heartbeat or not, as well as a "start" state. Here, a trial formally included two timesteps: 1) a "start" time point, followed by 2) the possibility of either a systole or diastole.

A vector **D** encoding prior expectations over initial states, $P(s_{t=1})$, specified that the participant always started the trial in the start state with perfect confidence. The relationships between states and observations at a time (t) within a trial are described by a set of matrices. The **A** matrix encodes how perceptual states are informed by observations, $P(o_t|s_t)$ – here corresponding to the probability of receiving an afferent cardiac signal reflecting systole/diastole if a heartbeat did or did not occur. The precision of this matrix was controlled by an "interoceptive precision" (IP) parameter. A value of 0.5 would indicate minimal precision (i.e., leading the individual to simply "guess" whether he/she experienced a heartbeat), whereas a value approaching 0 or 1 indicates high precision (i.e., leading the individual to more consistently respond before, or consistently respond after, a heartbeat has occurred). The **B** matrix encodes the probability that one state will transition into another, $P(s_{t+1}|s_t)$ – here corresponding to the probability of transitioning from the "start" state to a "heartbeat state" vs. a "no heartbeat" state. This probability was controlled

by a parameter pHB, where values above .5 indicate prior expectations favoring feeling a heartbeat (e.g., expecting a faster heart rate), and values less than .5 indicate stronger expectations not to feel a heartbeat (e.g., expecting slower heart rate). Both IP and pHB were estimated for each participant based on their tapping behavior, as described below.

Table 2. *Description of active inference model elements*

Model variable	**General Definition**	**Model-specific specification**
o_t	Observable outcomes at time t	Outcome modalities: 1. Diastole vs. systole
s_t	Hidden states at time t	Hidden state factors: 1. Feel heartbeat vs. no heartbeat
A matrix $p(o_t \mid s_t)$	A matrix encoding beliefs about the relationship between hidden states and observable outcomes (i.e., the probability that specific outcomes will be observed given specific hidden states).	Encodes beliefs about the relationship between felt heartbeats and diastole vs. systole. The precision of the relationship between heartbeats and diastole/systole was controlled by a parameter IP, which specified how much evidence systole provides for a heartbeat and how much evidence diastole provides for the absence

		of a heartbeat.
B matrix $p(s_{t+1} \mid s_t)$	A matrix encoding beliefs about how hidden states will evolve over time (transition probabilities).	Encodes the prior expectation that either a heartbeat or no heartbeat would occur on each trial, as controlled by a parameter pHB.
D vector $p(s_1)$	A matrix encoding beliefs about (a probability distribution over) initial hidden states.	Ensures the individual always begins in an initial starting state.

In this task, it is unclear whether or not learning over time (e.g., from paying attention to one's heartbeat) contributes to task performance. To assess this using Bayesian model comparison, we compared evidence for the "perception only" model described above with evidence for a model that included learning to update prior expectations over time. Learning within this model involves updating beliefs about the probability of feeling a heartbeat vs. no heartbeat at each time point, based on how frequently one believes they have felt their heartbeat in the past. Essentially, every time a heartbeat is felt, prior expectations favoring feeling a heartbeat go up, and every time no heartbeat is felt this (relative) expectation goes back down. Formally, this corresponds to updating the concentration parameters of Dirichlet (Dir) priors associated with the **B** matrix (b) that specify beliefs about state transitions. At $t = 0$:

$$P(\boldsymbol{B}) = Dir(b)$$

$$b(heartbeats) = P(s_{t+1}|s_t) = \begin{bmatrix} 0 & 0 & 0 \\ 1 - pHB & 1 & 0 \\ pHB & 0 & 1 \end{bmatrix} \times b_0$$

$$b_t = b_{t-1} + \sum_t s_t \otimes s_{t-1}$$

Here $\otimes$ indicates the cross-product, and b_0 is a scalar on the prior value for concentration parameters, where its value prior to learning encodes (inverse) sensitivity to information, such that higher values will reduce the rate at which prior expectations are updated over time with new observations. In the learning model, b_0 was also estimated for each individual to capture the possibility of different learning rates for updating prior expectations (this could also be thought of as differences in a kind of interoceptive "belief rigidity").

Thus, the final parameters estimated for each participant included the IP, pHB, and b_0 parameters. Our approach to parameter estimation used Bayesian inference at two levels (80). First, each participant's responses were modeled using the Bayesian model of perception described above. We then used a commonly used Bayesian optimization algorithm (called Variational Bayes) to estimate each participant's parameter values that maximized the likelihood of their responses (under the assumption that a higher/lower probability assigned to feeling a heartbeat corresponded to a higher/lower probability of choosing to tap), as described in (25). We optimized these parameters for each model using this likelihood and variational Laplace (81), implemented within the spm_nlsi_Newton.m parameter estimation routine available within the freely available SPM12 software package (Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm). This estimation approach has the advantage of preventing overfitting, due to the greater cost it assigns to moving parameters farther from their

prior values. Estimating parameters required setting prior means and prior variances for each parameter. The prior variance was set to a high precision value of 1/2 for each parameter (i.e., deterring overfitting), and the prior means were set as follows: IP = .5, pHB = .5, and b_0 = 1. Our decision for selecting these priors was motivated in part by initial simulations confirming that parameter values were recoverable under these prior values. The IP and pHB prior values were further chosen to minimize estimate bias, as pHB = .5 assumes flat prior expectations, and IP = .5 does not bias estimates in favor of values assuming anticipatory vs. reactive strategies. The b_0 prior of 1 is equivalent to this parameter having no effect on the model. After fitting parameters for each model, we then performed Bayesian model comparison (based on (82, 83)) to determine the best model. We then used classical inference to test for the effects of group differences in parameter estimates for the best model, using a standard summary statistic approach.

Before using these parameters in further analyses, however, the "raw" IP parameter values (IP_{raw}) were transformed to correctly capture 2 distinct constructs of interest. First, because IP_{raw} values both above and below .5 indicate higher precision (i.e., values below approaching 0 indicate reliable anticipatory tapping, whereas values approaching 1 indicate reliable tapping after a systole), our ultimate measure of precision was recalculated by centering IP_{raw} on 0 and taking its absolute value as follows:

$$IP = |IP_{raw} - 0.5|$$

This means that IP has a minimum value of 0 and a maximum value of 0.5. The IP_{raw} values were then instead used to assess individual differences in the tendency to tap in an anticipatory or reactive (*AvR*) fashion:

$$AvR = IP_{raw}$$

Higher AvR values (> 0.5) thus indicated a stronger tendency to reactively tap in response to a heartbeat as opposed to tapping in an anticipatory fashion (< 0.5).

Physiological measurements

Electrocardiography was used to assess the objective timing of participants' heartbeats throughout the task. A BIOPAC MP150 was used to collect a three lead EKG signal and the pulse oximeter signal, using a pulse plethysmography (PPG) device attached to the ear lobe. Response times were collected using a task implemented in PsychoPy, with data collection synchronized via a parallel port interface.

EKG and response data were scored using in-house developed MATLAB code. As described above in relation to modelling, each participant's pulse transit time (PTT) was estimated as the median delay between R wave and the corresponding inflection in the PPG signal. **Supplementary materials** display an example participant's EKG and PPG trace to illustrate how this was calculated (also supporting the 200ms delay assumption used in modelling and in individuals without usable PPG).

Quality Control and Final Sample Sizes

Prior to performing our analyses, several participants were removed due to quality control checks: 19 individuals were removed due to "cheating" (i.e., video revealed they were taking their pulse while performing the task); 3 individuals didn't complete the task; and 13 individuals had poor EKG across all trials that didn't allow reliable identification of heartbeat timing. An additional 31 individuals were removed due to being outliers when performing the tone task (using Iterative Grubb's with $p < 0.01$) – assumed to reflect inappropriate engagement during the task (e.g., being inattentive, tapping rapidly without listening to the tones, etc.). This resulted in

434 participants, including 52 HCs, 15 ANX, 69 DEP, 153 co-morbid ANX/DEP, 131 SUDs, and 14 eating disorders. In the three heartbeat tapping conditions, a few other participants were removed due to poor EKG, leading to the following participant numbers in each condition: 433 (guessing), 431 (no-guessing), and 427 (breath-hold). **Table 3** reports the final number of participants included for model-based analyses in each condition as well as descriptive statistics for all task-related measures.

Statistical analysis

Correlational analyses were first conducted to examine the relationships between parameters across conditions. For purposes of parameter validation, we ran further correlational analyses to examine the relationships between each parameter and task-specific measures, including the self-report ratings of difficulty, confidence, and heartbeat intensity collected after each trial.

Analyses of variance (ANOVAs) and covariance (ANCOVAs) were conducted to identify possible group differences (i.e., between HCs and the five patient groups) in each parameter and in how they differed across trials (i.e., guessing, no-guessing, breath-hold), while accounting for individual differences in age, sex, BMI, median PTT, number of heartbeats (and its interaction with condition), IP for the tone condition (to rule out potential effects of reaction time), and medication status (i.e., one analysis per parameter). Because the first half of the T1000 sample was intentionally designed as an exploratory sample, we report relationships at $p < .05$ as a potential basis for a priori hypotheses verification in the second (confirmatory) half of the T1000 dataset in subsequent work. Although our analyses are exploratory, we note that a Bonferroni corrected threshold for multiple comparisons with three parameters is $p < .017$ (α=0.05).

Results

Participant characteristics

Complete information on sample size, demographics, and symptom screening measures is provided in **Table 1**. Separate ANOVAs showed significant differences between groups in age ($F(5,428) = 2.98$, $p = .01$) and body mass index (BMI; $F(5,413) = 4.04$, $p = .001$). A chi-squared analysis also showed significant differences in the proportion of males to females between groups (chi-squared = 19.43, df = 5, $p = .002$). Therefore, in our analyses of model parameters, we also confirm our results after controlling for these other factors.

Interoceptive Perturbation Validation

As expected, across all participants self-reported heartbeat intensity, confidence in task performance, and task difficulty differed significantly between the three heartbeat tapping conditions in separate ANOVAs (intensity: $F(2,1288) = 42.67$, $p < .001$; confidence: $F(2,1288) = 24.94$, $p < .001$; difficulty: $F(2,1288) = 18.26$, $p < .001$), reflecting a greater perceived intensity of heartbeat sensations and greater confidence during the breath-hold perturbation than in the other 2 conditions, as well as lower difficulty in the breath-hold condition than the no-guessing condition (and greater difficulty in the no-guessing than guessing condition; Tukey post-hoc comparisons: $p < .001$ for all). Analysis of heart rate revealed a marginal difference in the number of heartbeats between conditions ($F(2,1381) = 2.49$, $p = .08$), reflecting a faster heart rate in the breath-hold than no-guessing conditions (Tukey post-hoc comparison: $p = .07$).

Table 3. *Summary statistics for all task-related variables*

Panel A.

Guessing Condition	**Healthy comparisons (N = 52)**	**Anxiety (N = 15)**	**Depression (N = 69)**	**Depression + Anxiety (N = 153)**	**Eating Disorder (N = 14)**	**Substance use disorder (N = 130)**	**p-value***
Number of taps	67.92 (29.58)	58.40 (28.40)	67.65 (31.97)	62.02 (32.58)	65.86 (27.86)	65.50 (30.82)	0.701

Number of heartbeats	66.67 (9.48)	67.13 (9.55)	70.23 (10.90)	70.34 (10.86)	67.21 (8.71)	71.23 (10.47)	0.1
Self-reported difficulty	56.87 (22.70)	53.20 (26.69)	57.55 (28.12)	53.18 (25.84)	71.57 (17.78)	48.24 (24.65)	0.009
Self-reported confidence	32.67 (19.27)	41.27 (21.54)	39.52 (22.67)	39.29 (21.20)	25.14 (21.36)	42.20 (22.39)	0.022
Self-reported intensity	20.71 (16.72)	26.73 (21.51)	26.26 (21.52)	28.27 (23.25)	15.36 (18.64)	35.94 (24.65)	<0.001
Counting accuracy	0.69 (0.33)	0.67 (0.30)	0.65 (0.29)	0.63 (0.30)	0.68 (0.30)	0.67 (0.29)	0.755

*p-values correspond to the results of ANOVAs comparing the groups (i.e., not including all task conditions within the analysis as reported in the main text).

Panel B.

No-Guessing Condition	**Healthy comparisons (N = 52)**	**Anxiety (N = 15)**	**Depression (N = 68)**	**Depression + Anxiety (N = 153)**	**Eating Disorder (N = 14)**	**Substance use disorder (N = 129)**	**p-value**
Number of taps	18.73 (22.69)	22.13 (24.04)	23.43 (24.22)	24.10 (24.26)	18.29 (21.74)	31.47 (25.46)	0.015
Number of heartbeats	65.88 (10.22)	66.73 (11.07)	69.97 (10.45)	69.78 (11.10)	66.14 (8.43)	70.69 (10.14)	0.066
Self-reported difficulty	67.83 (27.94)	59.73 (35.14)	68.00 (32.72)	62.88 (31.35)	69.86 (26.68)	55.14 (30.75)	0.039
Self-reported confidence	39.94 (30.32)	36.80 (35.54)	41.15 (31.36)	41.41 (29.49)	31.93 (32.16)	42.29 (28.27)	0.858
Self-reported intensity	18.15 (19.63)	16.80 (22.11)	25.82 (26.38)	25.31 (25.67)	13.50 (13.51)	33.64 (27.53)	0.001
Counting accuracy	0.28 (0.29)	0.31 (0.33)	0.32 (0.31)	0.33 (0.31)	0.28 (0.32)	0.43 (0.32)	0.028

Panel C.

Breath Hold Condition	**Healthy comparisons (N = 52)**	**Anxiety (N = 15)**	**Depression (N = 68)**	**Depression + Anxiety (N = 151)**	**Eating Disorder (N = 14)**	**Substance use disorder (N = 127)**	**p-value**
Number of taps	27.12 (21.47)	30.60 (42.93)	24.75 (21.07)	26.46 (23.58)	17.71 (16.80)	34.17 (25.05)	0.027
Number of heartbeats	69.40 (9.80)	66.87 (12.01)	71.26 (10.48)	70.59 (11.20)	66.29 (6.29)	72.96 (10.75)	0.061
Self-reported difficulty	46.44 (27.08)	54.20 (23.24)	56.53 (25.74)	53.50 (28.32)	55.64 (25.67)	45.86 (29.08)	0.07

Self-reported confidence	47.79 (26.20)	41.27 (26.06)	50.78 (26.49)	52.42 (27.42)	36.50 (33.37)	53.04 (25.55)	0.158
Self-reported intensity	38.19 (28.40)	26.40 (23.61)	41.85 (29.60)	41.09 (29.99)	32.71 (30.63)	46.81 (28.88)	0.074
Counting accuracy	0.39 (0.29)	0.28 (0.26)	0.34 (0.27)	0.35 (0.29)	0.28 (0.29)	0.44 (0.30)	0.03

Panel D.

Tone Condition	**Healthy comparisons (N = 52)**	**Anxiety (N = 15)**	**Depression (N = 69)**	**Depression + Anxiety (N = 153)**	**Eating Disorder (N = 14)**	**Substance use disorder (N = 131)**	**p-value**
Number of taps	77.90 (1.00)	77.40 (1.12)	77.71 (1.03)	77.82 (1.17)	77.57 (0.94)	78.06 (0.97)	0.072
Number of heartbeats	65.63 (11.10)	67.27 (9.81)	70.74 (9.94)	71.37 (11.14)	66.86 (8.16)	71.30 (10.62)	0.01
Self-reported difficulty	18.04 (15.84)	29.20 (21.12)	22.14 (19.52)	24.15 (22.00)	35.21 (22.33)	25.77 (22.40)	0.065
Self-reported confidence	77.56 (14.26)	73.93 (17.84)	78.84 (16.60)	74.22 (19.59)	59.79 (24.84)	75.16 (18.56)	0.017
Self-reported intensity	85.58 (13.76)	75.67 (15.19)	86.20 (15.09)	82.86 (17.83)	87.79 (12.46)	81.47 (16.51)	0.102
Counting accuracy	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.922

Bayesian Model Comparison

When comparing models (based on (82, 83)), the "perception only" model was better than the model that included learning prior expectations for the no-guessing and breath-hold conditions (protected exceedance probability = 1), whereas there was not a significantly better model for the guessing condition (protected exceedance probabilities = .46 vs. .54, slightly favoring the learning model). The learning model was better in the tone condition (protected exceedance probability = 1). No group differences were observed when comparing model fits between groups.

For consistency/comparability, we use the "perception only" model parameters to compare conditions in our analyses below, as this model best explained heartbeat tapping behavior overall. The accuracy of this model – defined as the percentage of choices to tap/not tap that matched the highest probability action in the model (e.g., a tap occurring when the highest probability percept in the model was a heartbeat) – was 74% across all conditions; by condition, model accuracy was: guessing condition = 67% (SD = 12%); tone condition = 67% (SD = 12%); no-guessing condition = 85% (SD = 14%); breath-hold condition = 84% (SD = 13%).

Relationship between parameters

As shown in **Figure 2**, correlations between IP across task conditions were generally low. Correlations between pHB estimates across conditions were moderate, most notably between the no-guessing and breath-hold conditions (which also included the no-guessing instruction). The tendency to tap in an anticipatory vs. reactionary manner (AvR) showed no relationships across conditions. Correlations between IP and pHB (or pTone) within each condition were also low ($0 \leq r \leq .27$), as were correlations between these parameters and AvR ($-.28 < r < .01$; see **supplementary materials**).

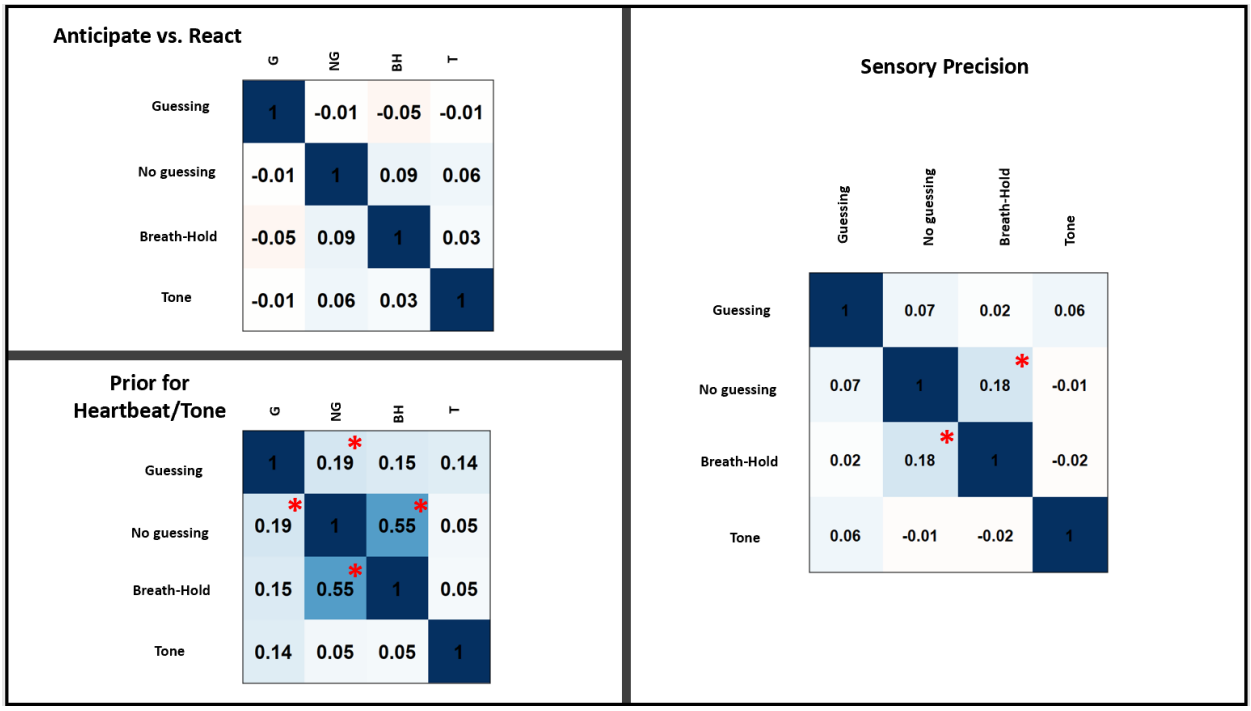


Figure 2. Pearson correlations between model parameters across task conditions across all participants. For reference, correlations at p < .001 (uncorrected) are marked with red asterisks.

Parameter face validity

Figure 3 shows the correlations, including some significant relationships ($p \leq .001$, uncorrected), across all participants between model parameters in each condition and several task-relevant variables. IP showed positive relationships with self-reported heartbeat intensity ratings in both the no-guessing and breath hold conditions. Additionally, pHB was lower in those self-reporting greater difficulty in the no-guessing condition, and higher in those self-reporting higher confidence and higher heartbeat intensity in both the no-guessing and breath hold conditions. Across heartbeat tapping conditions, model parameters were also weakly (IP) to strongly (pHB) related to the traditional counting accuracy measure (39).

In the breath-hold condition, t-tests revealed significantly greater IP in males than females ($t(251) = 2.30$, $p = .02$).

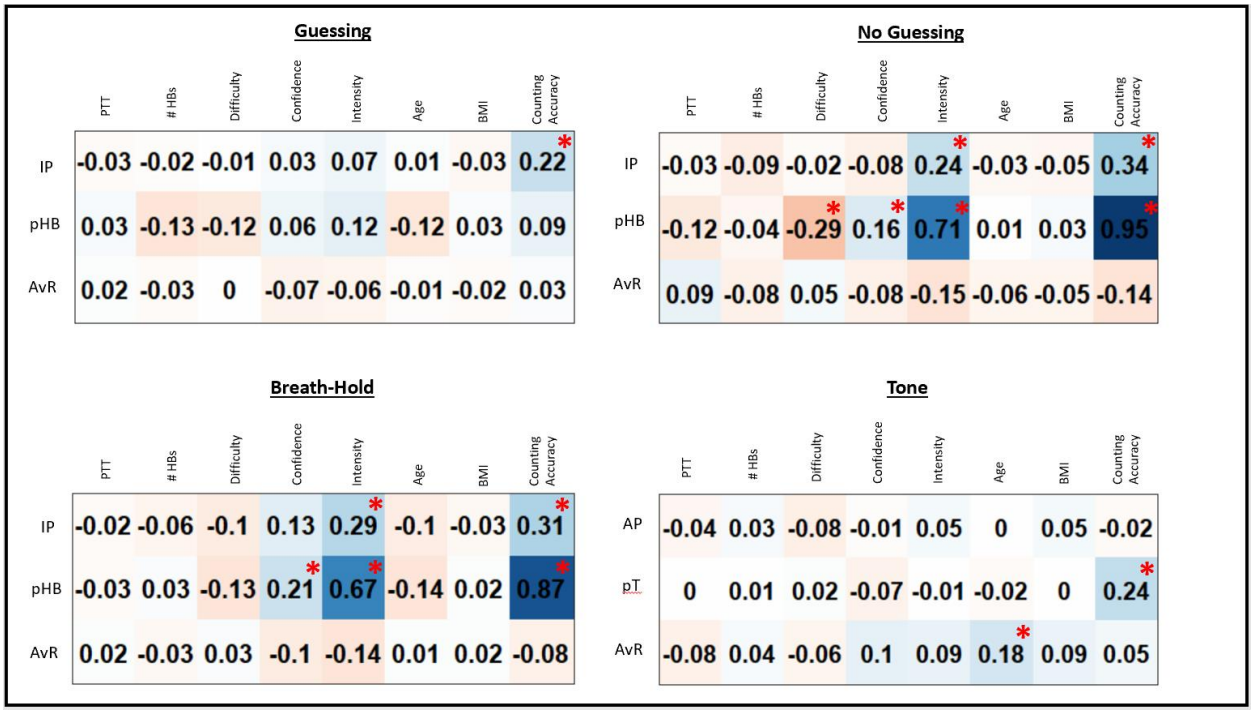


Figure 3. Exploratory Pearson correlations between model parameters and self-report and other task-relevant variables for each task condition across all participants. IP = interoceptive precision parameter, pHB = prior expectation for heartbeat parameter, pT = prior expectation for tone parameter, AvR = anticipate vs. react strategy parameter, PTT = median pulse transit time, #HBs = number of heartbeats during the task condition, BMI = body mass index. For reference, correlations at $p < .001$ (uncorrected) are marked with red asterisks.

Group by condition interactions in model parameters

Parameter values for each group and condition are listed in **Table 4**. Across conditions, all parameters were normally distributed (skew < 2).

***Table 4**. Mean (and standard deviation) for model parameters by group and condition*

Individual difference variable	**Healthy comparisons**	**Anxiety**	**Depression**	**Depression + Anxiety**	**Eating disorder**	**Substance use disorder**	**p-value***
Sensory precision							
Guessing	0.04 (0.05)	0.04 (0.02)	0.04 (0.03)	0.04 (0.04)	0.05 (0.05)	0.04 (0.04)	0.815

No-Guessing	0.04 (0.05)	0.04 (0.04)	0.03 (0.04)	0.03 (0.03)	0.06 (0.06)	0.04 (0.05)	0.09
Breath-hold	0.07 (0.07)	0.03 (0.03)	0.04 (0.04)	0.04 (0.04)	0.03 (0.03)	0.04 (0.04)	<0.001
Tone	0.18 (0.12)	0.12 (0.11)	0.18 (0.12)	0.14 (0.11)	0.21 (0.10)	0.16 (0.11)	0.052
Prior expectations							
Guessing	0.44 (0.16)	0.37 (0.16)	0.41 (0.17)	0.38 (0.18)	0.43 (0.18)	0.40 (0.16)	0.331
No-Guessing	0.15 (0.13)	0.15 (0.12)	0.17 (0.13)	0.17 (0.14)	0.14 (0.12)	0.21 (0.14)	0.05
Breath-hold	0.18 (0.11)	0.19 (0.19)	0.17 (0.11)	0.18 (0.13)	0.14 (0.11)	0.22 (0.13)	0.056
Tone	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.139
Anticipate vs. React							
Guessing	0.49 (0.06)	0.50 (0.04)	0.49 (0.05)	0.49 (0.06)	0.54 (0.07)	0.50 (0.06)	0.042
No-Guessing	0.50 (0.07)	0.52 (0.06)	0.49 (0.05)	0.50 (0.05)	0.50 (0.08)	0.50 (0.06)	0.683
Breath-hold	0.49 (0.10)	0.49 (0.04)	0.51 (0.06)	0.50 (0.06)	0.48 (0.04)	0.50 (0.05)	0.529
Tone	0.50 (0.22)	0.45 (0.16)	0.52 (0.21)	0.49 (0.18)	0.37 (0.20)	0.46 (0.19)	0.088

*p-values correspond to the results of ANOVAs comparing the groups (i.e., not including all task conditions within the analysis as reported in the main text).

An ANOVA revealed that sensory precision (IP or auditory precision in the tone condition) was significantly different between conditions ($F(3,1721) = 332$, $p < .001$; see **Figure 4**), reflecting greater values in the tone condition than in the three heartbeat tapping conditions (post-hoc Tukey comparisons indicated $p < .001$ in all cases). A subsequent ANOVA focused only on IP (excluding the tone condition), including condition, clinical group, and their interaction revealed a main effect of group ($F(5,1273) = 2.35$, $p = .04$) and a group by condition interaction ($F(10,1273) = 3.10$, $p < .001$); post-hoc Tukey comparisons indicated greater IP in HCs than DEP/ANX ($p = .05$) and DEP (p = .04), and that the interaction was driven by the following: 1) IP was significantly greater in the breath-hold condition than in the guessing ($p = .005$) and no-guessing ($p = .02$) conditions in HCs, but not in any of the other clinical groups, and 2) HCs had greater IP than ANX, DEP, DEP+ANX, and SUDs in the breath-hold condition ($p = .04$, $p = .008$, p = .001, and $p < .001$, respectively). We subsequently confirmed that the group by condition interaction remained significant ($F(10,1114) = 2.7$, $p = .003$) when also including age, sex, BMI, median PTT, precision within the tone condition, number of heartbeats (and it's

interaction with condition), and medication status within the model. No significant effect on IP was observed for any of these other variables, with the exception of a negative association with age ($F(1,1114) = 5.02$, $p = .03$). The main effect of group was only present at a trend level after accounting for these additional variables ($F(2,1114) = 2.7$, $p = .07$). See **Figure 5** for plots of IP values by group and condition.

pHB was significantly different between the different task conditions ($F(3,1721) = 704.48$, $p < .001$), indicating a lower prior expectation to feel a heartbeat in the no-guessing and breath-hold condition (which retained the no-guessing instruction) than in the guessing condition, as well as a higher prior expectation to hear a tone in the tone condition than to feel a heartbeat in the three interoception conditions (post-hoc Tukey comparisons indicated $p < .001$ in all cases). No effect of group, or group by condition interactions were observed for pHB in a subsequent interoception-focused model excluding the tone condition. See **supplementary materials** for plots of pHB values by group and condition.

No difference was observed between conditions in the tendency to anticipate vs. react (AvR) to the tone/heartbeat using an ANOVA. There was also no effect of group, or condition by group interactions for AvR in a subsequent interoception-focused model excluding the tone condition. See **supplementary materials** for plots of AvR values by group and condition.

Comparison to existing measures

To examine the ability of traditional measures to capture similar group differences, we ran analogous analyses using the traditional heartbeat counting task formula for interoceptive accuracy (39). An ANOVA revealed a main effect of trial condition ($F(2,1273) = 139.00$, $p < .001$) and group ($F(5,1273) = 3.57$, $p = .03$) on counting accuracy, but no group by condition

interaction. Tukey post-hoc comparisons revealed that counting accuracy was greater in the guessing condition than in the no-guessing and breath-hold conditions (both $p < .001$), and that accuracy was higher in SUDs than in DEP and DEP/ANX ($p = .04$ and .004, respectively). These effects remained significant after accounting for the other variables we controlled for above.

To assess potential group differences in the effect of task condition on self-reported experience and physiology, we also carried analogous analyses assessing confidence, intensity, and difficulty, as well as heart rate. These results are reported in the **supplementary materials**. No group by condition interactions in these variables were observed mirroring our IP results.

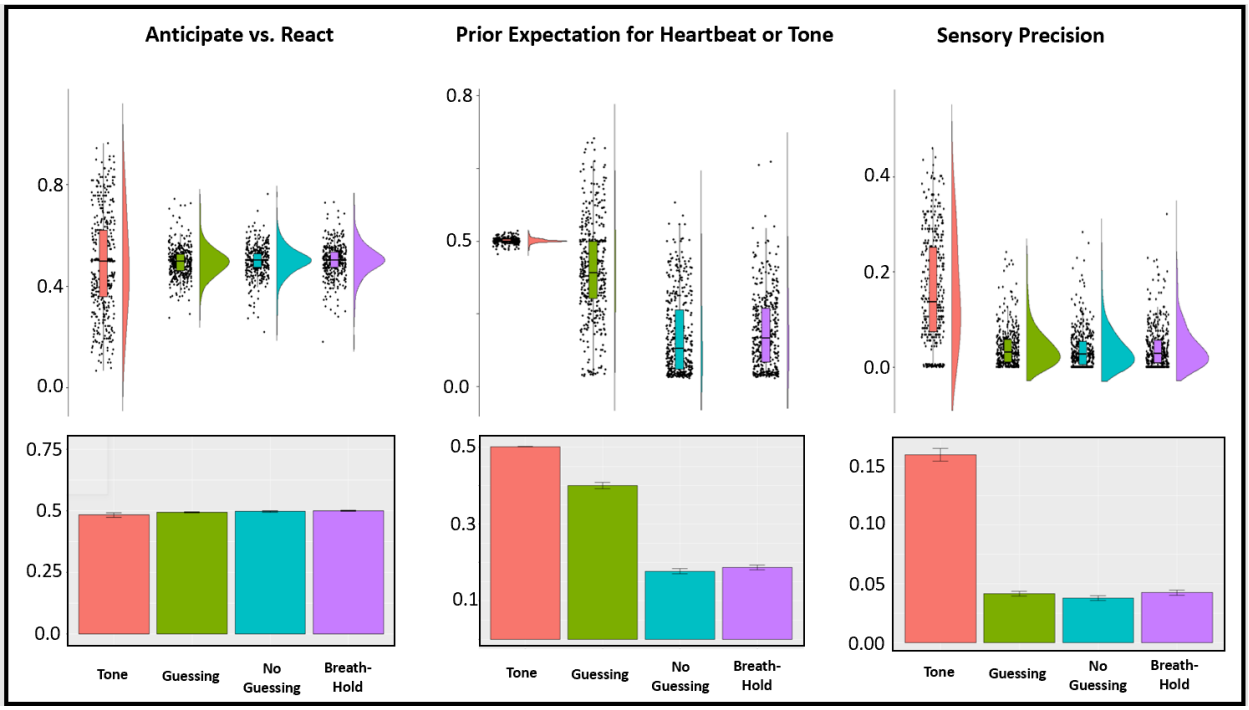


Figure 4. Bottom: Bar plots illustrating means and standard errors for model parameters by task condition across all participants. Prior expectations for heartbeats were lower in the no-guessing and breath-hold conditions. Sensory precision (i.e., interoceptive precision for the heartbeat or auditory precision for the tone condition) was much greater in the tone condition. This was expected given the unambiguous nature of this signal relative to the heartbeat signal. There were no significant differences in sensory precision between the heartbeat conditions. The Anticipate vs. React values revealed no

mean differences between conditions. Top: For more complete data characterization, we also show raincloud plots depicting the same results in terms of individual datapoints, boxplots (median and upper/lower quartiles), and distributions. These illustrate that, for the Anticipate vs. React parameter, nearly equally sized clusters of participants appeared to adopt more anticipatory (<.5) vs. reactive (>.5) strategies in the tone condition, and that prior expectations remained unbiased (.5; with little variance) in the tone condition relative to the heartbeat tapping conditions.

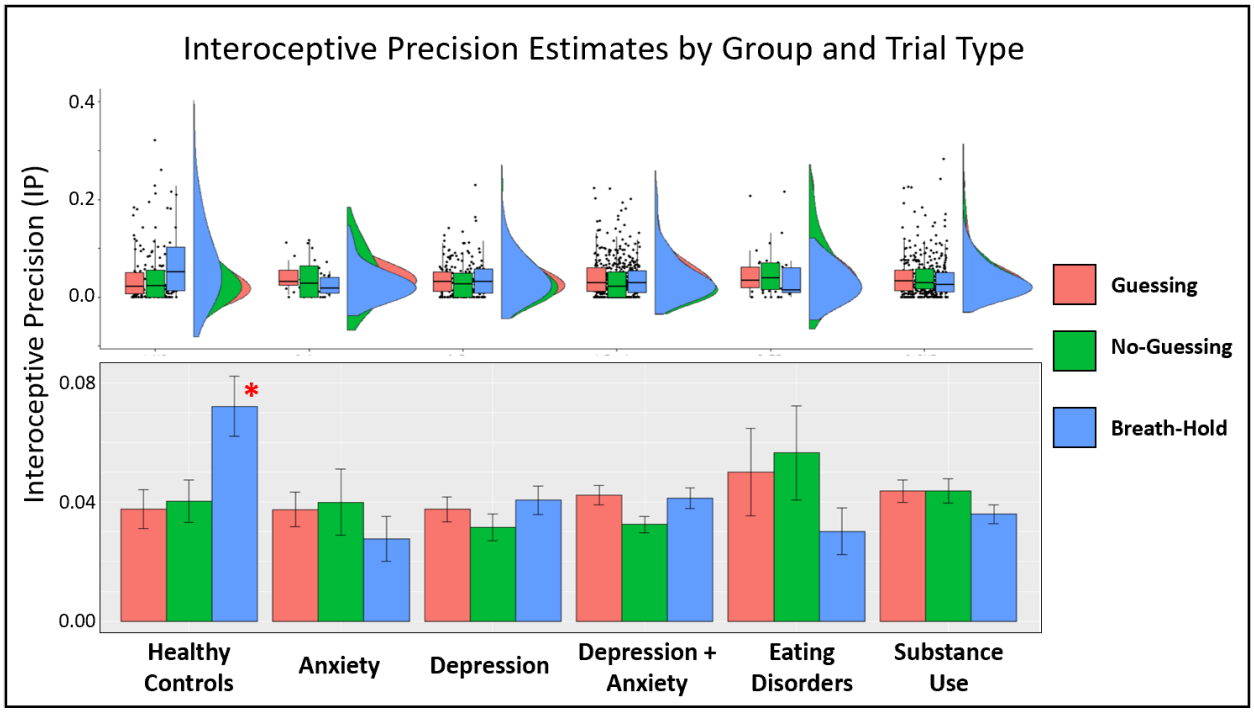


Figure 5. Bottom: Mean and standard error for interoceptive precision estimates by condition and clinical group. Interoceptive precision (IP) was significantly greater in healthy comparisons than all other groups in the breath-hold condition (*p < 0.05; with the exception of the eating disorders group, likely due to the small sample size for this group), and healthy comparisons showed a significant increase in IP from the guessing to breath-hold conditions that was absent in the other groups. Top: For more complete data characterization, we also show raincloud plots depicting the same results in terms of individual datapoints, boxplots (median and upper/lower quartiles), and distributions.

Associations with symptom severity measures and interoceptive awareness scales

Given the heterogeneity in our clinical sample, we ran subsequent exploratory correlational analyses with continuous scores on the clinical measures gathered, excluding HCs, to assess whether model parameters might provide additional information about symptom severity. We

note a few weak relationships at uncorrected levels in **supplementary materials**, but none survive correction for multiple comparisons.

The T1000 dataset also includes self-report measures commonly used in interoception research, including the Multidimensional Assessment of Interoceptive Awareness (MAIA; (84)), the Toronto Alexithymia Scale (TAS-20; (85)), and the Anxiety Sensitivity Index (ASI; (86)). For the interested reader, we also show exploratory correlation matrices between model parameters and these common measures within **supplementary materials**. While a couple of relationships were significant at uncorrected levels, the strength of these relationships was low.

Discussion

This investigation aimed to examine whether a novel computational (active inference) model of perception could provide a principled approach to empirically characterizing interoceptive processing dysfunctions that have previously been proposed. Specifically, we used behavior during a heartbeat perception task in conjunction with this model to estimate quantitative differences in the prior expectations and sensory precision estimates that individuals implicitly apply to afferent interoceptive (cardiac) signals – in both healthy individuals and a transdiagnostic sample of individuals with depression, anxiety, substance use, and/or eating disorder symptoms. We observed several relationships in the expected directions between model parameters and other task-related variables that supported the construct validity of our model parameters. Of greatest interest, we found evidence that an interoceptive (breath-hold) perturbation increased the precision estimates assigned to cardiac signals in healthy individuals, but that this perturbation had no effect on interoceptive precision in individuals with depression, anxiety, substance use, and/or eating disorders. In contrast, no significant differences in prior expectations were observed, suggesting sensory precision, and not prior expectations, best

accounts for interoceptive differences in the clinical groups. Model comparison also suggested that individuals did not update prior expectations over time during the task, which is perhaps unsurprising given the low cardiac awareness commonly seen in previous studies (6). That is, individuals may not have had sufficient signal to learn from (note that, in contrast, model comparison supported the presence of learning in our auditory control condition with a clear signal). We expand on these points below.

Parameter validity/sensitivity

IP showed positive relationships with self-reported heartbeat intensity ratings (no-guessing and breath-hold conditions) – as would be expected in the context of more precise cardiac signals. In the no-guessing and breath-hold conditions, pHB was lower than in the guessing condition. In other words, participants appear to have successfully adjusted their prior expectations to comply with the no-guessing instructions. Further, consistent with the role of prior expectations in perception, pHB was also lower in those reporting greater difficulty (no-guessing condition) and higher in those reporting greater confidence and higher heartbeat intensity (no-guessing and breath-hold conditions). Each of these results support the notion that our parameters tracked theoretically meaningful individual differences in perception/behavior and that our approach can disentangle the effects of sensory precision from those of prior expectations and anticipatory vs. reactive strategies.

Further, while model parameters had some shared variance with traditional accuracy measures (e.g., higher counting accuracy was weakly associated with higher interoceptive precision), they mainly captured unique variance that was not tracked by standard measures. Further, no group differences analogous to those seen with model parameters were found using the traditional interoceptive accuracy measure (i.e., across all participants, counting accuracy showed an

opposite pattern, being highest in the guessing condition). This highlights the unique ability of this computational method to uncover differences in perceptual decision making across different psychiatric subtypes.

It is also worth noting the strong positive correlation we observed between pHB and counting accuracy in the no-guessing and breath-hold conditions, suggesting that heartbeat counting accuracy primarily reflects prior expectations (as previously proposed in (38)). In the present task this is explained by the fact that higher pHB values led to a greater number of taps and the fact that the average number of taps in these conditions was low (i.e., because of the no-guessing instruction). Thus, those who tapped more often approached the actual number of heartbeats and therefore had higher counting accuracy scores. This highlights one specific way in which, in the context of restrictions on guessing, counting accuracy may be most closely associated with prior expectations.

Differences in Interoceptive Precision

Our primary results were that: 1) IP was higher in the healthy comparisons than in the clinical groups (in general for those with depression/anxiety, and for all clinical groups but eating disorders in the breath-hold condition; note that IP values in eating disorders were numerically comparable to the other clinical groups, but did not reach significance due to small sample size); and 2) there was a group by condition interaction, demonstrating that the interoceptive perturbation (breath-hold) increased IP (relative to resting conditions) in the healthy participants, whereas this perturbation had no effect in any of the clinical groups. These group differences were not accounted for by any other demographic (e.g., age, sex) or physiological variables (e.g., pulse transit time, changes in heart rate). The hypothesized finding that IP was reduced across psychiatric groups included in this study may be of clinical interest. First, multiple

neurocognitive (55, 87, 88) and computational (11, 12, 15, 27, 89) theories of emotion, and associated empirical findings (e.g.,(42, 90-94)), suggest that interoceptive awareness may be an important transdiagnostic factor in promoting emotional awareness. As low emotional awareness has been linked to multiple psychiatric and systemic medical conditions (reviewed in (95, 96)), reduced IP could contribute to low emotional awareness and its maladaptive consequences irrespective of diagnostic category. Second, visceral regulation might be expected to be less effective in the absence of precise feedback signals from the body, which could relate to visceral dysregulation in psychiatric conditions (e.g., see (56)). For example, to adaptively regulate visceral states, visceromotor control regions plausibly require reliable feedback signals from the body to determine whether descending visceromotor commands have successfully led to the intended adjustments in internal bodily states (10, 13, 36). Thus, the low IP observed across multiple psychiatric conditions could potentially contribute to visceral dysregulation in other contexts in which interoceptive perturbations occur, such as highly arousing, negatively valenced states (e.g. panic anxiety, irritability/anger etc.).

These results build on previous bodies of work suggesting associations between psychiatric disorders and interoceptive processing deficits (38, 56, 97-100). For example, previous cardiac perception studies have shown that depressed patients exhibit reduced accuracy on a heartbeat counting task (2-4), and that performance is negatively correlated with depressive symptoms (5) as well as associated with both lower positivity and poorer decision-making (2); although, the limitations of heartbeat counting tasks should be kept in mind when interpreting such findings (47, 49, 101, 102). While the literature on interoceptive dysfunction is mixed for anxiety disorders broadly, it is well-stablished in panic disorder (reviewed in (6)). A couple recent studies have also reported evidence of differences in interoceptive processing in both SUDs (e.g.,

blunted brain responses (7)) and eating disorders (e.g., stronger expectation effects on perception during low arousal (8)). The computational framework within which our findings were observed is also in line with several recent proposals about the role of interoceptive inference in guiding (predictive) autonomic control and the potential breakdown of this mechanism within different psychiatric conditions (10-15, 36, 38); however, in contrast to previous emphasis on altered prior expectations in these proposals, our results more selectively support the existence of deficits in adjusting precision estimates for afferent interoceptive signals – and do not suggest the presence of altered priors.

That said, the neurobiological mechanisms promoting reduced IP in mental disorders during interoceptive perturbation remains unclear. As IP increased during the interoceptive perturbation in HCs, but not in the clinical groups, it could be that altered brain processes in individuals with certain psychiatric disorders fail to update IP estimates during states of acute bodily arousal. The neural process theory associated with active inference suggests an inability to adjust IP estimates would correspond to reductions in synaptic plasticity in response to changes in patterns of interoceptive prediction-errors (10, 24, 76, 78, 103), most plausibly within neural networks supporting interoception and visceromotor control (e.g., insula and anterior cingulate (15, 104)). This could be tested using our model/task in conjunction with neuroimaging. Alternatively, IP estimates could be accurate, and afferent interoceptive signals may in fact be conveyed with less fidelity (i.e., greater noise) to the brain in the context of psychiatric conditions. For example, such differences could be due to altered signaling in interoceptive sensory axons, which could in principle be affected by many factors (genetic/epigenetic influences, early adversity and related socio-environmental factors, and/or effects of disease-related chronic stress, among others). Future research will need to investigate these different possibilities.

Strengths, Limitations, and Conclusion

This study has several major strengths. First is the novel application of a computational (active inference) model to behavior on an interoceptive awareness task, which allowed for model comparison (e.g., allowing us to rule out learning effects) and parameter estimates to disentangle distinct computational mechanisms (e.g., the role of prior expectations vs. interoceptive precision). A second strength is the application of this model to interoceptive processing in individuals with psychiatric disorders, which to our knowledge, has never been reported. While accounting for other influences, this approach allowed us to test a theoretical prediction – that afferent interoceptive signals (and resulting prediction-errors) are assigned low precision estimates in those with psychiatric disorders, leading those signals to be under-weighted during interoceptive inference (although, as discussed above, other interpretations are possible). While an important first step, a further test of this hypothesis would require replication in a confirmatory sample, something we plan to investigate in the second cohort of participants in the Tulsa 1000 project. Additionally, combining this approach with neuroimaging measures would help to establish the expected relationship between IP and neural activity within interoceptive cortices. Active inference models may be especially useful in this regard, as they allow for simulation of predicted neuronal responses (78).

Some limitations of our modeling approach are also important to keep in mind. First, we were required to make specific choices about model structure. For example, while priors during estimation were fixed at plausible values (justification explained within the methods section), other values might have been chosen. Also, we chose not to include an explicit action model. While this was possible, in simple perceptual tasks like ours that lack reward feedback, any variability in the precision of action selection would be correlated with sensory precision;

therefore, we chose a simpler model that treated behavior as a direct readout of confidence in perception (and controlled for sensory precision in the tone condition to account for individual differences in motor stochasticity). However, the novelty of our approach entails that it should be replicated in future studies.

Another limitation is that, while we did compare models with vs. without learning, we did not compare our perceptual model to other existing approaches. The heartbeat tapping task was not well-suited for the several alternatives we considered. For example, it did not have a sufficient number of trials to be suited for traditional signal detection approaches, which would be most comparable to our model-based precision measure (105). Computational models based on reinforcement learning were also inappropriate as the task did not include planning or learning from reward, and instead dealt mainly with uncertainty in perception. Consequently, we employed an active inference model that only includes perceptual inference. This involved discretizing timepoints where responses were considered co-occurrent (or not) with diastole vs. systole; but other such discretizations could have been chosen. That said, given the relationships we observed between parameters and other task measures, our choices appear to have led to estimates that track meaningful individual differences in task behavior. The task measurement conditions also had certain limitations. For example, many individuals had low IP values – reflecting the low cardiac awareness commonly seen at rest in previous studies (6) – which may have limited the variability necessary to assess relationships with other variables. Finally, low power, due to a relatively lower sample size in the eating disorders and anxiety disorders groups, may have prevented us from detecting some effects in these groups.

In summary, this study 1) demonstrated the sensitivity of individual difference measures (parameter estimates) derived from a novel active inference-based computational model, with a

focus on estimating the precision weighting of interoceptive signals across a transdiagnostic sample of individuals with psychiatric disorders, and 2) tested—and found evidence supporting—the hypothesis that individuals with psychiatric disorders fail to update the precision weighting of afferent interoceptive signals during homeostatic perturbations. While the underlying neurophysiological mechanisms leading to this difference remain unidentified, these results point to a potential origin of visceral dysregulation (and perhaps its influence on maladaptive behavior) across multiple psychiatric conditions. This represents an important step towards a primary goal of computational psychiatry – computationally phenotyping individuals with psychiatric disorders with the tools of computational neuroscience in hopes of using this information to guide the development of precision medicine interventions.

Software Note: All model simulations were implemented using standard routines (**spm_MDP_VB_X.m**) that are available as Matlab code in the latest version of SPM academic software: http://www.fil.ion.ucl.ac.uk/spm/.

Funding: This work has been supported in part by The William K. Warren Foundation, the National Institute of Mental Health, K23MH112949 (SSK), and the National Institute of General Medical Sciences Center Grant, P20GM121312 (MPP, SSK, JF, JLS).

Conflict of Interest: None of the authors have any conflicts of interest to disclose.

Supplementary Files

Supplementary materials. This file includes results of supplementary analyses as well as supplementary figures, as referred to in the main text.

Study data. This file includes all data used in the analyses reported in the manuscript.

References

1. Khalsa SS, Adolphs R, Cameron OG, Critchley HD, Davenport PW, Feinstein JS, et al. Interoception and Mental Health: A Roadmap. Biol Psychiatry Cogn Neurosci Neuroimaging. 2018;3(6):501-13.
2. Furman D, Waugh C, Bhattacharjee K, Thompson R, Gotlib I. Interoceptive awareness, positive affect, and decision making in major depressive disorder. Journal of affective disorders. 2013;151:780-5.
3. Terhaar J, Viola FC, Bär K-J, Debener S. Heartbeat evoked potentials mirror altered body perception in depressed patients. Clinical neurophysiology. 2012;123:1950-7.
4. Dunn BD, Stefanovitch I, Evans D, Oliver C, Hawkins A, Dalgleish T. Can you feel the beat? Interoceptive awareness is an interactive function of anxiety- and depression-specific symptom dimensions. Behav Res Ther. 2010;48(11):1133-8.
5. Pollatos O, Traut-Mattausch E, Schandry R. Differential effects of anxiety and depression on interoceptive accuracy. Depression and anxiety. 2009;26:167-73.
6. Khalsa SS, Lapidus RC. Can Interoception Improve the Pragmatic Search for Biomarkers in Psychiatry? Front Psychiatry. 2016;7:121.
7. Stewart JL, Khalsa SS, Kuplicki R, Puhl M, Investigators T, Paulus MP. Interoceptive attention in opioid and stimulant use disorder. Addict Biol. 2019:e12831.
8. Khalsa SS, Craske MG, Li W, Vangala S, Strober M, Feusner JD. Altered interoceptive awareness in anorexia nervosa: effects of meal anticipation, consumption and bodily arousal. International Journal of Eating Disorders. 2015;48(7):889-97.
9. Allen M, Levy A, Parr T, Friston KJ. In the Body's Eye: The Computational Anatomy of Interoceptive Inference. bioRxiv. 2019:603928.
10. Owens AP, Allen M, Ondobaka S, Friston KJ. Interoceptive inference: From computational neuroscience to clinic. Neurosci Biobehav Rev. 2018;90:174-83.
11. Seth A. Interoceptive inference, emotion, and the embodied self. Trends in cognitive sciences. 2013;17:565-73.
12. Seth AK, Critchley HD. Extending predictive processing to the body: Emotion as interoceptive inference. Behavioral and brain sciences. 2013;36:47-8.
13. Stephan K, Manjaly Z, Mathys C, Weber L, Paliwal S, Gard T, et al. Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression. Frontiers in human neuroscience. 2016;10:550.
14. Petzschner F, Weber L, Gard T, Stephan K. Computational Psychosomatics and Computational Psychiatry: Toward a Joint Framework for Differential Diagnosis. Biological Psychiatry. 2017;82:421-30.
15. Barrett L, Simmons W. Interoceptive predictions in the brain. Nature reviews Neuroscience. 2015;16:419-29.
16. Bastos A, Usrey W, Adams R, Mangun G, Fries P, Friston K. Canonical microcircuits for predictive coding. Neuron. 2012;76:695-711.
17. Friston K. A theory of cortical responses. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2005;360:815-36.
18. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I. Reinforcement learning in depression: A review of computational research. Neurosci Biobehav Rev. 2015;55:247-67.
19. Clark J, Watson S, Friston K. What is mood? A computational perspective. Psychological Medicine. 2018:1-8.
20. Friston K, Stephan K, Montague R, Dolan R. Computational psychiatry: the brain as a phantastic organ. The lancet Psychiatry. 2014;1:148-58.

21. Huys Q, Maia T, Frank M. Computational psychiatry as a bridge from neuroscience to clinical applications. Nature Neuroscience. 2016;19:404-13.
22. Montague P, Dolan R, Friston K, Dayan P. Computational psychiatry. Trends in Cognitive Sciences. 2012;16:72-80.
23. Moutoussis M, Shahar N, Hauser T, Dolan R. Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. Computational Psychiatry. 2017:1-21.
24. Parr T, Friston K. The Anatomy of Inference: Generative Models and Brain Structure. Frontiers in Computational Neuroscience. 2018;12:90.
25. Schwartenbeck P, Friston K. Computational Phenotyping in Psychiatry: A Worked Example. eNeuro. 2016;3:ENEURO.0049-16.2016.
26. Sharp PB, Eldar E. Computational models of anxiety: nascent efforts and future directions. Current Directions in Psychological Science. 2019;28(2):170-6.
27. Smith R, Parr T, Friston KJ. Simulating Emotions: An Active Inference Model of Emotional State Inference and Emotion Concept Learning. Front Psychol. 2019;10:2844.
28. Smith R, Lane RD, Parr T, Friston KJ. Neurocomputational mechanisms underlying emotional awareness: Insights afforded by deep active inference and their potential clinical relevance. Neurosci Biobehav Rev. 2019;107:473-91.
29. Adams R, Shipp S, Friston K. Predictions not commands: active inference in the motor system. Brain Structure and Function. 2013;218:611-43.
30. Adams R, Perrinet L, Friston K. Smooth Pursuit and Visual Occlusion: Active Inference and Oculomotor Control in Schizophrenia. PLoS ONE. 2012;7:e47502.
31. Edwards M, Adams R, Brown H, Pareés I, Friston K. A Bayesian account of 'hysteria'. Brain. 2012;135:3495-512.
32. Schwartenbeck P, FitzGerald T, Mathys C, Dolan R, Friston K. The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. Cerebral Cortex. 2015;25:3434-45.
33. Iglesias S, Mathys C, Brodersen K, Kasper L, Piccirelli M, den Ouden H, et al. Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. Neuron. 2013;80:519-30.
34. Diaconescu A, Mathys C, Weber L, Kasper L, Mauer J, Stephan K. Hierarchical prediction errors in midbrain and septum during social learning. Social Cognitive and Affective Neuroscience. 2017;12:nsw171.
35. Petzschner FH, Weber LA, Wellstein KV, Paolini G, Do CT, Stephan KE. Focus of attention modulates the heartbeat evoked potential. Neuroimage. 2019;186:595-606.
36. Smith R, Thayer JF, Khalsa SS, Lane RD. The hierarchical basis of neurovisceral integration. Neurosci Biobehav Rev. 2017;75:274-96.
37. Owens A, Friston K, Low D, Mathias C, Critchley H. Investigating the relationship between cardiac interoception and autonomic cardiac control using a predictive coding framework. Autonomic Neuroscience. 2018;210:65-71.
38. Paulus MP, Feinstein JS, Khalsa SS. An Active Inference Approach to Interoceptive Psychopathology. Annu Rev Clin Psychol. 2019;15:97-122.
39. Schandry R. Heart beat perception and emotional experience. Psychophysiology. 1981;18(4):483-8.
40. Ludwick-Rosenthal R, Neufeld RW. Heart beat interoception: a study of individual differences. Int J Psychophysiol. 1985;3(1):57-65.
41. Garfinkel S, Seth A, Barrett A, Suzuki K, Critchley H. Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. Biological Psychology. 2014;104:65-74.

42. Simmons W, Avery J, Barcalow J, Bodurka J, Drevets W, Bellgowan P. Keeping the body in mind: insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. Human brain mapping. 2013;34:2944-58.
43. Brener J, Kluvitse C. Heartbeat detection: judgments of the simultaneity of external stimuli and heartbeats. Psychophysiology. 1988;25(5):554-61.
44. Katkin ES, Morell MA, Goldband S, Bernstein GL, Wise JA. Individual differences in heartbeat discrimination. Psychophysiology. 1982;19(2):160-6.
45. Whitehead WE, Drescher VM, Heiman P, Blackwell B. Realtion of heart rate control to heartbeat perception. Biofeedback Self Regul. 1977;2(4):317-92.
46. Phillips GC, Jones GE, Rieger EJ, Snell JB. Effects of the presentation of false heart-rate feedback on the performance of two common heartbeat-detection tasks. Psychophysiology. 1999;36(4):504-10.
47. Ring C, Brener J, Knapp K, Mailloux J. Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: a cautionary tale about interoceptive awareness. Biol Psychol. 2015;104:193-8.
48. Windmann S, Schonecke OW, Frohlig G, Maldener G. Dissociating beliefs about heart rates and actual heart rates in patients with cardiac pacemakers. Psychophysiology. 1999;36(3):339-42.
49. Ring C, Brener J. Heartbeat counting is unrelated to heartbeat detection: A comparison of methods to quantify interoception. Psychophysiology. 2018;55(9):e13084.
50. Murphy J, Millgate E, Geary H, Ichijo E, Coll MP, Brewer R, et al. Knowledge of resting heart rate mediates the relationship between intelligence and the heartbeat counting task. Biol Psychol. 2018;133:1-3.
51. Khalsa SS, Rudrauf D, Sandesara C, Olshansky B, Tranel D. Bolus isoproterenol infusions provide a reliable method for assessing interoceptive awareness. Int J Psychophysiol. 2009;72(1):34-45.
52. Hassanpour MS, Yan L, Wang DJ, Lapidus RC, Arevian AC, Simmons WK, et al. How the heart speaks to the brain: neural activity during cardiorespiratory interoceptive stimulation. Philos Trans R Soc Lond B Biol Sci. 2016;371(1708).
53. Schandry R, Bestler M, Montoya P. On the relation between cardiodynamics and heartbeat perception. Psychophysiology. 1993;30(5):467-74.
54. Huang H, Thompson W, Paulus MP. Computational dysfunctions in anxiety: Failure to differentiate signal from noise. Biological psychiatry. 2017;82(6):440-6.
55. Barrett L. How emotions are made: The secret life of the brain. 2017.
56. Barrett LF, Quigley KS, Hamilton P. An active inference theory of allostasis and interoception in depression. Philos Trans R Soc Lond B Biol Sci. 2016;371(1708).
57. Critchley H, Garfinkel S. Interoception and emotion. Current Opinion in Psychology. 2017;17:7-14.
58. Smith R, Lane RD. The neural basis of one's own conscious and unconscious emotional states. Neurosci Biobehav Rev. 2015;57:1-29.
59. Craig AD. How do you feel--now? The anterior insula and human awareness. Nature Reviews Neuroscience. 2009;10:59-70.
60. Domschke K, Stevens S, Pfleiderer B, Gerlach AL. Interoceptive sensitivity in anxiety and anxiety disorders: an overview and integration of neurobiological findings. Clin Psychol Rev. 2010;30(1):1-11.
61. Yoris A, Esteves S, Couto B, Melloni M, Kichic R, Cetkovich M, et al. The roles of interoceptive sensitivity and metacognitive interoception in panic. Behav Brain Funct. 2015;11:14.
62. Bogacz R. A tutorial on the free-energy framework for modelling perception and learning. Journal of mathematical psychology. 2017;76:198-211.
63. Victor TA, Khalsa SS, Simmons WK, Feinstein JS, Savitz J, Aupperle RL, et al. Tulsa 1000: a naturalistic study protocol for multilevel assessment and outcome prediction in a large psychiatric sample. BMJ open. 2018;8(1):e016620.

64. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-13.
65. Norman SB, Cissell SH, Means-Christensen AJ, Stein MB. Development and validation of an Overall Anxiety Severity And Impairment Scale (OASIS). Depress Anxiety. 2006;23(4):245-9.
66. Staley D, el-Guebaly N. Psychometric properties of the Drug Abuse Screening Test in a psychiatric patient population. Addict Behav. 1990;15(3):257-64.
67. Morgan JF, Reid F, Lacey JH. The SCOFF questionnaire: a new screening tool for eating disorders. West J Med. 2000;172(3):164-5.
68. Aupperle RL, Paulus MP, Kuplicki R, Touthang J, Victor T, Yeh HW, et al. Web-Based Graphic Representation of the Life Course of Mental Health: Cross-Sectional Study Across the Spectrum of Mood, Anxiety, Eating, and Substance Use Disorders. JMIR Ment Health. 2020;7(1):e16919.
69. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of clinical psychiatry. 1998;59 Suppl 20:22-33;quiz 4-57.
70. Smith R, Feinstein J, Kuplicki R, Forthman KL, Stewart JL, Paulus MP, et al. Perceptual Insensitivity to the Modulation of Interoceptive Signals in Depression, Anxiety, and Substance Use Disorders. PsyArXiv. 2020.
71. DeVille DC, Kuplicki R, Stewart JL, Tulsa I, Aupperle RL, Bodurka J, et al. Diminished responses to bodily threat and blunted interoception in suicide attempters. Elife. 2020;9.
72. Canales-Johnson A, Silva C, Huepe D, Rivera-Rei Á, Noreika V, Garcia MdC, et al. Auditory Feedback Differentially Modulates Behavioral and Neural Markers of Objective and Subjective Performance When Tapping to Your Heartbeat. Cerebral Cortex. 2015;25(11):4490-503.
73. Allen J, Murray A. Age-related changes in the characteristics of the photoplethysmographic pulse shape at various body sites. Physiol Meas. 2003;24(2):297-307.
74. Ring C, Brener J. The temporal locations of heartbeat sensations. Psychophysiology. 1992;29(5):535-45.
75. Friston K, Lin M, Frith C, Pezzulo G, Hobson J, Ondobaka S. Active Inference, Curiosity and Insight. Neural Computation. 2017;29:2633-83.
76. Friston K, Parr T, de Vries B. The graphical brain: Belief propagation and active inference. Network Neuroscience. 2017;1:381-414.
77. Da Costa L, Parr T, Sajid N, Veselic S, Neacsu V, Friston K. ACTIVE INFERENCE ON DISCRETE STATE-SPACES – A SYNTHESIS. arXiv. 2020:2001.07203v2 [q-bio.NC]
78. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active Inference: A Process Theory. Neural Computation. 2017;29:1-49.
79. Parr T, Friston K. Working memory, attention, and salience in active inference. Scientific Reports. 2017;7:14678.
80. Daunizeau J, den Ouden HE, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ. Observing the observer (I): meta-bayesian models of learning and decision-making. PLoS One. 2010;5(12):e15554.
81. Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. Neuroimage. 2007;34(1):220-34.
82. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - revisited. Neuroimage. 2014;84:971-85.
83. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage. 2009;46(4):1004-17.
84. Mehling WE, Price C, Daubenmier JJ, Acree M, Bartmess E, Stewart A. The Multidimensional Assessment of Interoceptive Awareness (MAIA). PLoS One. 2012;7(11):e48230.

85. Parker J, Taylor G, Bagby R. The 20-Item Toronto Alexithymia Scale: III. Reliability and factorial validity in a community population. Journal of Psychosomatic Research. 2003;55:269-75.
86. Sandin B, Chorot P, McNally RJ. Anxiety sensitivity index: normative data and its differentiation from trait anxiety. Behav Res Ther. 2001;39(2):213-9.
87. Smith R, Killgore WDS, Lane RD. The structure of emotional experience and its relation to trait emotional awareness: A theoretical review. Emotion. 2018;18(5):670-92.
88. LeDoux JE, Brown R. A higher-order theory of emotional consciousness. Proc Natl Acad Sci U S A. 2017;114(10):E2016-E25.
89. Smith R, Lane RD, Parr T, Friston KJ. Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. Neurosci Biobehav Rev. 2019.
90. Terasawa Y, Moriguchi Y, Tochizawa S, Umeda S. Interoceptive sensitivity predicts sensitivity to the emotions of others. Cognition & emotion. 2014;28:1435-48.
91. Gu X, Hof P, Friston K, Fan J. Anterior insular cortex and emotional awareness. The Journal of comparative neurology. 2013;521:3371-88.
92. Zaki J, Davis JI, Ochsner K. Overlapping activity in anterior insula during interoception and emotional experience. NeuroImage. 2012;62:493-9.
93. Herbert B, Herbert C, Pollatos O. On the Relationship Between Interoceptive Awareness and Alexithymia: Is Interoceptive Awareness Related to Emotional Awareness? Journal of Personality. 2011;79:1149-75.
94. Smith R, Lane RD, Sanova A, Alkozei A, Smith C, Killgore WDS. Common and Unique Neural Systems Underlying the Working Memory Maintenance of Emotional vs. Bodily Reactions to Affective Stimuli: The Moderating Role of Trait Emotional Awareness. Front Hum Neurosci. 2018;12:370.
95. Lane RD, Weihs KL, Herring A, Hishaw A, Smith R. Affective agnosia: Expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. Neurosci Biobehav Rev. 2015;55:594-611.
96. Kashdan T, Barrett L, McKnight P. Unpacking Emotion Differentiation: Transforming Unpleasant Experience by Perceiving Distinctions in Negativity. Current Directions in Psychological Science. 2015;24:10-6.
97. Paulus MP, Stein MB. Interoception in anxiety and depression. Brain Struct Funct. 2010;214(5-6):451-63.
98. Paulus MP, Stewart JL. Interoception and drug addiction. Neuropharmacology. 2014;76 Pt B:342-50.
99. Paulus MP, Tapert SF, Schulteis G. The role of interoception and alliesthesia in addiction. Pharmacol Biochem Behav. 2009;94(1):1-7.
100. Verdejo-Garcia A, Clark L, Dunn BD. The role of interoception in addiction: a critical review. Neurosci Biobehav Rev. 2012;36(8):1857-69.
101. Ring C, Brener J. Influence of beliefs about heart rate and actual heart rate on heartbeat counting. Psychophysiology. 1996;33(5):541-6.
102. Desmedt O, Luminet O, Corneille O. The heartbeat counting task largely involves non-interoceptive processes: Evidence from both the original and an adapted counting task. Biol Psychol. 2018;138:185-8.
103. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, O Doherty J, Pezzulo G. Active inference and learning. Neuroscience and biobehavioral reviews. 2016;68:862-79.
104. Kleckner I, Zhang J, Touroutoglou A, Chanes L, Xia C, Simmons W, et al. Evidence for a Large-Scale Brain System Supporting Allostasis and Interoception in Humans. Nature human behaviour. 2017;1:0069.

105. Kleckner IR, Wormwood JB, Simmons WK, Barrett LF, Quigley KS. Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. Psychophysiology. 2015;52(11):1432-40.

Supplementary material

Supplementary Analyses

Group by condition interactions in other task-relevant variables

To assess potential group differences in the effect of task condition on self-reported experience and physiology, we also carried analogous analyses assessing confidence, intensity, and difficulty, as well as heart rate.

Heart beat intensity. A main effect of both condition ($F(2,1273) = 44.14$, $p < .001$) and group ($F(5,1273) = 9.9$, $p < .001$) was found on self-reported heart beat intensity, but no group by condition interaction; post-hoc Tukey comparisons indicated greater intensity in the breath-hold than guessing or no-guessing conditions (*p*s < .001), and that SUDs reported greater intensity than HCs ($p < .001$), ANX ($p = .002$), DEP ($p = .01$), DEP/ANX ($p < .001$), and eating disorders ($p < .001$).

Confidence. A main effect of both condition ($F(2,1273) = 25.11$, $p < .001$) and group ($F(5,1273) = 3.25$, $p = .006$) was found on self-reported confidence, but no group by condition interaction; post-hoc Tukey comparisons indicated greater intensity in the breath-hold than guessing or no-guessing conditions (*p*s < .001), and that eating disorders reported lower confidence than SUDs and DEP/ANX ($p = .008$ and .02, respectively).

Difficulty. A main effect of both condition ($F(2,1273) = 18.62$, $p < .001$) and group ($F(5,1273) = 5.94$, $p < .001$) was found on self-reported difficulty, but no group by condition interaction; post-hoc Tukey comparisons indicated greater difficulty in the no-guessing condition than in the

guessing or breath-hold conditions (*p*s < .001), and that SUDs reported less difficulty than DEP, DEP/ANX, and eating disorders (*ps* < .001, .007, and .007, respectively).

Heart rate. A main effect of group ($F(5,1273) = 5.60$, $p < .001$) was found on heart rate, as well as a marginal effect of condition ($F(2,1273) = 2.55$, $p = .08$), but no group by condition interaction; post-hoc Tukey comparisons indicated marginally faster heart rate in the breath-hold condition than in the no-guessing condition ($p = .07$), and that SUDs and DEP/ANX had faster heart rate than HCs (*p*s < .001and .04, respectively), and that SUDs had faster heart rate than eating disorders ($p = .04$).

Associations with symptom severity

Given the heterogeneity in our clinical sample, we ran subsequent exploratory correlational analyses with continuous scores on the clinical measures gathered, excluding the healthy comparisons, to assess whether model parameters in the heartbeat tapping conditions might provide additional information about symptom severity. The only notable (but weak) relationships observed were as follows. In the no-guessing condition, IP was negatively associated with both depression (PHQ; $r = -.13$, $p = .01$) and anxiety (OASIS; $r = -.11$, $p = .03$) severity, and positively associated with substance use severity (DAST; $r = .10$, $p = .04$); pHB was also positively associated with substance use severity (DAST; $r = .17$, $p = .001$). In the breath-hold condition, pHB was positively associated with substance use severity (DAST; $r = .16$, $p = .002$). Higher anxiety (OASIS) was also associated with a slightly more anticipatory response pattern (AvR; $r = -.12$, $p = .02$).

Supplementary Figures

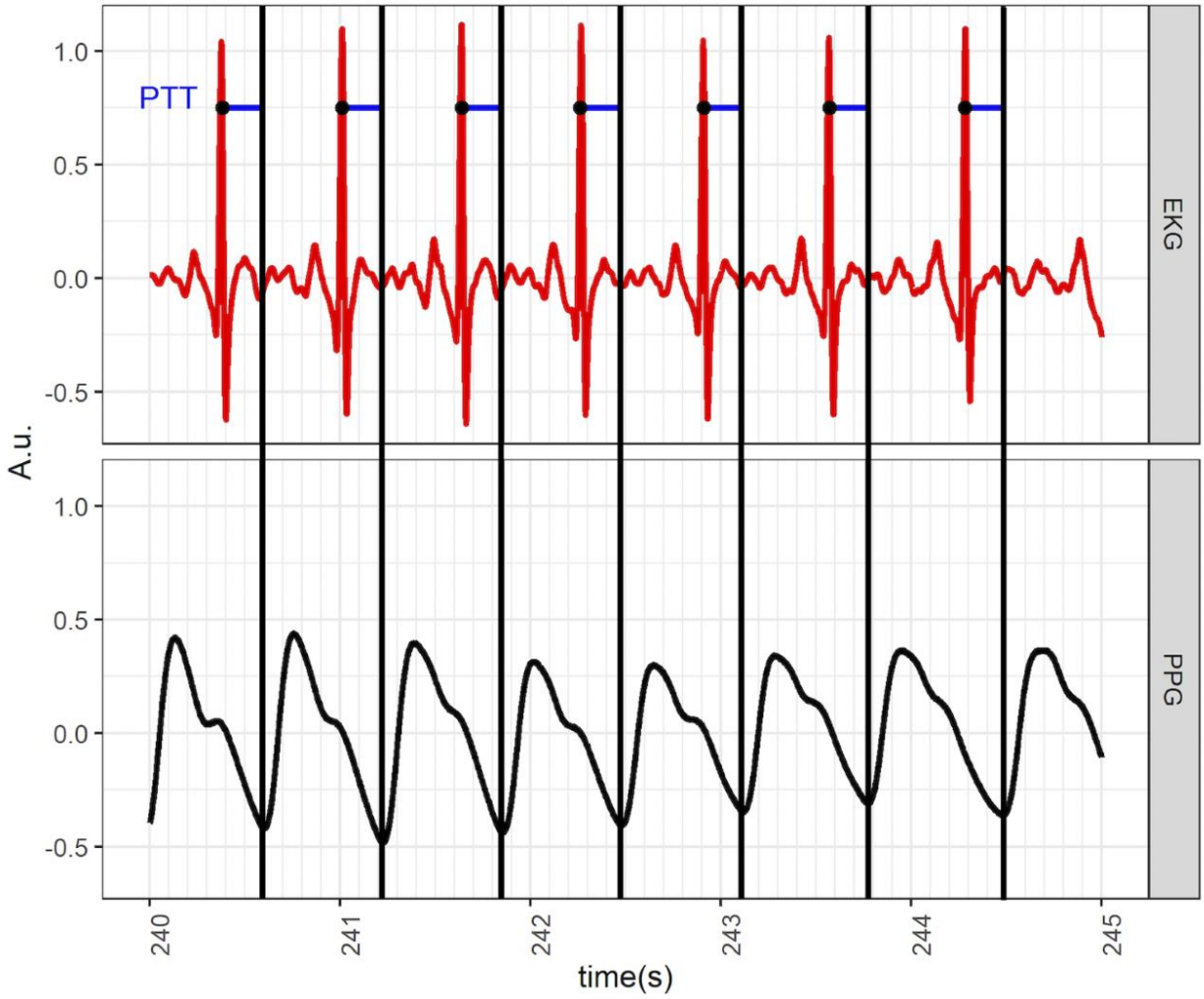


Figure S1. Top: Example EKG trace segment from one participant (red). Bottom: Simultaneous PPG trace from the same participant. Each vertical black line denotes the upswing in the PPG signal after each EKG R-spike. Each cyan line indicates the length of the delay (i.e., the pulse transit time; PTT) between each R-spike and each subsequent upswing in PPG, the median of which was calculated for each participant. These showed that a delay of 200ms was a good estimate of PTT, which is what was assumed for computational modelling.

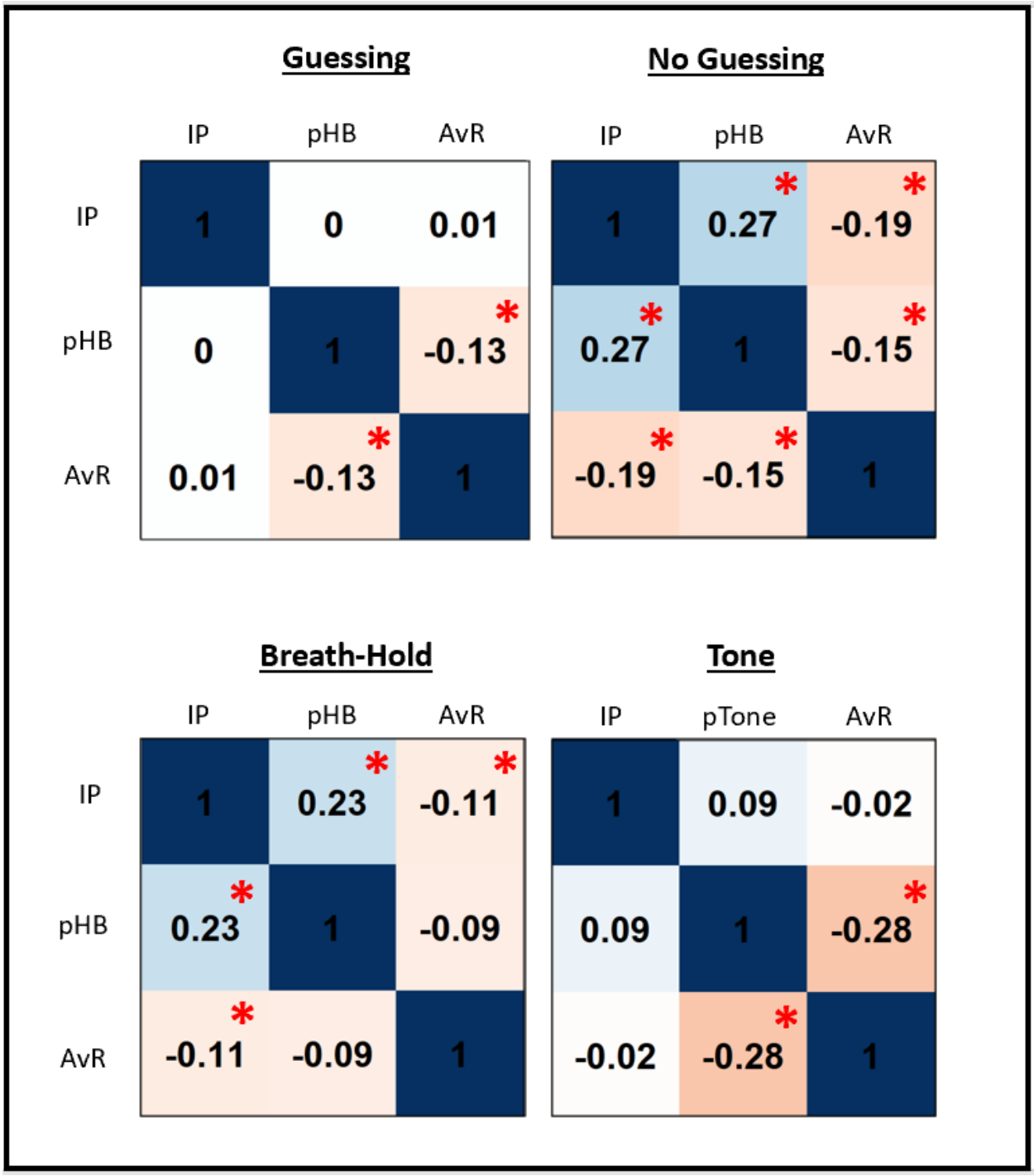


Figure S2. Correlations between model parameters by condition across all participants. For reference, correlations at $p < .05$ are marked with red asterisks.

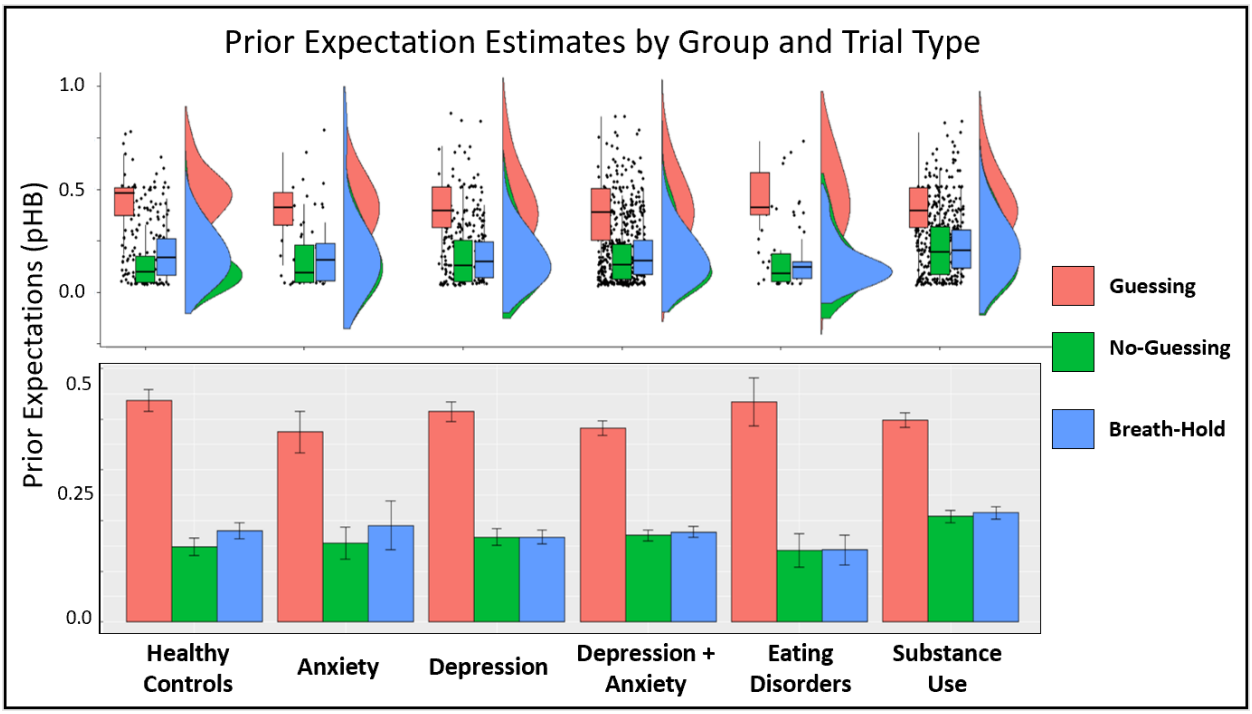


Figure S3. Bottom: Mean and standard error for prior expectation estimates by condition and clinical group. Prior expectations (pHB) were significantly higher in the guessing condition, but were not significantly different between groups in the analyses reported in the main text (although marginally greater pHB in the no-guessing and breath-hold conditions can be seen in the substance use group relative to the other groups). Top: Raincloud plots showing the same results in terms of individual datapoints, boxplots, and distributions.

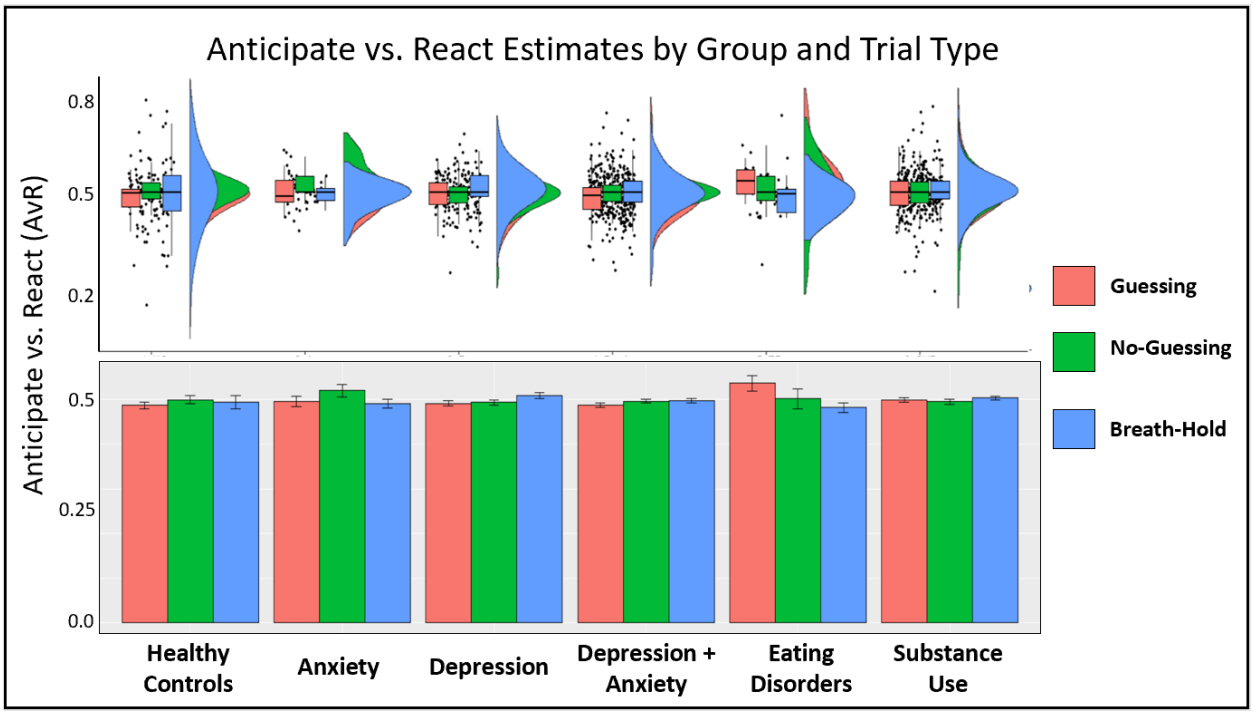


Figure S4. Bottom: Mean and standard error for Anticipate vs. React (AvR) parameter estimates by condition and clinical group. AvR was not significantly different between groups in the analyses reported in the main text. Top: Raincloud plots showing the same results in terms of individual datapoints, boxplots, and distributions.

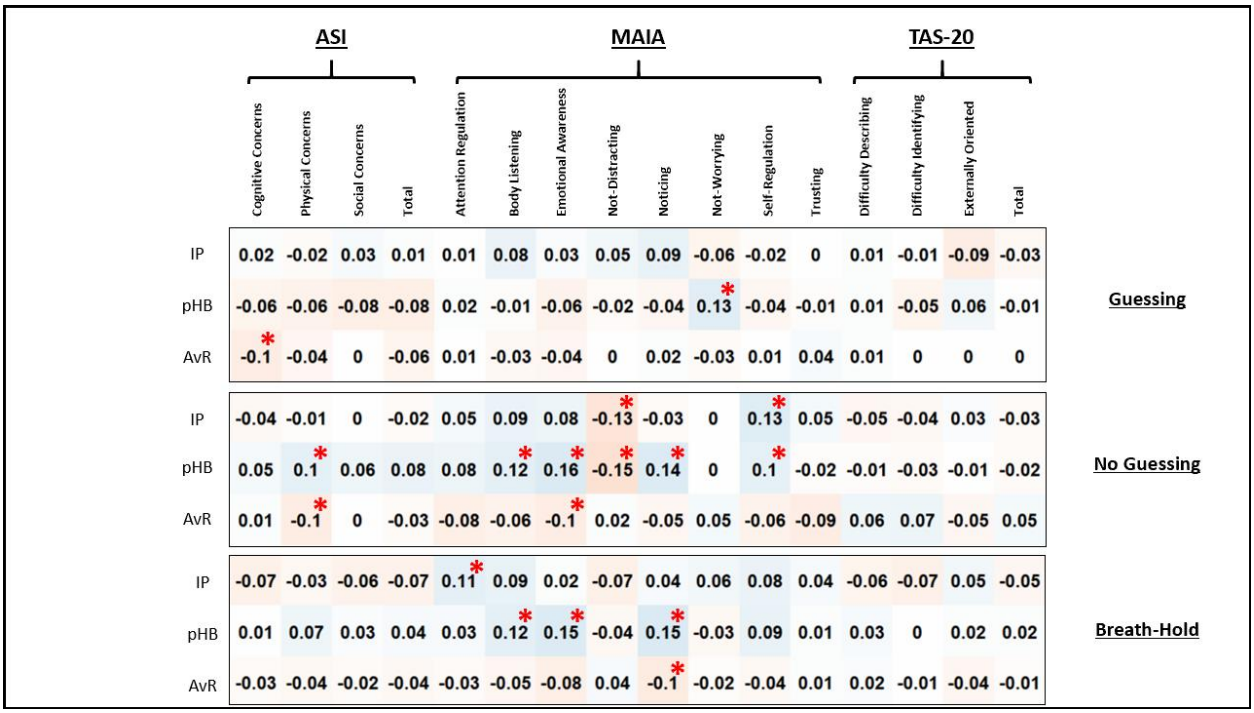


Figure S5. Correlations between model parameters by condition across all participants. For reference, correlations at $p < .05$ (uncorrected) are marked with red asterisks. IP = interoceptive precision, pHB = prior expectation for heartbeat, AvR = anticipatory vs. reactive tapping strategy parameter.

$P(s_{t=1})$

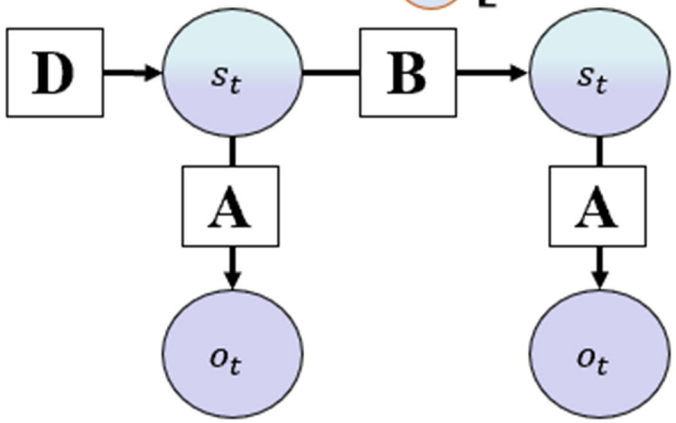
$$\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{array}{l}\text{Start}\\\text{No Heartbeat}\\\text{Heartbeat}\end{array}$$

$P(s_{t=2})$

$$\begin{bmatrix}0\\P(nHB)\\P(HB)\end{bmatrix}$$

$$\begin{array}{cc} & s_t \\ & \begin{array}{ccc}\text{Start} & \text{No Heartbeat} & \text{Heartbeat}\end{array} \\ \begin{array}{c}\text{Start}\\ s_{t+1}\end{array} & \begin{bmatrix}0 & 0 & 0\\ 1-pHB & 1 & 0\\ pHB & 0 & 1\end{bmatrix}\end{array}$$

$$\mathbf{B} = P(s_{t+1}|s_t)$$

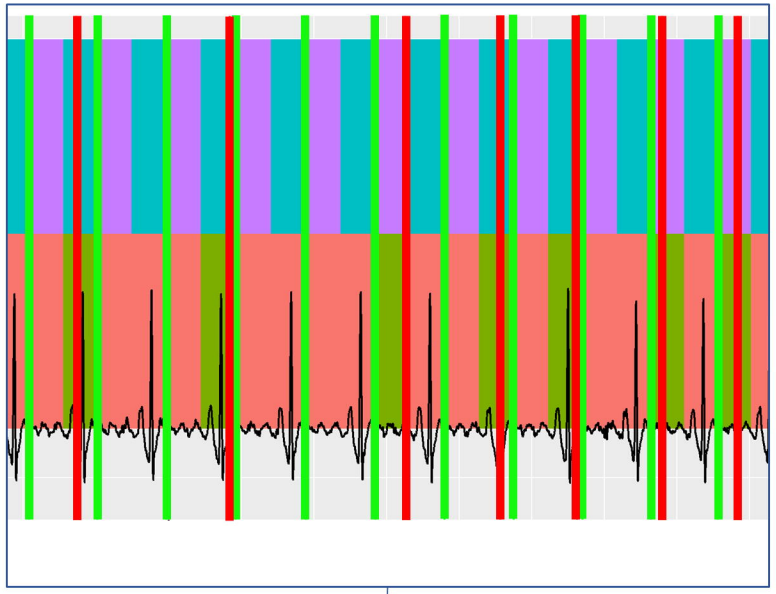


$$\begin{array}{l}\text{Start}\\\text{Diastole}\\\text{Systole}\end{array}\begin{bmatrix}1\\0\\0\end{bmatrix}\quad\begin{bmatrix}0\\1\\0\end{bmatrix}\quad\begin{bmatrix}0\\0\\1\end{bmatrix}$$

Diastole Systole

$$\begin{array}{cc} & \begin{array}{ccc}\text{Start} & \text{No Heartbeat} & \text{Heartbeat}\end{array} \\ \begin{array}{c}\text{Start}\\ \text{Diastole}\\ \text{Systole}\end{array} & \begin{bmatrix}1 & 0 & 0\\ 0 & IP & 1-IP\\ 0 & 1-IP & IP\end{bmatrix}\end{array}$$

$$\mathbf{A} = P(o_t|s_t)$$



IP = interoceptive precision
pHB = prior expectation for heartbeat

Blue = diastole
Purple = systole
Light Red = no tap
Dark Green = tap

Green vertical lines = R-wave + 200ms
Red vertical lines = tap location

Heartbeat perception:

$$\bar{s}_{t=1} = \boldsymbol{\sigma}\left(\frac{1}{2}(\ln \mathbf{D} + \ln \mathbf{B} \cdot s_{t+1}) + \ln \mathbf{A} \cdot o_t\right) \qquad \bar{s}_{t=2} = \boldsymbol{\sigma}(\ln \mathbf{B}\, s_{t-1} + \ln \mathbf{A} \cdot o_t)$$

Anticipate vs. React

	G	NG	BH	T
Guessing	1	-0.01	-0.05	-0.01
No guessing	-0.01	1	0.09	0.06
Breath-Hold	-0.05	0.09	1	0.03
Tone	-0.01	0.06	0.03	1

Prior for Heartbeat/Tone

	G	NG	BH	T
Guessing	1	0.19*	0.15	0.14
No guessing	0.19*	1	0.55*	0.05
Breath-Hold	0.15	0.55*	1	0.05
Tone	0.14	0.05	0.05	1

Sensory Precision

	Guessing	No guessing	Breath-Hold	Tone
Guessing	1	0.07	0.02	0.06
No guessing	0.07	1	0.18*	-0.01
Breath-Hold	0.02	0.18*	1	-0.02
Tone	0.06	-0.01	-0.02	1

Guessing

	PTT	# HBs	Difficulty	Confidence	Intensity	Age	BMI	Counting Accuracy
IP	-0.03	-0.02	-0.01	0.03	0.07	0.01	-0.03	0.22*
pHB	0.03	-0.13	-0.12	0.06	0.12	-0.12	0.03	0.09
AvR	0.02	-0.03	0	-0.07	-0.06	-0.01	-0.02	0.03

No Guessing

	PTT	# HBs	Difficulty	Confidence	Intensity	Age	BMI	Counting Accuracy
IP	-0.03	-0.09	-0.02	-0.08	0.24*	-0.03	-0.05	0.34*
pHB	-0.12	-0.04	-0.29*	0.16*	0.71*	0.01	0.03	0.95*
AvR	0.09	-0.08	0.05	-0.08	-0.15	-0.06	-0.05	-0.14

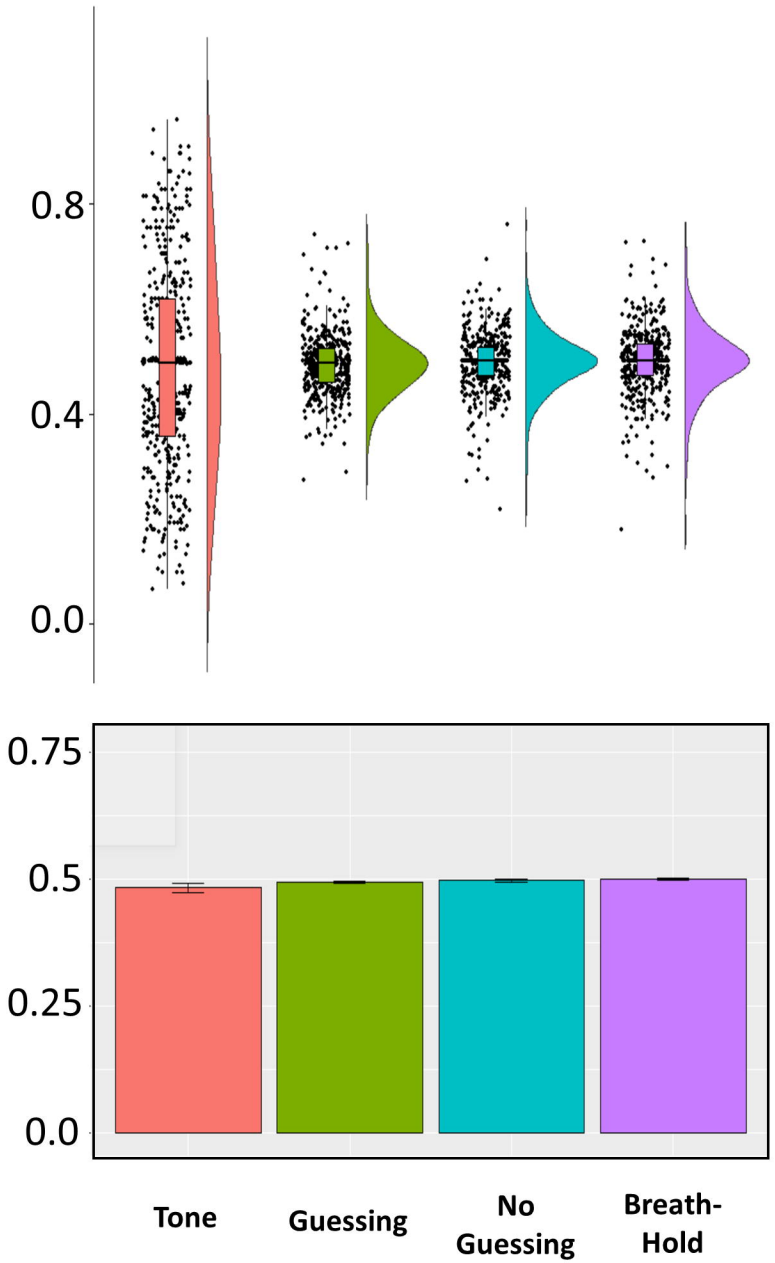
Breath-Hold

	PTT	# HBs	Difficulty	Confidence	Intensity	Age	BMI	Counting Accuracy
IP	-0.02	-0.06	-0.1	0.13	0.29*	-0.1	-0.03	0.31*
pHB	-0.03	0.03	-0.13	0.21*	0.67*	-0.14	0.02	0.87*
AvR	0.02	-0.03	0.03	-0.1	-0.14	0.01	0.02	-0.08

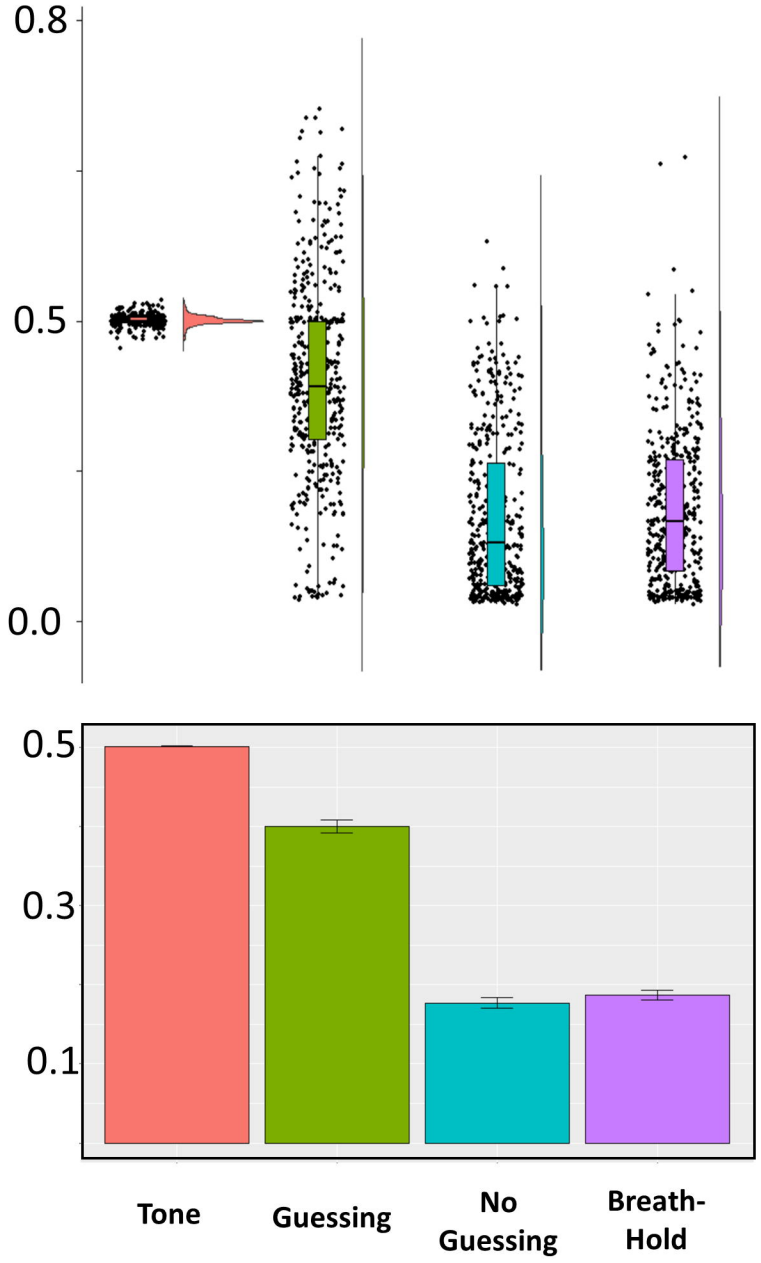
Tone

	PTT	# HBs	Difficulty	Confidence	Intensity	Age	BMI	Counting Accuracy
AP	-0.04	0.03	-0.08	-0.01	0.05	0	0.05	-0.02
pT	0	0.01	0.02	-0.07	-0.01	-0.02	0	0.24*
AvR	-0.08	0.04	-0.06	0.1	0.09	0.18*	0.09	0.05

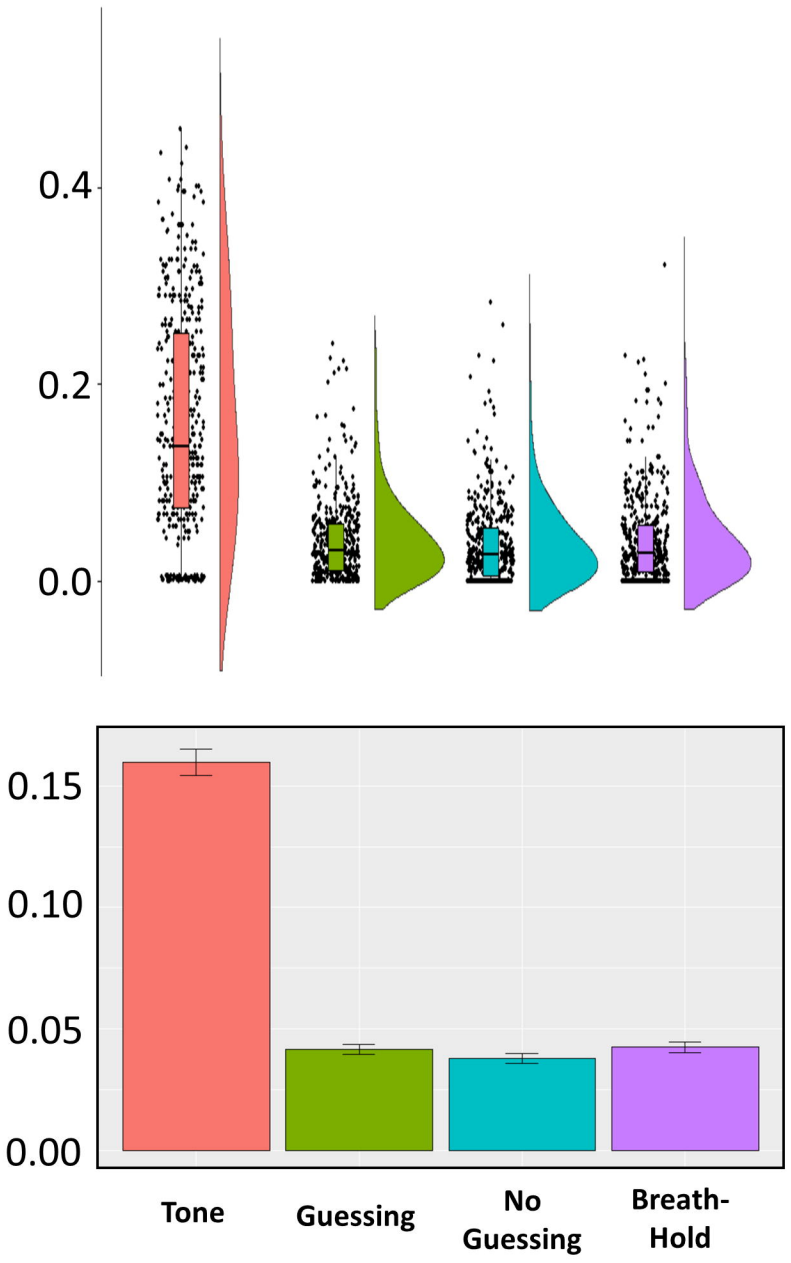
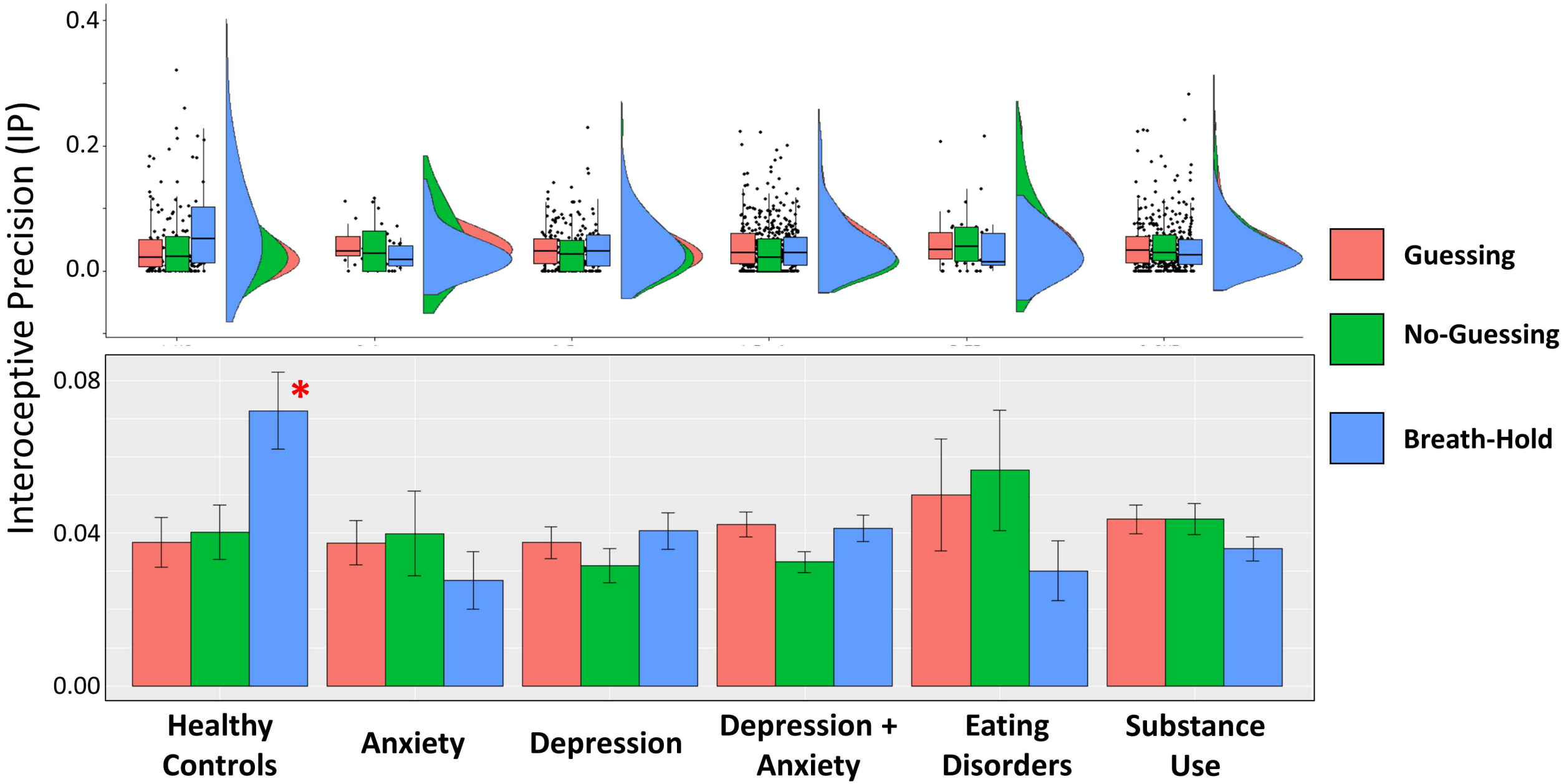
Anticipate vs. React



Prior Expectation for Heartbeat or Tone



Sensory Precision



Interoceptive Precision Estimates by Group and Trial Type