Transcutaneous vagus nerve stimulation modulates adaptive empathic learning under uncertainty

Vassilis Kotsaris[1], Dimitris Bolis[2], Ozan Cem Ozturk[1], Andrea Salaris[1,3] and Ruben T Azevedo[1]

[1]School of Psychology, University of Kent, UK

[2]Italian Institute of Technology, Italy

[3]Department of Psychology, Sapienza University of Rome, Italy

corresponding author:

Ruben T Azevedo, School of Psychology, University of Kent, UK

Email: r.a.teixeira-azevedo@kent.ac.uk

Abstract

Transcutaneous auricular vagus nerve stimulation (taVNS) has emerged as a promising tool for modulating interoception and social cognition, yet its effects on adaptive social learning remain unexplored. This study investigates the impact of taVNS on social reinforcement learning under uncertainty using an adaptive empathy task. Participants (N = 68) were assigned to either active or sham stimulation groups. During the task, participants learned the emotion regulation preferences (reappraisal or distraction) of a virtual character under stable and volatile conditions. Behavioural data were modelled using the Hierarchical Gaussian Filter (HGF), capturing learning rates and belief updates across different levels of uncertainty. While taVNS did not influence overall task accuracy or mean learning rates, it enhanced adaptability to volatile conditions, reflected in greater learning rate modulation of other's emotion regulation preferences between stable and volatile phases. Regression analyses revealed that individual differences in empathic concern and interoceptive sensibility interacted with taVNS effects, further modulating learning parameters. These findings suggest that taVNS modulates adaptive learning by enhancing sensitivity to environmental volatility. This study highlights the therapeutic potential of taVNS for modulating social cognition and adaptability to uncertainty, with implications for interpersonalised psychiatry.

Keywords: empathy, social cognition, learning, vagus nerve stimulation, Hierarchical Gaussian Filter

Several bodily functions, like respiration and cardiac activity, are controlled by the vagus nerve, i.e., the tenth cranial nerve and a major part of the parasympathetic nervous system.

Comprising approximately 75% afferent and 25% efferent fibres, the vagus nerve is essential for bidirectional communication between the brain and the body (Berthoud and Neuhuber, 2000). Through its predominantly afferent fibres, it conveys information from major internal organs to the brainstem, a central interoceptive hub (Critchley & Harrison, 2013), and from there to several other areas including the insula, hippocampus and locus coeruleus - noradrenergic system (Van Leusden et al., 2015). These brain networks are critically implicated in social attention, learning, emotion processing, empathy and interoception - the perception and regulation of internal body states (Critchley & Harrison, 2013; Critchley & Garfinkel, 2017; Singer et al., 2009).

Leveraging these insights, non-invasive methods for stimulating the vagus nerve, specifically transcutaneous auricular vagus nerve stimulation (taVNS), were developed in the early 2000s (Ventureyra, 2000) and have since been widely used in basic and clinical research (Johnson & Wilson, 2018; Yap et al., 2020). taVNS delivers low-intensity electrical stimulation to the left auricular branch of the vagus nerve, targeting afferent pathways without interfering with efferent cardiac fibres. TaVNS has also been widely investigated as a non-invasive neuromodulation approach across clinical and translational contexts (e.g., mood, pain), with ongoing evaluation of its efficacy and mechanisms (Beekwilder & Beems, 2010; De Smet et al., 2021; Guerriero et al., 2023).Recent studies further demonstrate enhanced cardiac interoceptive processing following taVNS application, reflected in improved performance on interoceptive accuracy tasks (Villani et al., 2019) and modulation of heart-evoked potentials (HEP)—an EEG marker of cortical cardiac signal processing (Poppa et al., 2022).

Importantly, interoception has been associated with various aspects of social cognition, such emotion perception, empathic understanding and social inference (Hubner et al., 2021; Grynberg & Pollatos, 2015; Shah et al., 2017), whereas afferent interoceptive signals found to influence socio-affective processing (Arslanova et al., 2023; Azevedo et al., 2017; 2023).

Moreover, meta-analytic evidence indicates that interoceptive dimensions (e.g., accuracy and sensibility) show reliable, though modest, associations with empathy and theory of mind measures (Canino et al., 2024). Given interoception's role in social cognition, taVNS might present a valuable tool for enhancing social functioning by improving access to interoceptive states (Salaris & Azevedo, 2024).

Social interactions are inherently dynamic and require not only the understanding of others' current mental and affective states, but also the ability to flexibly adapt one's responses as these states and interpersonal contingencies evolve. Empathic exchanges, for example, unfold in real time and often involve iterative cycles of action and feedback, demanding rapid and context-sensitive adjustments to others' needs (Kozakevich Arbel et al., 2021; Shamay-Tsoory & Hertz, 2022). Such adaptive social cognition depends on learning mechanisms that support the refinement of interpersonal responses based on experience and social cues, including learning which empathic response best supports the other. Within such dynamic contexts, uncertainty arises in many forms - whether from probabilistic but stable patterns (expected uncertainty) or from environments where contingencies change more abruptly (unexpected uncertainty; Angela & Dayan, 2005; Soltani & Izquierdo, 2019). Effective social adaptation thus requires accurate estimation and updating of these uncertainties, to guide learning and behaviour in complex interpersonal settings (Diaconescu et al., 2014; FeldmanHall & Shenhav, 2019; Sevgi et al., 2020). Computational accounts propose that unexpected uncertainty (or volatility) should increase learning rates, enabling rapid updating when contingencies change.

Interoceptive and noradrenergic systems have been proposed to contribute to these adaptive processes. Afferent interoceptive signals may support adaptive predictions and intuitive decisions by informing affective salience and allostatic regulation (Pinna & Edwards, 2020; Ventura-Bort & Weymar, 2024), especially in uncertain (social) environments (Critchley &

Garfinkel, 2017; Damasio, 1996; Dunn et al., 2010; 2012; Katkin et al., 2001; Lenggenhager et al., 2013; Siegel et al., 2018). Several studies also indicate a role for interoceptive signals in learning and memory (Fouragnan et al., 2024; Pfeifer et al., 2017; Werner & Shandry, 2024), while vagus nerve stimulation (VNS) has been suggested to facilitate these processes (Bowles et al., 2022; Driskill et al., 2022; Pena et al., 2014). However, human studies examining taVNS effects on learning have produced mixed results. Specifically, taVNS improved performance in two different reward-based learning tasks (Çakır et al., 2025; Weber et al., 2021), but impaired learning in a go/no-go task (Kühnel et al., 2020). Extending this mixed picture, a taVNS study using a predictive reversal-learning paradigm found no overall improvement in reversal performance relative to sham, with only subtle and partly unexpected cue-type differences (D'Agostini et al., 2021). Thus, it remains unclear when and how reinforcement learning is affected by taVNS. Consistent with this heterogeneity, VNS is reported to influence dopaminergic (Han et al., 2018), noradrenergic (Roosevelt et al., 2006) and cholinergic systems (Bowles et al., 2022), each with distinct computational roles in uncertainty-guided cognition. Specifically, hierarchical Bayesian modelling suggests that noradrenaline is implicated in learning driven by unexpected change/volatility, acetylcholine in attributing uncertainty to within-context noise versus contextual change, and dopamine in adaptive responding based on uncertainty representations (Marshall et al., 2016). Additionally, computational and neurophysiological work further links LC–noradrenergic arousal to unexpected uncertainty/volatility processing and learning-rate regulation in dynamic environments (Nassar et al., 2012; Payzan-LeNestour et al.,2013; Vincent et al., 2019; Angela & Dayan, 2005). Despite this evidence, we do not yet know whether and how taVNS can impact social reinforcement learning, particularly under varying uncertainty conditions.

Despite evidence suggesting taVNS effects on social cognition (Colzato et al., 2017; Sellaro et al., 2018), no study has thus far investigated the impact of taVNS on social functioning under volatility. In particular, taVNS could be used to examine whether it can influence social reinforcement learning when the relevant contingencies, such as another person's regulatory preferences, change over time, requiring volatility-sensitive adaptation. Despite the postulated importance of interoception in socio-affective functioning, it remains challenging to manipulate (the processing of) afferent interoceptive activity to individuate its impact on key processes of social learning and decision-making. taVNS can thus be a promising tool for modulating both interoceptive and neuromodulatory pathways, and in turn adaptive (social) cognition.

Here, capitalizing on evidence that vagus stimulation influences learning and social cognition potentially through interoceptive and noradrenergic pathways, we examined whether, and how, taVNS can impact adaptive decision-making in a social context under changing contingencies and volatility. Specifically, we used a reinforcement learning task of empathic responses while applying active or sham taVNS in a between-participants design. In this adaptive empathy task (AET), adapted from Kozakevich Arbel and colleagues (2021), participants have to learn the preferred emotion regulation strategy (between reappraisal and distraction) of a virtual character based on probabilistic facial-expression feedback. Critically, these preference contingencies were not stable but changed over time under different volatility phases, i.e. changing every few trials (volatile block) requiring faster learning or after longer periods of stable preference (stable block) requiring more a consistent response strategy. Behaviour was modelled using a hierarchical Bayesian learning model (the Hierarchical Gaussian Filter), where social beliefs and the related uncertainty are represented at 3 different levels (current preference, preference tendency, volatility of preference; Mathys et al., 2011; 2014). This modelling approach captures individuals learning parameters,

adaptive learning rates at multiple representation levels, prediction errors (PEs) and decision-making indices. Based on the predominant proposal that taVNS effects are mostly mediated by noradrenergic modulations (Roosevelt et al., 2006, Zhang et al., 2019), our main hypothesis was that taVNS would increase volatility-dependent adaptation of emotion regulation preferences which could potentially be reflected on overall task performance too. We also explored how individual differences in interoceptive sensibility (IS), empathy and autistic traits may moderate the effects of taVNS on learning.

Methods

Participants

At the time of study design, there were no published effect sizes for taVNS on a directly comparable adaptive empathy task or on volatility-dependent computational parameters. We therefore based our a priori recruitment target on the closest available human taVNS learning studies using between-subject designs, which generally reported effects in the medium range (Salaris and Azevedo, 2025; D'Agostini et al., 2021; Giraudier et al., 2020), and aimed to recruit approximately 72 participants. Because our primary hypothesis concerned volatility-dependent modulation of learning rather than a simple overall group difference, the most relevant inferential target is the within–between interaction in the repeated-measures analyses of learning adaptation. Using G*Power (v3.1.9; Faul et al., 2009), assuming a medium interaction effect (f=.25), a=.05, power = .90, two groups, two repeated measurements, and an estimated correlation among repeated measures of r =.27, the required total sample size is N=62. Our final analysed sample (N=69) exceeded this threshold. Analyses of moderation by individual-difference measures were exploratory and not powered to detect small effects.

Our participants were psychology students from the University of Kent. In total, 69 participants were included in our analyses [aged 18-35, (M=20.5, SD=2.5), 47 females]. Their participation was rewarded with credits in the research participation scheme of the University of Kent. All participants provided written informed consent before the beginning of the experiment. The study was approved by the School of Psychology University of Kent Ethics Committee and all safety requirements were ensured. TaVNS is a non-invasive technique that posits no major risk to the participants but to minimize any potential implications. Inclusion criteria required: 1) no history of neurological disorders, 2) no history of brain surgery, tumour, or intracranial metal implantation, 3) no known cardiovascular abnormalities, 4) no pregnancy, 5) no susceptibility to seizures or migraine, and 6) no pacemaker or other implanted devices; 7) no history of syncope; 8) no particularly irritable/sensitive skin. Before the beginning of the experimental procedure, participants were informed regarding possible adverse effects (i.e. itch, minor skin irritation and mild pain at the point of stimulation that typically disappears a few minutes after stimulation offset).

Online questionnaires

Participants completed 4 questionnaires prior to the empathy task in the lab session. All questionnaires were created and presented in Qualtrics platform (https://www.qualtrics.com/en-gb/).

Emotion Regulation Preference questions (ERP)

Participants were presented with the 60 distress scenarios, that were later presented in the AET, and instructed to report their preferred emotion regulation response if they were in these scenarios. They responded on a scale from 1 to 10, with the 1 being “always prefer the first option” (distraction) and 10 being “always prefer the second option” (reappraisal). Then, they were asked to evaluate again 10 of these scenarios under two different imagined

conditions: while experiencing economical scarcity; and while having a family member facing a severe health issue. These additional contexts were included to illustrate how contextual parameters might change emotion regulation preferences. This was done to demonstrate the potential variability in ER preferences, making the volatility of the virtual character's ER preferences in the subsequent AET and the related cover story more realistic.

Autism-Quotient (AQ)

The Autism-Quotient is a self-report questionnaire to measuring autistic traits in normal adult populations (Baron-Cohen et al, 2001). It includes 50 statements to which participants have to report their agreement in 4-point Likert scale from ‘definitely agree’ to ‘definitely disagree’. The scores can be calculated as total sum of all answers or divided into 5 dimensions, assessing: social skills, attention switching, attention to detail, communication, imagination.

Interpersonal Reactivity Index (IRI)

The IRI is a self-report questionnaire that measures difference in empathic social interaction and included 4 subscales assessing perspective taking, fantasy (imagination), empathic concern (EC) and personal distress (PD; Davis, 1980). The first two subscales assess cognitive empathy while the last two affective empathy. It includes 28 items to which participants respond in 5-point Likert-type scale from 0 (does not describes me well) to 4 (describes me very well).

Multidimensional Assessment of Interoceptive Awareness (MAIA)

The MAIA-2 assesses various aspects of self-reported interoceptive awareness, also referred to as interoceptive sensibility (MAIA-2, Mehling et al., 2018). This test includes 37 questions assessing, across 8 dimensions, the different ways people process and pay attention to their

bodily sensations. Participants are asked to indicate how often each statement applies to them generally in daily life, in a 5-point Likert scale ranging from 0 (never) to 5 (always).

Adaptive empathy task

Procedure

After completing the questionnaires at home, participants came to the lab to perform the AET while being stimulated by sham or active taVNS. In this task, adapted from Kozakevich Arbel and colleagues (2021), participants are first shown a description of a distress scenario for 3s accompanied by an image of the virtual character "Amy", displaying a facial expression indicating distress (Fig. 1A). Two empathic response options are then presented in left and right boxes on either side of the image. One option corresponds to cognitive reappraisal, i.e., suggesting a change in how the situation is interpreted to reduce its emotional impact (cognitive reframing), whereas the other corresponds to distraction, i.e., suggesting a shift of attention away from the distressing event toward another activity or topic (attentional redeployment).The positions of these responses are randomized and counterbalanced to avoid an association between response type and box side. Examples of responses include: 'We can go on a weekend trip away from the city. I have a few good places to suggest' (distraction) and 'The most important thing is that you were not there at the time and weren't hurt' (reappraisal, reframing toward safety/benign aspects). Importantly, participants are never made aware that the answers fall into these two emotion regulation categories. Participants had 12s to read the responses and make a decision, but they were only allowed to respond after the first 6s. Following their selection, feedback appeared on the screen for 3s, showing Amy's face turning happy if the correct response was chosen and sad if the response was incorrect. Then follows an inter-trial-interval (ITI) of 2s. Of note, participants were instructed that their goal was to choose the response that Amy would find most helpful, and to learn her

preferences from the feedback provided. They were explicitly told that Amy's preference could change over time (e.g., due to changing contextual circumstances such as those presented in the online ERP questionnaire), and therefore they should continuously adapt their choices based on recent feedback rather than aiming for consistency across repeated scenarios.

Participants were seated comfortably at a 60-70 cm distance from the computer monitor. The task lasted around 35mins, including two breaks. At the end, participants were asked a few debriefing questions (see supplementary material S2).

Design

Sixty different everyday distress scenarios of medium severity were selected based on the evaluation of 20 people in a pilot study. Each scenario was presented twice (with the same empathic responses), resulting in a total of 120 trials. The task began with 4 practice trials. The main task started with a stable condition block in which one empathic response (order counterbalanced across participants) was correct with a probability of ~73%. This was followed by a volatile phase consisting of 5 mini-blocks, where the probability of one strategy being correct was 80% and alternated between the two strategies every 10 trials. Finally, the second stable block followed, but with the alternative empathic response now correct in ~73% of trials (see Fig. 1B). In total, 70 trials were categorized as stable and 50 as volatile. taVNS was applied in a between-subjects single-blind design, with participants randomly assigned to the sham or active stimulation groups.

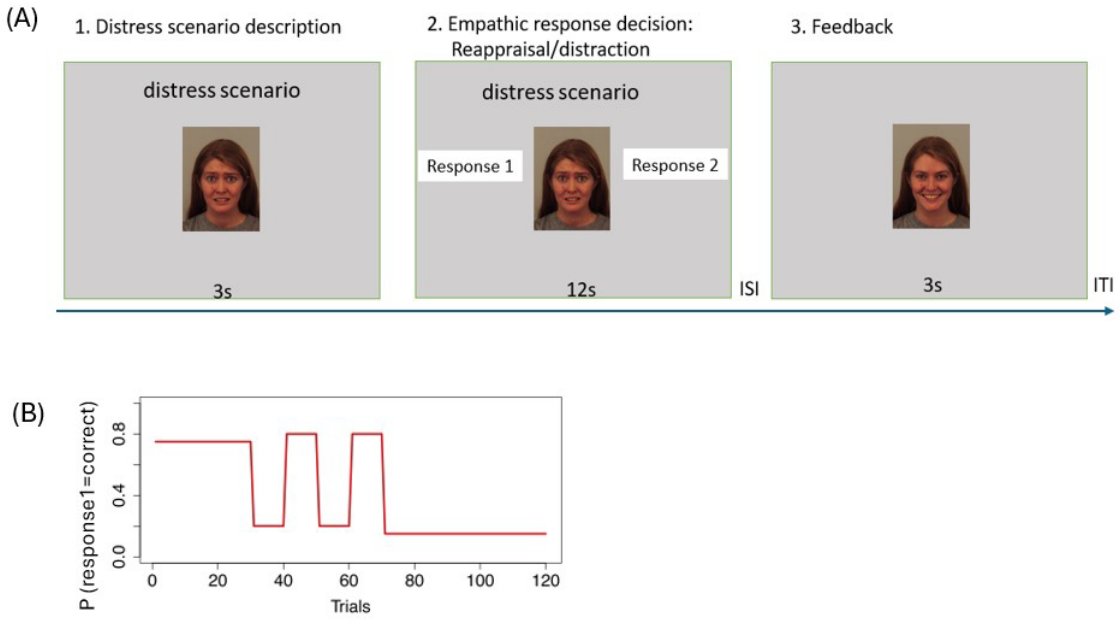

Figure 1. Experimental procedure and design. (A) Experimental design. The adaptive empathy task comprises three phases: scenario presentation; emotion regulation (ER) options and decision; feedback/outcome presentation. Participants first read the distress scenario and then select among two ER options (reappraisal and distraction) presented randomly and counterbalanced in a left and right box. If the correct option is selected a happy face is presented, otherwise a sad face is displayed. ER strategies are not labelled to avoid participants being aware of the existence of these 2 response categories. (B) Probability schedule. Probability of one of the two ER responses being the correct one, the alternative ER strategy follows the opposite probability schedule. The schedule comprised three phases: stable1 (trials 1–30), volatile (trials 31–80), and stable2 (trials 81–120). For half of the participants a reappraisal response is coded as "1" and for the other half is distraction is coded as "1", due to the counterbalanced presentation order.

Stimuli and paradigm

Our tasks were programmed and presented in Matlab (https://uk.mathworks.com/) using the Psychtoolbox (http://psychtoolbox.org/). The stimuli presented in this task were taken from the FACES database (Ebner et al., 2010). We used fearful (i.e. during the presentation of the distress scenario), happy and sad (i.e. as feedback) expressions of one white female model.

The task is a social variant of two-choice probabilistic learning that operationalizes adaptive empathy as the ability to learn, from socio-affective feedback, which interpersonal response best alleviates another person's distress (i.e., empathic response selection rather than affect sharing). In the original AET (Kozakevich Arbel et al., 2021), adaptive-empathy accuracy exceeded a non-social control condition and was uniquely associated with trait cognitive

empathy (online simulation), yet not with a measure of affective empathy, providing initial construct validity that task performance partly reflects mentalizing processes engaged during interpersonal emotion regulation. The AET shares core computations with standard reinforcement learning/reversal learning, while differing in that outcomes are social-affective (facial feedback) and the learning target is, ostensively, a person-specific preference.

Our version extends this paradigm by introducing stable and volatile phases in which the character's preferred strategy changes over time. Specifically, in the original version, there were two virtual characters, instead of one in ours, with stable ER preferences throughout the 20 trials of task, such that one ER strategy was correct for each character at 80% of times. In contrast, our task focused on a single character and incorporated volatility, requiring participants to learn the character's ER strategy and detect when it changed due to unknown contextual factors. The distress scenarios used in the original study were based on "everyday life situations related to relationships, work, daily routines, and the like". The use of the two ER strategies was decided based on the related literature suggesting that reappraisal and distraction (or expressive suppression) to be the most commonly used ER approaches. According to Gross's (2001) process model of emotion regulation, cognitive reappraisal involves changing the interpretation/meaning of an event to alter its emotional impact (cognitive change), whereas distraction involves redirecting attention away from the emotional aspects of the situation toward an alternative focus (attentional deployment).

For our task, we used the 20 distress scenarios from Kozakevich Ardel's and colleagues (2021) study and added 40 new scenarios that were tested in a pilot study. We aimed for scenarios of moderate severity that allowed both ER strategies to be perceived as plausible responses. Therefore, we created 50 distress scenarios, each paired with 50 different reappraisal and distraction responses and asked the participants of the pilot study to rate each scenario's severity and the likelihood of choosing each response. Responses were given in

10-point Likert scale from 1 (“I would always choose distraction”) to 10 (“I would always choose reappraisal”). Scenarios where the two ER strategies appeared less equiprobable (usually the most preferred one was appraisal) were excluded.

TAVNS set up and device

We used the Transcutaneous Electrical Nerve Stimulation device (V-TENS Plus; https://bodyclock.co.uk/) for stimulation with custom-built electrodes (Villani et al, 2019). For the active stimulation, the electrodes were placed on the anterior wall of the external left ear canal where the tragus is, a region considered to have high density ABVN contributions but also mixed innervation from other nerves. Sham stimulation was performed by placing the electrode on the left earlobe, as this area is free of vagal innervation (Peuker & Filler, 2002). Here, we applied continuous current stimulation with the following parameters for all participants: pulse width=250μs and frequency=25Hz. The intensity of the stimulation was adjusted across participants so that it remained just below their individual perceptual threshold. To this aim, the experimenter slowly increased the intensity level up to the level that the participant felt some sensation (e.g., tingling, itching), which was barely detected without causing any pain or discomfort. We then decreased the intensity and then increased again repeating the process 3 times to ensure the intensity stimulation level for each participant, i.e. just below perceptual awareness. The intensity was applied continuously at that specified level continuously throughout the task. The average intensity was 2.0 mA (SD=0.7) for the active stimulation group and 1.9 mA (SD=0.6) for the sham stimulation group. Previous research has shown that sub-perceptual threshold stimulation is effective as indexed, for example, by changes in pupil dilation (Phillips et al, 2025; Villani et al, 2022; Pandža et al, 2020), a proxy for LC-noradrenergic activity. This approach was adopted to avoid the side-effects of more intense stimulations, such as discomfort and distraction, which could alter the participants motivational-affective states and impact on task performance.

330 ***Computational modelling***

331 We observed how participants learned to recognize the preferred emotion regulation (ER)
332 strategy of a virtual character, using clear and immediate feedback to allow them to associate
333 their actions with outcomes. However, the relationship between empathic actions and
334 emotional outcomes was probabilistic and changed throughout the task, presenting
335 participants with limited information to interpret social cues under varying conditions of
336 uncertainty. By accumulating available information (e.g. the distress scenario and the target's
337 previous ER strategy preferences), participants could infer the character's unobserved mental
338 states (i.e. their preferred ER strategy). To examine how participants utilized the available
339 social information to learn and make decisions in our AET, we used a hierarchical Bayesian
340 learning model (HGF; Mathys et al., 2011; 2014). Such Bayesian models provide a formal
341 framework for modelling the evolution of an observer's beliefs about another's mental states
342 over time. Since each belief state depends on the previous one, this approach can be viewed
343 as a partially observable Markov decision process (POMDP), which defines the relationship
344 between observed environmental states and the unobserved mental states of the other (Baker
345 et al., 2011).

346 The HGF is a generic Bayesian learning model that quantifies belief states and their
347 associated uncertainties across multiple levels (Fig.2). Each belief state is represented as a
348 gaussian distribution, where the mean value represents the current belief state, while the
349 variance reflects the uncertainty around this belief, with precision ($\pi_i^{(k)}=1/\sigma_i^{(k)}$) being the
350 inverse variance. In our 3-level HGF model, each level is denoted as x_1^k, x_2^k, x_3^k, where k is
351 the trial number. The model's lowest level x_1 represents outcome uncertainty (or irreducible
352 uncertainty) of the effectiveness of the ER responses coded as binary categories (e.g. 1 for
353 reappraisal being the correct response, 0 for distraction). In the HGF version implemented
354 here, these beliefs are stochastically mapped to corresponding binary inputs with reappraisal

coded as $u^k = 1$ and distraction as $u^k = 0$. The second level x_2 represents the character's tendency to prefer one ER response over the other, which fluctuates over the task according to a predefined schedule (estimation uncertainty). The third level x_3 represents the volatility of these preferences, capturing how frequently the character's ER preferences change over time (volatility uncertainty).

The model's three levels capture different forms of uncertainty to produce predictions for the specific social environment. The three representation levels are hierarchically coupled so that each one is influenced by the one above it. Specifically, the first level is coupled to the second according to the sigmoid function of Equation 1:

$$\hat{x}_1^{(k)} = s\left(x_2^{(k-1)}\right) = \frac{1}{1+exp(-x_2^{(k-1)})} \quad \text{(Equation 1)}$$

At the second and third level, beliefs evolve as Gaussian random walks in the following way. The second level's step size is controlled by the third level through the equation: $x_2^k \sim N(x_2^{k-1}, \exp(\kappa x_3 + \omega_2))$, and similarly, the third level's step size is determined by $x_3^k \sim N(x_3^{k-1}, \exp(\omega_3))$. The three parameters κ, ω_2, and ω_3 capture individual differences in belief updating within this social context: κ represents the influence of volatility at the third level on the belief updating rate at the second level; ω_2 is the tonic volatility at level two, reflecting beliefs about how easily ER preferences may change; and ω_3, or 'meta-volatility', captures beliefs about environmental volatility, with higher values indicating a greater readiness to update x_3 in response to perceived instability. Here, the parameter κ is fixed to 1, defining an invariable relationship between the second and third level across participants.

In this framework, individuals perform trial-by-trial belief updating by calculating posterior probabilities, applying an approximation of ideal Bayesian inference. Here, a mean-field approximation is applied, yielding the following update equations:

$$\Delta\mu_i^{(k)} = \mu_i^{(k)} - \mu_i^{(k-1)} \propto \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} \quad \text{(Equation 2)}$$

According to Eq. 2, beliefs (posterior means) at each level (i) are proportional to the prediction error (δ_{i-1}) weighted by a precision ratio. The prediction error term quantifies the difference between the lower-level prediction $\hat{\mu}_{i-1}^{(k)}$ before the input and the posterior expectation $\mu_{i-1}^{(k)}$ afterwards. This error term is weighted by the ratio of the precision of the prediction of the lower level $\hat{\pi}_{i-1}^{(k)}$ before the input to the precision of the current belief $\pi_i^{(k)}$. Precision here refers to the inverse variance of the posterior probability distribution. This ratio of precision estimates represents the adaptive learning rate at each level, where the second level's learning rate can be transformed to reflect that of the first level.

x1	Outcome uncertainty: represents whether reappraisal or distraction is the correct response
x2	Estimation uncertainty: Belief about the virtual character's preference for an ER strategy
x3	Volatility uncertainty: how often ER preferences change
ω2	Tonic log-volatility: how easily ER preferences may change
ω3	Meta-volatility: how volatile ER preference changes are
y	Observed responses
ζ	Response model decision parameter: how deterministic is the mapping of beliefs to responses

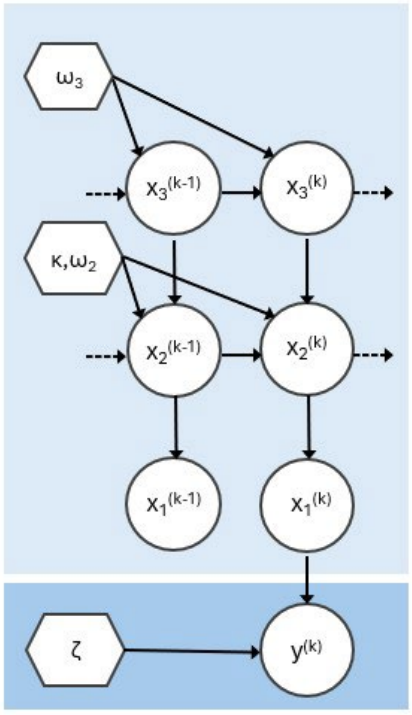

Figure 2. **Schematic of the 3-level Hierarchical Gaussian Filter (HGF; light blue) and response model (dark blue) used in the analysis**. The table on the right summarizes the key hidden states and parameters. In our task, the first level x_1 represents the binary outcome (emotion regulation choice), coded as 1 for reappraisal and 0 for distraction. The second level x_2 encodes the virtual character's preference for reappraisal, while the third level x_3 reflects the volatility of this preference over time.

Circles denote model states that evolve across time points *k*, and hexagons indicate fixed learning parameters (ω_2, ω_3). At each trial $x_2^{(k)}$ and $x_3^{(k)}$ are updated on their respective values at *k-1* modulated by ω_2 (tonic log-volatility of x_2) and ω_3 (meta-volatility of x_3) respectively. Behavioural responses ($y^{(k)}$) are generated by mapping $x_1^{(k)}$ through the response model, incorporating a decision noise parameter (ζ). Observed variables are the participant's choices and related facial feedback, whereas preference and volatility are latent states inferred from the feedback history.

The HGF serves as the perceptual model in our study and is used in conjunction with a response model. Response models are functions for mapping beliefs generated by the perceptual model to the observed responses. Here, we used as a response model the unit-square sigmoid function for binary responses (Ingesias et al, 2013), which has the following form:

$$f(y=1) = \frac{\hat{\mu}_1^{\zeta}}{\mu_1^{\zeta} + (1-\hat{\mu}_1)^{\zeta}} \quad \text{(equation 3)}$$

f(y=1) represents the probability of choosing a particular response (reappraisal) based on the belief state x_1, with values closer to 1 indicating a higher probability of selecting reappraisal. The sensitivity parameter ζ, reflects exploration tendencies and decision noise. Higher values of ζ make choices more deterministic, while lower values indicate increased randomness or exploration. This response model enables us to investigate how participants' inferred beliefs about the correct response translate into action probabilities, with the sensitivity parameter ζ adjusting the influence of these beliefs in dynamic and uncertain contexts.

In sum, the model can be read as formalising a stability–flexibility trade-off in this task. At the preference level (x_2), the participant's belief corresponds to which response (reappraisal vs distraction) is currently more likely to be preferred by the character; at the volatility level (x_3), beliefs reflect the likelihood of this preference to change over time. Behaviourally, surprising facial feedback (i.e. outcome not aligned with expectations) induces larger belief updating and thus a higher probability of switching to the alternative response on subsequent trials. However, while such surprising outcomes are best treated as noise in stable phases they

may be perceived as cues for genuine change in volatile phases. Higher second-level learning rates (LR2) mean that recent feedback has a stronger impact on beliefs about the character's current preference and therefore predict faster switching after surprising outcomes. Higher third-level learning rates (LR3) indicate stronger updating of volatility beliefs, which in turn modulates LR2 across stable versus volatile contexts. In addition, the parameters ω_2 and ω_3 capture fixed participant-specific baseline assumptions about changeability at the preference (x_2) and volatility (x_3) levels, respectively, thereby also shaping the magnitude of trialwise learning rates (LR2/LR3) and the degree to which prediction errors translate into belief updating and switching behaviour. The response parameter ζ captures choice consistency (how deterministically beliefs translate into choices), separating learning dynamics from response noise.

Model inversion

The priors of our model parameters were selected based on the experimental design and the questionnaires regarding ER preferences. We opted for relatively uninformative priors, setting large variances values to allow for adjustments based on individual differences in empathic learning and decision-making. Similarly, for parameters bounded between 0 and 1, we set the prior mean at 0.5. Details on all priors used in the HGF are response model are provided in Table S1 of the Supplemental Materials. Parameters and states were estimated in unbounded spaces, thus confined parameters (between 0-1) were log-transformed. Using the TAPAS toolbox (Frässle et al., 2021), we estimated the maximum-a-posteriori (MAP) of model parameters, given the priors provided and input sequence for each participant. Optimization was performed with a quasi-Newton algorithm, maximizing the log-joint posterior density over all perceptual and observation parameters based on the data and generative model.

Statistical analysis

With the behavioural data obtained on the empathic learning task, we first ran a 2 (taVNS group: active, sham) × 2 (Volatility Phase: stable, volatile) mixed ANOVA on accuracy to examine whether task performance varied as a function of stimulation and social volatility. As a robustness check addressing possible confounds introduced by fixed phase order (stable-volatile-stable) we imposed on our design and the trial-wise nature of accuracy, we additionally analysed trial-by-trial accuracy using mixed-effects logistic regression. As fixed effects we included stimulation, volatility phase (stable1, volatile, stable2), their interaction and the trial-wise correct strategy (reappraisal-correct vs distraction-correct). Random intercepts were added for participant and scenario: Accuracy ~ tVNS * phase + CorrectStrategy + (1|Scenario) + (1|ParticipantID).

For the computational model parameters and estimates extracted from HGF, we summarised learning-rate trajectories within each uncertainty level and experimental phase and submitted these participant-level repeated measures to repeated-measures models. Specifically, we ran a 2 (learning-rate level: LR2, LR3) × 2 (taVNS group: active, sham) mixed ANOVA on mean learning rates to test for global stimulation effects across levels. To assess whether taVNS influenced volatility sensitivity (phase-dependent modulation), we computed the difference in learning rate between volatile and stable blocks at the second and third level (ΔLR2 = LR2_volatile – LR2_stable; ΔLR3 = LR3_volatile – LR3_stable), such that positive values index stronger upregulation of updating under volatility. We submitted these two values to another 2×2 mixed ANOVA with ΔLR level (2, 3) as the within-subjects factor and taVNS group as the between-subjects factor. In addition, we ran an independent samples t-test to compare the response parameter ζ between the sham and active stimulation groups.

Moreover, as exploratory analysis we tested how individual differences affect task performance and potentially moderate taVNS effects, by running 4 linear regressions with the following DVs: the mean learning rate of the second (LR2) and third (LR3) HGF levels; the learning rate change in the second (ΔLR2) and third level (ΔLR3) between the stable and volatile phase. The predictors were the questionnaires scores and their interaction with stimulation group. Specifically, as predictors we used the affective dimensions of the IRI (i.e. EC and PD), and the total AQ and MAIA score. Emotion regulation preferences (ERP) were entered as covariate. Finally, taVNS group was entered as an interaction term. Thus, the general formula of these regressions was the following:

DV ~ (IRI_EC + IRI_PD + AQ + MAIA) * taVNS + ERP

All the above statistical analyses (ANOVAs and regressions) we conducted using the R software.

Results

taVNS effects on accuracy and on learning and response model variables

First, the 2x2 mixed ANOVA on accuracy revealed the expected main effect of *phase* ($F(1,67)=49.250$, $p<0.001$, $\eta^2p =0.424$), with higher performance in the stable phase (M=57.3, SD=8.12) compared to the volatile phase (M=49.5, SD=7.30). Conversely, no main effect of *taVNS* ($F(1,67)=0.405$, $p=0.527$, $\eta^2p =0.006$), or interaction effect between taVNS and *phase* were observed ($F(1,67)=0.541$, $p=0.465$, $\eta^2p =0.008$). This shows no general taVNS effect on performance in terms of accuracy.

As a robustness check, we also ran a mixed-effects logistic regression on trial-wise accuracy (including phase as stable1/volatile/stable2) to test whether the fixed phase order we imposed

in our task induced order confounds and to account for repeated trials within participants and scenarios. This analysis showed lower accuracy in the volatile phase relative to the first stable phase ($\beta = -0.284$, SE = 0.089, $z = -3.17$, $p = .001$), whereas accuracy did not differ between the two stable phases ($\beta = 0.041$, SE = 0.096, $z = 0.43$, $p = .670$), indicating no evidence for a generic fatigue/practice effect across the session. There was no main effect of stimulation and no stimulation-by-phase interaction in this model (ps ≥ .290). Accuracy was substantially lower on trials where reappraisal was the correct response compared to distraction-correct trials ($\beta = -0.936$, SE = 0.048, $z = -19.64$, $p < .001$), indicating a strong strategy-related asymmetry in objective trial difficulty.

To further investigate how participants' learning was affected by taVNS, we entered the mean learning rate of the second and third levels (LR2, LR3) in a 2x2 mixed ANOVA. The results showed a significant main effect of *LR* type ($F(1,67)=29.625$, $p<0.001$, $\eta^2p =0.313$) but no significant main effect of *taVNS* ($F(1,67)=1.661$, $p=0.202$, $\eta^2p =0.026$) nor interaction effect between LR type and *taVNS* ($F(1,67)=0.058$, $p=0.942$, $\eta^2p =0.001$). Therefore, no effect of taVNS on mean learning rates was observed.

We also analysed how participants modulated their learning rate according to environmental volatility by entering the learning rate difference (ΔLR2, ΔLR3) between the stable and volatile phase parameters in a 2x2 mixed ANOVA. The ANOVA showed a significant main effect of *ΔLR* type ($F(1,67)=18.957$, $p<0.001$, $\eta^2p =0.221$) and a significant main effect of *taVNS* ($F(1,67)=4.686$, $p=0.034$, $\eta^2p =0.065$). Importantly, we also found an interaction effect between *ΔLR* and taVNS ($F(1,67)=4.817$, $p=0.032$, $\eta^2p =0.067$). Pairwise comparisons (with Bonferroni corrections) showed that ΔLR2 was significantly higher for the active stimulation group (M=0.209, SD=0.357) compared to the sham stimulation group (M=0.073, SD=0.076; $t(1,66.695)=2.387$, $p=0.037$, Cohen's $d=0.544$; Fig. 3). Conversely, ΔLR3 did not differ between the active and sham stimulation groups ($t(1,66.804)=-0.2073$, $p=0.836$, Cohen's

d=0.049). These results suggest that the taVNS effect on learning pertains mainly to the learning of second-level contingencies (with a moderate to large effect size) by helping participants adapt to the virtual character's changes in ER preferences.

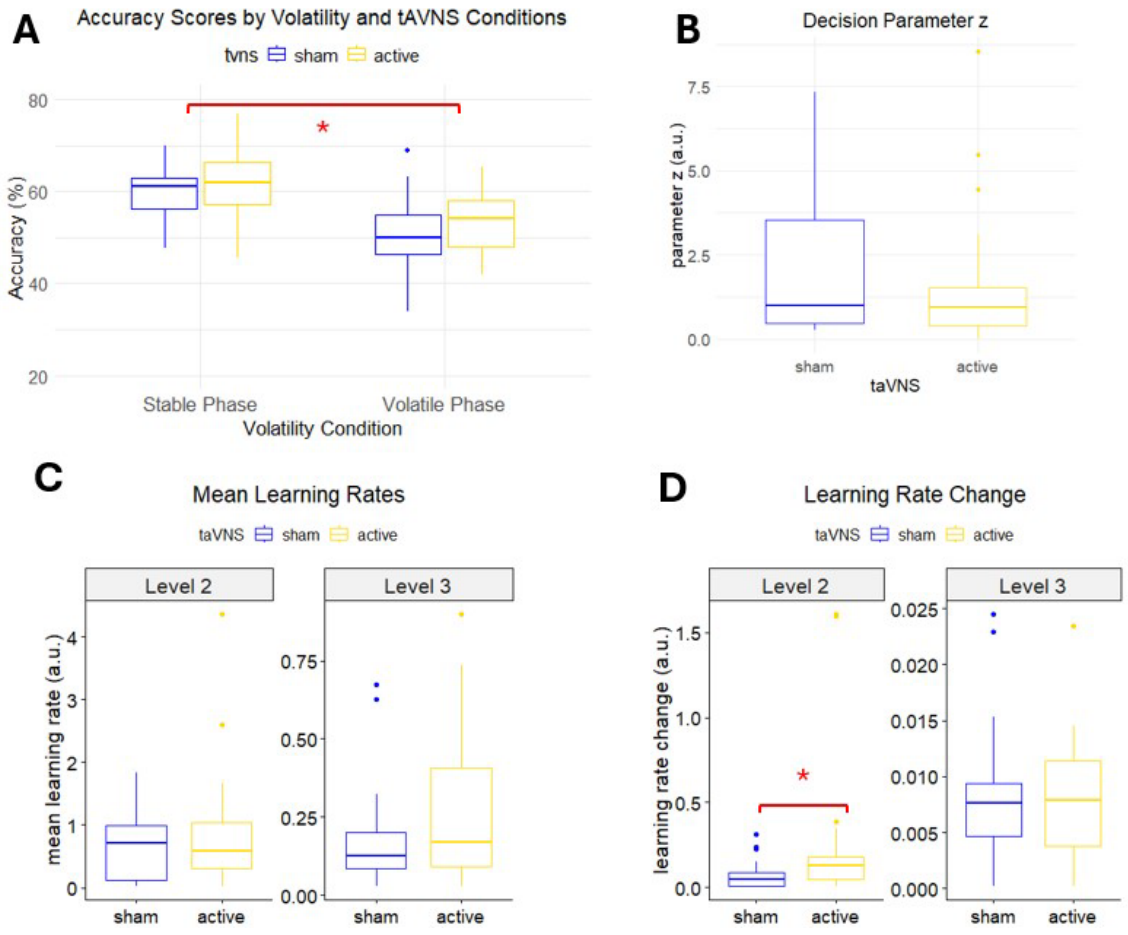

Figure 3. Group comparisons of behavioural and computational model parameters across taVNS conditions. A) Mean task accuracy for stable and volatile phases, showing a main effect of phase but no effect of taVNS or interaction. B) Inverse decision temperature parameter ζ from the response model, reflecting choice consistency, did not differ between the active and sham stimulation groups. C) Mean learning rates at the second and third levels (LR2, LR3), showing no effect of taVNS. D) Learning rate differences (ΔLR2 and ΔLR3) between volatile and stable phases. A significant interaction between ΔLR type and taVNS was found, with pairwise comparisons indicating that ΔLR2 was significantly higher in the active group compared to sham. Error bars represent standard error of the mean.

Regarding the response model results, we found that there was no difference between the active and sham stimulation for the model parameter ζ ($t(1,66.339)=0.455$, $p=0.649$, Cohen's

d=0.057). This indicates that taVNS did not influence participants' decision-making consistency, suggesting that its effects were specific to learning dynamics rather than response behaviour.

Individual differences

To explore how the learning parameters of the HGF model were related to individual traits, we ran linear regressions on the learning parameters presented above, with the taVNS group and the individual traits measured in the questionnaires as predictors. Prior to this, we conducted a series of t-tests to rule out possible initial differences between the two groups of participants for the traits measured. No significant differences were observed in any of these traits (ps > 0.493).

The regressions on mean learning rates showed that EC was the only significant predictor (t = 2.390, p = 0.022) of LR2 with a small effect size ($f^2 = 0.03$), suggesting that participants with higher EC learned the ER contingencies faster. The interaction between taVNS and EC approached significance (t = -1.866, p = 0.071) and showed a small-to-moderate effect size ($f^2 = 0.11$); all other predictors were not close to significance (ps > 0.124). Conversely, individual traits did not influence the volatility (LR3) learning rate (ps>0.296).

The regression on the learning rate difference ΔLR2 showed that it was predicted by taVNS (t = 2.423, p = 0.021), with medium effect size (f^2=0.18), confirming our ANOVA results. EC approached significance (t = 1.984, p = 0.053, $f^2 = 0.09$), well as the taVNS x AQ interaction (t = -1.980, p = 0.056, f^2=0.12), while all other predictors did not (ps>0.095).

Interestingly, and in contrast to the results observed in the ANOVA, the volatility learning rate change (ΔLR3) was significantly predicted by taVNS activation (t = 3.494, p = 0.001), showing a greater change in volatility learning rate between stable and volatile phases in the active stimulation group compared to the sham group. This effect was associated with a large

effect size ($f^2 = 0.38$), indicating a substantial impact of taVNS on belief updating about volatility when accounting for individual differences. AQ score also emerged as a significant predictor ($t = 3.620$, $p = 0.001$, $f^2 = 0.03$), with individuals higher in autistic traits showing greater updating of ΔLR3. The interaction between AQ score and taVNS was also significant ($t = -3.636$, $p = 0.001$), with a large effect size ($f^2 = 0.41$). In the sham condition, higher AQ scores were associated with more adaptive volatility learning, while this association was absent in the taVNS group. A similar moderation effect was observed for IS (MAIA), with a significant interaction between MAIA and taVNS ($t = -2.060$, $p = 0.047$) and a small-to-moderate effect size ($f^2 = 0.13$). In the sham group, MAIA scores were positively correlated with ΔLR3, whereas in the taVNS condition, this correlation was abolished. Together, these findings suggest that while individual traits such as autistic characteristics and IS are associated with increased volatility learning in more naturalistic conditions, taVNS appears to attenuate this trait dependence, promoting more uniform learning across individuals.

Regression Model	Predictor	t-value	p-value	Effect size (f^2)	Interpretation
LR2	Empathic Concern	2.390	0.029	0.03	Higher EC → Faster learning of ER preferences
ΔLR3	MAIA x taVNS	-2.060	0.047	0.13	IS positively predicted ΔLR3 mainly in the sham group
	Autism	3.620	0.001	0.03	Higher AQ → Higher volatility learning rate change (ΔLR3)

	Autism x taVNS	-3.636	0.001	0.41	AQ correlated with ΔLR3 mainly in the sham group

Table 1. Regression Results on Individual Differences. This table presents the significant regression results examining the relationship between individual traits (e.g., IS, EC, autistic traits) and learning trajectories. Predictors include questionnaire scores and their interactions with taVNS stimulation (active vs. sham). LR2 = Second-level learning rate and ΔLR3 = Learning rate change between stable and volatile phases at the third level, respectively. Significant interaction effects indicate that the relationship between individual traits and learning dynamics differ between the active and sham taVNS groups.

Discussion

Adaptive social cognition relies on the ability to flexibly update our understanding of others' internal states, preferences and intentions as social contexts evolve. This dynamic process becomes especially important in situations of uncertainty, where social and, as we show here, visceral cues must be integrated with changing contextual information to guide appropriate behaviour. Interoceptive signals have been proposed to support social cognition (Grynberg & Pollatos, 2015; Ondobaka et al., 2017; Shah et al., 2017) and decision making under uncertainty (Dunn et al., 2010; Katkin et al, 2001; Kandasamy et al, 2016), while taVNS has emerged as a promising tool to intervene in the brain-body pathways involved in these processes (Paciorek & Skora, 2020; Weng et al., 2021). Here, we demonstrate that taVNS modulates adaptive social behaviour in an empathy task where participants had to (implicitly) learn the emotion regulation preferences of a virtual character across stable and volatile phases. To examine learning under different uncertainty conditions, we applied a hierarchical Bayesian learning model (HGF; Mathys et al., 2014) to capture changes in belief updating across nested levels of inference. While taVNS did not influence task accuracy and average learning rates, it significantly modulated volatility-dependent learning flexibility, specifically increasing the adjustment of second-level learning rates between stable and volatile contexts. Moreover, taVNS modulated the relationship between individual differences in EC and IS and social learning. These results extend previous findings suggesting a positive impact of

vagus stimulation on associative learning (Weber et al., 2021) and social perception (Colzato et al., 2017), showing for the first time how taVNS can influence social learning under volatile conditions. These effects likely reflect taVNS' neuromodulatory influence on the noradrenergic system and/or its modulation of interoceptive signal processing.

While several animal studies have demonstrated a positive impact of vagal stimulation on learning and memory (Clark et al., 1998, 1999; Driskill et al., 2022; Frausto Pena et al., 2014), human research is limited and marked by inconsistent findings. For instance, Weber and colleagues (2021) reported that taVNS improved reinforcement learning by enhancing reward sensitivity and inducing a shift on accuracy-speed trade-offs towards maximizing rewards. Conversely, another study observed impaired decision-making and reduced learning rate in an approach-avoidance task following taVNS (Kühnel et al., 2020). Burger et al. (2018) found no effect of taVNS on fear extinction in humans, failing to replicate positive findings reported in animal research in rats (Alvarez-Dieppa et al., 2016). Similarly, D'Agostini et al. (2021) found no overall improvement in reversal learning following taVNS and reported no clear enhancement of putative noradrenergic markers such, despite some cue-specific behavioural effects and tentative post hoc effects on cortisol. Together, these mixed findings highlight that taVNS does not appear to exert a uniform facilitatory effect on human learning, but may instead influence specific components of adaptation depending on task demands, the structure of uncertainty, and stimulation parameters.

In our study investigating associative learning within a social context, we found no differences in total accuracy scores or average learning rates between active and sham stimulation groups. Critically, however, we observed greater modulation of the learning rate of the target's ER preferences between high and low volatility phases under active stimulation. This effect was particularly evident at the second hierarchical level, which captures beliefs about the target's current ER preference and determines how strongly recent

feedback updates these preference beliefs. Notably, the absence of group differences in overall accuracy or mean learning rates suggests that taVNS did not produce a global improvement in task performance but instead altered how participants adjusted updating as a function of environmental volatility. Such mechanistic modulation can be detectable in latent learning dynamics even when aggregate performance remains stable, because similar mean accuracy can arise from different combinations of updating and decision policies. A similar pattern also emerged for volatility-related learning rate adjustments (ΔLR3) when individual differences were taken into account. These two levels are inherently linked, as second-level learning is modulated by higher-order expectations about volatility. In social contexts, anticipated volatility determines how prediction errors are interpreted. In volatile contexts, unexpected outcomes are more likely to signal a meaningful change and promote faster updating, whereas in stable contexts they are more readily attributed to noise.

A plausible interpretation is that taVNS altered catecholaminergic regulation of belief updating under uncertainty. Animal studies with invasive VNS suggest that vagal stimulation can influence both noradrenergic (Roosevelt et al., 2006) and dopaminergic signalling (Han et al., 2018). In humans, however, direct evidence for comparable effects of auricular taVNS remains limited and mixed. There is some preliminary evidence consistent with noradrenergic engagement, especially under specific stimulation conditions. By contrast, comparable human evidence for dopaminergic modulation is still lacking. Importantly, each neuromodulatory system appears to have distinct effects on learning under volatile conditions (Angela & Dayan, 2005; Marshall et al., 2016; Sales et al., 2019; Vincent et al., 2019). Pharmacological studies using hierarchical Bayesian models indicate that noradrenergic manipulations alter learning under contextual uncertainty and volatility-linked updating, whereas dopaminergic manipulations more strongly affect the translation of higher-order beliefs into behaviour (Marshall et al., 2016; Lawson et al., 2021). Furthermore, in the context of RL, dopaminergic

influences have often been linked to reward prediction and value-guided action selection (Bayer & Glimcher, 2005; Glimcher, 2011; Jocham et al., 2011), thereby consistently implicated in adaptive behaviour. Relatedly, recent rodent work showed that VTA dopamine neuron activity encodes social prediction error and drives social reinforcement learning, suggesting that dopaminergic mechanisms may also contribute to adaptive learning in social contexts (Solié et al., 2022). In the present study, taVNS modulated volatility-dependent learning-rate adjustment without changing overall performance or response consistency. This pattern appears more compatible with an effect on precision-weighting and contextual belief updating than with a broad effect on action selection. Accordingly, our findings are more consistent with a noradrenergic contribution than with a dopaminergic one. At the same time, the present design does not allows to make strong inferences on the specific neurotransmitters involved, and overlapping noradrenergic and dopaminergic contributions cannot be excluded.

This interpretation also fits the computational demands of our adaptive empathy task. Although the AET shares core features with probabilistic reinforcement-learning and reversal-learning paradigms (D'Agostini et al., 2021; Kühnel et al., 2020), it places particular demands on ongoing volatility inference across stable and volatile social contexts, rather than on a single reversal after acquisition. This makes noradrenergic processes involved in unexpected uncertainty especially relevant. Moreover, because the outcomes are social-affective, arousal and interoceptive signals may contribute more strongly to valuation and precision-weighting than in more abstract non-social tasks or in associative tasks without reinforcement.

These taVNS effects on empathic learning may therefore also reflect changes in brain-body signalling and the use of interoceptive information during belief updating. Prior work suggests that taVNS can improve interoceptive processing (Villani et al., 2019; Poppa et al., 2022; Richter et al., 2021), enhance sensitivity to emotional and social cues (Colzato et al.,

2017; Maraver et al., 2020), and promote prosocial behaviour (Oehrn et al., 2022). More broadly, interoception has been implicated in emotion perception (Hubner et al., 2021), empathy (Fukushima et al., 2011; Grynberg & Pollatos, 2015), and emotion regulation (Zamariola et al., 2019), and has been linked to intuitive decision-making and learning under uncertainty (Dunn et al., 2010; Katkin et al., 2001; Fouragnan et al., 2024; Werner & Shandry, 2024). It has been further suggested that alignment between subjective arousal and physiological state may scaffold the updating of uncertainty representations in dynamic environments (Biddell et al., 2024). One possibility is that, by enhancing the perception (Villani et al., 2019) and integration (Salaris & Azevedo, 2025) of visceral signals, taVNS may help individuals to more precisely gauge their bodily responses in uncertain social contexts. In this sense, visceral signals may provide part of the embodied salience generated by surprising social outcomes (Allen et al., 2016; Fouragnan et al., 2024), while LC-NE modulation may regulate the precision assigned to those signals and thus how strongly they drive belief updating under volatility (Angela & Dayan, 2005; Sales et al., 2019). This could modulate the perceived reliability of internal signals, recalibrate subjective precision estimates, and thereby alter how much weight is assigned to social prediction errors. On this account, taVNS may calibrate emotional responsiveness and confidence in evolving social predictions, promoting greater flexibility in empathic learning under volatile conditions.

A parallel taVNS-related moderation emerged for volatility learning when accounting for participants' IS (MAIA). In the sham group, higher MAIA scores predicted a larger increase in ΔLR3, suggesting that individuals with heightened interoceptive abilities were more attuned to subtle contextual shifts during volatile phases. Notably, this correlation was absent in the active stimulation group, indicating that taVNS may regulate autonomic activity in a way that lessens the influence of somatic processing differences on social learning. While the positive association between MAIA and volatility learning was not directly foreseen based on

previous literature, it aligns with evidence linking interoceptive accuracy to learning and decision-making in uncertain environments (Dunn et al., 2010; Katkin et al., 2001; Kandasamy et al., 2016). The absence of a similar association with second-level learning suggests that this effect may be more specific to volatility estimation than to lower-level updating more generally.

A similar moderation effect emerged for autistic traits (AQ) in relation to volatility learning. Individuals with higher AQ scores exhibited a more pronounced increase in ΔLR3, echoing previous findings that people on the autism spectrum tend to overlearn about volatility, possibly due to an overestimation of environmental change coupled with heightened phasic noradrenaline responses (Lawson et al., 2017). Strikingly, this positive correlation was largely absent in participants receiving active taVNS, suggesting that vagal stimulation may regulate noradrenergic mechanisms in a way that reduces the impact of autistic traits on volatility-driven belief updating (Jin & Kong, 2017; Yap et al., 2020). Although autistic individuals often show lower vagal tone (Ming et al., 2005), atypical locus coeruleus activity (Chatham et al., 2022), difficulties navigating high-uncertainty contexts (Jenkinson et al., 2020), as well as aberrant belief updating (Lawson et al., 2014; Van de Cruys et al., 2014), taVNS could potentially alleviate these challenges by "optimizing" uncertainty estimations. Nevertheless, the notion of what constitutes an "optimal" volatility estimate could be heavily context-dependent. Psychopathology in autism has been reframed as not merely a suboptimal belief-updating process, but rather an interpersonal mismatch thereof (Bolis et al., 2018). As such, taVNS may offer a promising route toward an "interpersonalised psychiatry" (Bolis et al., 2023), aiming to harmonize interactions by modulating both internal autonomic states and interpersonal predictive processes.

We also found that EC correlated positively with both mean second-level learning rate (LR2) and its volatility-driven change (ΔLR2), suggesting that individuals high in EC more readily

integrate social cues when inferring others' emotion regulation preferences. These findings extend earlier work showing that empathic traits predict performance in this paradigm, but not on analogous associative learning tasks lacking social meaning, highlighting the usefulness of this paradigm to the study of adaptive social behaviour (Kozakevich Arbel et al, 2021). Our findings further suggest that people more attuned and concerned with other's feelings are better at learning others' emotion regulation preferences and more sensitive to unexpected shifts, two abilities that are crucial for flexible empathic behaviour. By contrast, EC did not predict higher-level volatility learning, indicating that this facet of affective empathy primarily supports the processing of context-specific socio-affective cues rather than the broader monitoring of environmental instability.

However, it is important to note that the interpretations of how taVNS modulates the relationship between individual differences, as measured by questionnaires, and task performance needs to be approached cautiously due to the limited sample size, which restricts our ability to draw confident conclusions. Moreover, although we found a taVNS effect on adaptive empathic learning, it was only observed in the change of learning rates in response to volatility shifts and not in other indices of overall task performance. This could be explained by a specific stimulation-dependent effect related to the update of volatility representations. However, alternative explanations cannot be excluded. For instance, performance might have been affected by the order in which the virtual character initially preferred a specific emotion regulation strategy (e.g., reappraisal vs. Distraction, see supplementary material S4). This potential confound could be more systematically assessed in future studies with larger samples, enabling more precise modelling of order effects and their interaction with stimulation.

It should also be noted that: i) there is considerable debate regarding the optimal stimulation site for taVNS, with many researchers preferring to stimulate the cymba conchae region

rather than the tragus (for discussions see: Butt et al., 2020; Verma et al., 2021); ii) inter-individual variability in vagus nerve innervation of the external ear may contribute to heterogeneity in stimulation efficacy; iii) and there is still uncertainty regarding the optimal stimulation parameters and protocol (e.g. Phillips et al., 2025; D'Agostitni et al, 2023). Moreover, we did not assess physiological proxies of vagal engagement, such as pupil diameter or LC activity, which could have provided more direct evidence of effective modulation of vagal–noradrenergic networks. Together, these limitations warrant caution in attributing the observed modulation in adaptive learning solely to vagus nerve stimulation. Future studies incorporating concurrent measures of pupil dynamics or LC-related activity would help to clarify the extent to which vagal network engagement mediates these effects.

Lastly, even if this task has been shown to be particularly sensitive to learning in the social domain (Kozakevich Arbel et al, 2021), we cannot be certain whether the taVNS effects observed here are specific to social cognition or, instead, reflect broader enhancements in associative learning. Indeed, the AET shares core computations with probabilistic reinforcement learning and can be viewed as a social and more complex variant of simpler reversal learning tasks, differing primarily in the social-affective meaning of outcomes and uncertainty structure. VNS has been shown to impact both domain-general associative or RL learning (Bowles et al., 2022; Çakır et al., 2025; Weber et al., 2021) and social engagement and perception (Colzato et al., 2017; Maraver et al., 2020). Applying a comparable task structure to a non-social learning context would help clarify the specificity of taVNS's influence on social versus general adaptive behaviour.

Conclusion

In summary, this study is the first to demonstrate a relationship between vagus nerve stimulation and adaptive learning under volatility within a social context. Specifically, taVNS

was found to modulate empathic learning flexibility, particularly in response to changes in environmental uncertainty. taVNS also influenced how individual differences in autistic and interoceptive traits shape learning behaviour. Our modelling approach highlighted specific facets of uncertainty representations affected by taVNS while suggesting potential underlying mechanisms. While these findings have promising implications for clinical and subclinical applications, future research integrating computational, behavioural and neuro-physiological methods will be essential to clarify the neurocognitive processes through which taVNS supports adaptive social learning.

Author contributions (CRediT)

Vassilis Kotsaris: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing, Project administration. **Ruben Azevedo:** Conceptualization, Methodology, Supervision, Resources, Writing – review & editing. **Ozan Cem Ozturk:** Investigation, Data Curation. **Andrea Salaris:** Investigation, Data Curation. **Dimitris Bolis:** Formal analysis, Writing – review & editing.

Acknowledgments

This research was conducted as part of VK PhD at the School of Psychology, University of Kent, and was supported by a Graduate Teaching Assistant (GTA) scholarship awarded by the university. I would like to thank Dr. Kozakevich Arbel for kindly providing the emotion regulation scenarios used in her original research, which were adapted in the present study. I also thank all participants for their time and engagement. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

References

Alvarez-Dieppa, A. C., Griffin, K., Cavalier, S., & McIntyre, C. K. (2016). Vagus nerve stimulation enhances extinction of conditioned fear in rats and modulates arc protein, CaMKII, and GluN2B-containing NMDA receptors in the basolateral amygdala. *Neural Plasticity*, *2016*(1), 4273280.

Angela, J. Y., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681-692.

D'Agostini, M., Burger, A. M., Franssen, M., Claes, N., Weymar, M., von Leupoldt, A., & Van Diest, I. (2021). Effects of transcutaneous auricular vagus nerve stimulation on reversal learning, tonic pupil size, salivary alpha-amylase, and cortisol. *Psychophysiology*, *58*(10), e13885.

Arslanova, I., Kotsaris, V., & Tsakiris, M. (2023). Perceived time expands and contracts within each heartbeat. *Current biology*, *33*(7), 1389-1395.

Azevedo, R. T., Garfinkel, S. N., Critchley, H. D., & Tsakiris, M. (2017). Cardiac afferent activity modulates the expression of racial stereotypes. *Nature Communications*, *8*(1), 13854.

Azevedo, R. T., von Mohr, M., & Tsakiris, M. (2023). From the viscera to first impressions: phase-dependent cardio-visual signals bias the perceived trustworthiness of faces. *Psychological science*, *34*(1), 120-131.

Baker, C., Saxe, R., & Tenenbaum, J. (2011). Bayesian theory of mind: Modelling joint belief-desire attribution. In Proceedings of the annual meeting of the cognitive science society (Vol. 33, No. 33).

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*, 5-17.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Beffara, B., Bret, A. G., Vermeulen, N., & Mermillod, M. (2016). Resting high frequency heart rate variability selectively predicts cooperative behavior. *Physiology & Behavior*, *164*, 417-428.

Beekwilder, J. P., & Beems, T. (2010). Overview of the clinical applications of vagus nerve stimulation. *Journal of clinical neurophysiology*, *27*(2), 130-138.

Berthoud, H. R., & Neuhuber, W. L. (2000). Functional and chemical anatomy of the afferent vagal system. *Autonomic Neuroscience*, *85*(1-3), 1-17.

Biddell, H., Solms, M., Slagter, H., & Laukkonen, R. (2024). Arousal coherence, uncertainty, and well-being: an active inference account. *Neuroscience of consciousness*, *2024*(1), niae011.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2018). Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. *Psychopathology*, *50*(6), 355-372.

Bolis, D., Dumas, G., & Schilbach, L. (2023). Interpersonal attunement in social interactions: from collective psychophysiology to inter-personalized psychiatry and beyond. *Philosophical Transactions of the Royal Society B*, *378*(1870), 20210365.

Bonaz, B., Sinniger, V., & Pellissier, S. (2019). Vagus nerve stimulation at the interface of brain–gut interactions. *Cold Spring Harbor Perspectives in Medicine*, *9*(8), a034199.

Bowles, S., Hickman, J., Peng, X., Williamson, W. R., Huang, R., Washington, K., ... & Welle, C. G. (2022). Vagus nerve stimulation drives selective circuit modulation through cholinergic reinforcement. *Neuron*, *110*(17), 2867-2885.

Burger, A. M., Van Diest, I., van der Does, W., Hysaj, M., Thayer, J. F., Brosschot, J. F., & Verkuil, B. (2018). Transcutaneous vagus nerve stimulation and extinction of prepared fear: a conceptual non-replication. *Scientific reports*, *8*(1), 11471.

Butt, M. F., Albusoda, A., Farmer, A. D., & Aziz, Q. (2020). The anatomical basis for transcutaneous auricular vagus nerve stimulation. *Journal of anatomy*, *236*(4), 588-611.

Çakır, R., Büyükgüdük, İ., Bilim, P., Erdinç, A., & Veldhuizen, M. G. (2025). Transcutaneous Vagus Nerve Stimulation Enhances Probabilistic Learning. *Psychophysiology*, *62*(3), e70037.

Canino, S., Torchia, V., Gaita, M., Raimo, S., & Palermo, L. (2024). Linking the inner and outer mental representations of the body to social cognition skills: A systematic review and meta-analysis. *Neuropsychologia*, *204*, 108989.

Chatham, C. H., Huppert, T. J., & Müller, R. A. (2022). Measures of tonic and phasic activity of the locus coeruleus—norepinephrine system in children with autism spectrum disorder: An event-related potential and pupillometry study. Autism Research, 15(10), 1792-1807.

Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A., & Jensen, R. A. (1999). Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nature neuroscience*, *2*(1), 94-98.

Clark, K. B., Smith, D. C., Hassert, D. L., Browning, R. A., Naritoku, D. K., & Jensen, R. A. (1998). Posttraining electrical stimulation of vagal afferents with concomitant vagal efferent inactivation enhances memory storage processes in the rat. *Neurobiology of learning and memory*, *70*(3), 364-373.

874 Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current opinion in*
875 *psychology*, *17*, 7-14.

876 Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron, 77(4),*
877 624-638.

878 Colzato, L. S., Sellaro, R., & Beste, C. (2017). Darwin revisited: The vagus nerve is a causal element
879 in controlling recognition of other's emotions. *Cortex*, *92*, 95-102.

880 Davis, M. H. (1980). Interpersonal reactivity index.

881 De Smet, S., Baeken, C., Seminck, N., Tilleman, J., Carrette, E., Vonck, K., & Vanderhasselt, M. A.
882 (2021). Non-invasive vagal nerve stimulation enhances cognitive emotion
883 regulation. *Behaviour research and therapy*, *145*, 103933.

884 Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., ... &
885 Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian
886 learning. *PLoS computational biology*, *10*(9), e1003810.

887 Driskill, C. M., Childs, J. E., Itmer, B., Rajput, J. S., & Kroener, S. (2022). Acute vagus nerve
888 stimulation facilitates short term memory and cognitive flexibility in rats. *Brain sciences*, *12*(9),
889 1137.

890 Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., ... & Dalgleish, T. (2010).
891 Listening to your heart: How interoception shapes emotion experience and intuitive decision
892 making. *Psychological science*, *21*(12), 1835-1844.

893 Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in
894 young, middle-aged, and older women and men: Development and validation. *Behavior*
895 *research methods*, *42*(1), 351-362.

896 Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power
897 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-
898 1160.

FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature human behaviour*, *3*(5), 426-435.

Fouragnan, E. F., Hosking, B., Cheung, Y., Prakash, B., Rushworth, M., & Sel, A. (2024). Timing along the cardiac cycle modulates neural signals of reward-based learning. *Nature Communications*, *15*(1), 2976.

Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., ... & Stephan, K. E. (2021). TAPAS: an open-source software package for translational neuromodeling and computational psychiatry. *Frontiers in psychiatry*, *12*, 680811.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, 108(supplement_3)*, 15647-15654.

Gross, J. J. (2001). Emotion regulation in adulthood: Timing is everything. *Current directions in psychological science*, *10*(6), 214-219.

Grynberg, D., & Pollatos, O. (2015). Perceiving one's body shapes empathy. *Physiology & behavior*, *140*, 54-60.

Guerriero, G., Liljedahl, S., Carlsen, H., Muñoz, M. L., Daros, A. R., Ruocco, A. C., & Steingrimsson, S. (2023). Transcutaneous vagus nerve stimulation (tVNS) to acutely reduce emotional vulnerability and improve emotional regulation in borderline personality disorder (tVNS-BPD): study protocol for a randomized, single-blind, sham-controlled trial.

Han, W., Tellez, L. A., Perkins, M. H., Perez, I. O., Qu, T., Ferreira, J., ... & de Araujo, I. E. (2018). A neural circuit for gut-induced reward. *Cell*, *175*(3), 665-678.

Hübner, A. M., Trempler, I., & Schubotz, R. I. (2022). Interindividual differences in interoception modulate behavior and brain responses in emotional inference. *NeuroImage*, *261*, 119524.

Jenkinson, Richard, Elizabeth Milne, and Andrew Thompson. "The relationship between intolerance of uncertainty and anxiety in autism: A systematic literature review and meta-analysis." *Autism* 24.8 (2020): 1933-1944.

Jin, Y., & Kong, J. (2017). Transcutaneous vagus nerve stimulation: a promising method for treatment of autism spectrum disorders. *Frontiers in Neuroscience*, *10*, 609.

Jocham, G., Klein, T. A., & Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *Journal of Neuroscience, 31(5)*, 1606-1613.

Johnson, R. L., & Wilson, C. G. (2018). A review of vagus nerve stimulation as a therapeutic intervention. *Journal of inflammation research*, 203-213.

Kandasamy, N., Garfinkel, S. N., Page, L., Hardy, B., Critchley, H. D., Gurnell, M., & Coates, J. M. (2016). Interoceptive ability predicts survival on a London trading floor. *Scientific reports*, *6*(1), 32986.

Katkin, E. S., Wiens, S., & Öhman, A. (2001). Nonconscious fear conditioning, visceral perception, and the development of gut feelings. *Psychological Science*, *12*(5), 366-370.

Kühnel, A., Teckentrup, V., Neuser, M. P., Huys, Q. J., Burrasch, C., Walter, M., & Kroemer, N. B. (2020). Stimulation of the vagus nerve reduces learning in a go/no-go reinforcement learning task. *European Neuropsychopharmacology*, *35*, 17-29.

Kogan, A., Oveis, C., Carr, E. W., Gruber, J., Mauss, I. B., Shallcross, A., ... & Keltner, D. (2014). Vagal activity is quadratically related to prosocial traits, prosocial emotions, and observer perceptions of prosociality. *Journal of personality and social psychology*, *107*(6), 1051.

Kozakevich Arbel, E., Shamay-Tsoory, S. G., & Hertz, U. (2021). Adaptive empathy: Empathic response selection as a dynamic, feedback-based learning process. *Frontiers in psychiatry*, *12*, 706474.

946 Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in*
947 *human neuroscience*, *8*, 302.

948 Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the
949 sensory environment. *Nature neuroscience*, *20*(9), 1293-1299.

950 Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N., & Rees, G. (2021). The computational,
951 pharmacological, and physiological determinants of sensory learning under uncertainty. *Current*
952 *Biology*, *31*(1), 163-172.

953 Mai, Sandra, Chung Ki Wong, Eleana Georgiou, and Olga Pollatos. "Interoception is associated with
954 heartbeat-evoked brain potentials (HEPs) in adolescents." *Biological Psychology* 137 (2018):
955 24-33.

956 Maraver, M. J., Steenbergen, L., Hossein, R., Actis-Grosso, R., Ricciardelli, P., Hommel, B., &
957 Colzato, L. S. (2020). Transcutaneous vagus nerve stimulation modulates attentional resource
958 deployment towards social cues. *Neuropsychologia*, *143*, 107465.

959 Marshall, L., Mathys, C., Ruge, D., De Berker, A. O., Dayan, P., Stephan, K. E., & Bestmann, S.
960 (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS Biology*, *14*(11),
961 e1002575.

962 Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for
963 individual learning under uncertainty. *Frontiers in human neuroscience*, *5*, 39.

964 Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., &
965 Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers*
966 *in human neuroscience*, *8*, 825.

967 Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional
968 assessment of interoceptive awareness, version 2 (MAIA-2). *PloS one*, *13*(12), e0208034.

969 Ming, X., Julu, P. O., Brimacombe, M., Connor, S., & Daniels, M. L. (2005). Reduced cardiac
970 parasympathetic activity in children with autism. *Brain and Development*, *27*(7), 509-516.

971 Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational
972 regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, *15*(7),
973 1040-1046.

974 Oehrn, C. R., Molitor, L., Krause, K., Niehaus, H., Schmidt, L., Hakel, L., ... & Weber, I. (2022).
975 Non-invasive vagus nerve stimulation in epilepsy patients enhances cooperative behavior in the
976 prisoner's dilemma task. *Scientific reports*, *12*(1), 10255.

977 Ondobaka, S., Kilner, J., & Friston, K. (2017). The role of interoceptive inference in theory of
978 mind. *Brain and cognition*, *112*, 64-68.

979 Paciorek, A., & Skora, L. (2020). Vagus nerve stimulation as a gateway to interoception. *Frontiers in*
980 *Psychology*, *11*, 1659.

981 Pandža, N. B., Phillips, I., Karuzis, V. P., O'Rourke, P., & Kuchinsky, S. E. (2020). Neurostimulation
982 and pupillometry: New directions for learning and research in applied linguistics. *Annual*
983 *Review of Applied Linguistics, 40,* 56-77.

984 Peuker, E. T., & Filler, T. J. (2002). The nerve supply of the human auricle. *Clinical Anatomy, 15(1*),
985 35–37.

986 Peña, D. F., Childs, J. E., Willett, S., Vital, A., McIntyre, C. K., & Kroener, S. (2014). Vagus nerve
987 stimulation enhances extinction of conditioned fear and modulates plasticity in the pathway
988 from the ventromedial prefrontal cortex to the amygdala. *Frontiers in behavioral*
989 *neuroscience*, *8*, 327.

990 Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent
991 prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–
992 1045.

993 Pfeifer, G., Garfinkel, S. N., van Praag, C. D. G., Sahota, K., Betka, S., & Critchley, H. D. (2017).
994 Feedback from the heart: Emotional learning and memory is controlled by cardiac cycle,
995 interoceptive accuracy and personality. *Biological Psychology*, *126*, 19-29.

Pollatos, O., Kirsch, W., & Schandry, R. (2005). Brain structures involved in interoceptive awareness and cardioafferent signal processing: a dipole source localization study. *Human brain mapping*, *26*(1), 54-64.

Roosevelt, R. W., Smith, D. C., Clough, R. W., Jensen, R. A., & Browning, R. A. (2006). Increased extracellular concentrations of norepinephrine in cortex and hippocampus following vagus nerve stimulation in the rat. *Brain research*, *1119*(1), 124-132.

Pinna, T., & Edwards, D. J. (2020). A systematic review of associations between interoception, vagal tone, and emotional regulation: Potential applications for mental health, wellbeing, psychological flexibility, and chronic conditions. *Frontiers in psychology*, *11*, 1792.

Phillips, I., Johns, M. A., Pandža, N. B., Calloway, R. C., Karuzis, V. P., & Kuchinsky, S. E. (2025). Three Hundred Hertz Transcutaneous Auricular Vagus Nerve Stimulation (taVNS) Impacts Pupil Size Non-Linearly as a Function of Intensity. *Psychophysiology, 62(2)*, e70011.

Poppa, T., Benschop, L., Horczak, P., Vanderhasselt, M. A., Carrette, E., Bechara, A., ... & Vonck, K. (2022). Auricular transcutaneous vagus nerve stimulation modulates the heart-evoked potential. *Brain Stimulation*, *15*(1), 260-269.

Richter, F., García, A. M., Rodriguez Arriagada, N., Yoris, A., Birba, A., Huepe, D., ... & Sedeño, L. (2021). Behavioral and neurophysiological signatures of interoceptive enhancements following vagus nerve stimulation. *Human brain mapping*, *42*(5), 1227-1242.

Salaris, A., & Azevedo, R. T. (2025). Investigating the modulation of gastric sensations and disposition toward food with taVNS. *Psychophysiology*, *62*(2), e14735.

Sales, A. C., Friston, K. J., Jones, M. W., Pickering, A. E., & Moran, R. J. (2019). Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: An Active Inference model. *PLoS computational biology*, *15*(1), e1006267.

Sellaro, R., de Gelder, B., Finisguerra, A., & Colzato, L. S. (2018). Transcutaneous vagus nerve stimulation (tVNS) enhances recognition of emotions in faces but not bodies. *Cortex*, *99*, 213-223.

Sevgi, M., Diaconescu, A. O., Henco, L., Tittgemeyer, M., & Schilbach, L. (2020). Social Bayes: using Bayesian modeling to study autistic trait–related differences in social cognition. *Biological Psychiatry*, *87*(2), 185-193.

Shah, P., Catmur, C., & Bird, G. (2017). From heart to mind: Linking interoception, emotion, and theory of mind. *Cortex; a journal devoted to the study of the nervous system and behavior*, *93*, 220.

Shamay-Tsoory, S. G., & Hertz, U. (2022). Adaptive empathy: a model for learning empathic responses in response to feedback. *Perspectives on Psychological Science*, *17*(4), 1008-1023.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in cognitive sciences*, *13*(8), 334-340.

Solié, C., Girard, B., Righetti, B., Tapparel, M., & Bellone, C. (2022). VTA dopamine neuron activity encodes social interaction and promotes reinforcement learning through social prediction error. *Nature neuroscience*, *25*(1), 86-97.

Stellar, J. E., Cohen, A., Oveis, C., & Keltner, D. (2015). Affective and physiological responses to the suffering of others: compassion and vagal activity. *Journal of personality and social psychology*, *108*(4), 572.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., De-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological review*, *121*(4), 649.

Van Leusden, J. W., Sellaro, R., & Colzato, L. S. (2015). Transcutaneous Vagal Nerve Stimulation (tVNS): a new neuromodulation tool in healthy humans?. *Frontiers in psychology*, *6*, 102.

Ventureyra, E. C. G. (2000). Transcutaneous vagus nerve stimulation for partial onset seizure therapy. Child's Nervous System, 16(2), 101–102.

Ventura-Bort, C., & Weymar, M. (2024). Transcutaneous auricular vagus nerve stimulation modulates the processing of interoceptive prediction error signals and their role in allostatic regulation. *Human Brain Mapping*, *45*(3), e26613.

Verma, N., Mudge, J. D., Kasole, M., Chen, R. C., Blanz, S. L., Trevathan, J. K., ... & Ludwig, K. A. (2021). Auricular vagus neuromodulation—a systematic review on quality of evidence and clinical effects. *Frontiers in Neuroscience*, *15*, 664740.

Villani, V., Tsakiris, M., & Azevedo, R. T. (2019). Transcutaneous vagus nerve stimulation improves interoceptive accuracy. *Neuropsychologia*, *134*, 107201.

Villani, V., Finotti, G., Di Lernia, D., Tsakiris, M., & Azevedo, R. T. (2022). Event-related transcutaneous vagus nerve stimulation modulates behaviour and pupillary responses during an auditory oddball task. *Psychoneuroendocrinology, 140,* 105719.

Vincent, P., Parr, T., Benrimoh, D., & Friston, K. J. (2019). With an eye on uncertainty: Modelling pupillary responses to environmental volatility. *PLoS computational biology*, *15*(7), e1007126.

Weber, I., Niehaus, H., Krause, K., Molitor, L., Peper, M., Schmidt, L., ... & Oehrn, C. R. (2021). Trust your gut: vagal nerve stimulation in humans improves reinforcement learning. *Brain communications*, *3*(2), fcab039.

Weng, H. Y., Feldman, J. L., Leggio, L., Napadow, V., Park, J., & Price, C. J. (2021). Interventions and manipulations of interoception. *Trends in neurosciences*, *44*(1), 52-62.

Werner, N. S., & Schandry, R. (2024). The Impact of Interoception on Learning, Memory, and Decision-Making. *Interoception: A Comprehensive Guide*, 151-184.

Yap, J. Y., Keatch, C., Lambert, E., Woods, W., Stoddart, P. R., & Kameneva, T. (2020). Critical review of transcutaneous vagus nerve stimulation: challenges for translation to clinical practice. *Frontiers in neuroscience*, *14*, 284.

Zamariola, G., Frost, N., Van Oost, A., Corneille, O., & Luminet, O. (2019). Relationship between interoception and emotion regulation: New evidence from mixed methods. *Journal of Affective Disorders*, *246*, 480-485.

Zhang, Y., Liu, J., Li, H., Yan, Z., Liu, X., Cao, J., ... & Kong, J. (2019). Transcutaneous auricular vagus nerve stimulation at 1 Hz modulates locus coeruleus activity and resting state functional connectivity in patients with migraine: an fMRI study. *NeuroImage: Clinical, 24*, 101971.