UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Measuring interoception

Plans, D.; Ponzo, S.; Morelli, D.; Cairo, M.; Ring, C.; Keating, C.t.; Cunningham, A.C.; Catmur, C.; Murphy, J.; Bird, G.

DOI:
10.1016/j.biopsycho.2021.108171

License:
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Plans, D, Ponzo, S, Morelli, D, Cairo, M, Ring, C, Keating, CT, Cunningham, AC, Catmur, C, Murphy, J & Bird, G 2021, 'Measuring interoception: the phase adjustment task', *Biological Psychology*, vol. 165, 108171. https://doi.org/10.1016/j.biopsycho.2021.108171

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 25. Apr. 2024

Biological Psychology 165 (2021) 108171

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Measuring interoception: The phase adjustment task

D. Plans [a,b,c,*,1], S. Ponzo [b,*,1], D. Morelli [b,d], M. Cairo [b], C. Ring [e], C.T. Keating [f], A. C. Cunningham [b], C. Catmur [g], J. Murphy [h,1], G. Bird [c,i,1]

[a] INDEX Group, Department of Science, Innovation, Technology, and Entrepreneurship, University of Exeter, United Kingdom
[b] Huma Therapeutics Ltd, London, United Kingdom
[c] Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[d] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom
[e] School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, United Kingdom
[f] School of Psychology, University of Birmingham, Birmingham, United Kingdom
[g] Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom
[h] Department of Psychology, Royal Holloway University of London, London, United Kingdom
[i] Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom

ARTICLE INFO

Keywords:
Interoception
Phase adjustment task
PAT task
Cardioception
Cardiac perception
Mental health
Depression
Anxiety
Stress
Heartbeat counting task
Heartbeat detection task
Empathy
Telemetrics

ABSTRACT

Interoception, perception of one's bodily state, has been associated with mental health and socio-emotional processes. However, several interoception tasks are of questionable validity, meaning associations between interoception and other variables require confirmation with new measures. Here we describe the novel, smartphone-based Phase Adjustment Task (PAT). Tones are presented at the participant's heart rate, but out of phase with heartbeats. Participants adjust the phase relationship between tones and heartbeats until they are synchronous. Data from 124 participants indicates variance in performance across participants which is not affected by physiological or strategic confounds. Associations between interoception and anxiety, depression and stress were not significant. Weak associations between interoception and mental health variables may be a consequence of testing a non-clinical sample. A second study revealed PAT performance to be moderately stable over one week, consistent with state effects on interoception.

1. Introduction

Interoception, the perception of the body's internal state (Craig, 2002), has been linked to various aspects of mental health. Interoception is thought to play a central role in emotion regulation (Füstös et al., 2013; Kever et al., 2015), social ability, particularly empathy (Bird & Viding, 2014; Quattrocki & Friston, 2014; Stevens et al., 2011), craving and motivation (Schmidt et al., 2013; Verdejo-Garcia et al., 2012), learning and decision making (Dunn, Galton, et al., 2010; Paulus, 2007; Paulus et al., 2009; Paulus & Stein, 2010; Paulus & Stewart, 2014; Shah et al., 2017; Werner et al., 2009), attachment (Quattrocki & Friston, 2014), and monitoring of one's own arousal, hunger/satiety, and pain (Hechler, 2021; Herbert et al., 2012; Pollatos et al., 2012; Werner et al., 2009). It is therefore unsurprising that impaired or otherwise atypical interoception has been reported in numerous psychiatric conditions, including anxiety, panic disorder, depression, schizophrenia, alcohol and substance abuse, somatoform disorders, personality disorders, and eating disorders (Ardizzi et al., 2016; Barrett et al., 2016; Khalsa et al., 2018; Mussgay et al., 1999; Naqvi & Bechara, 2010; Paulus & Stein, 2010; Schaefer et al., 2012). It has even been suggested that interoception may represent a common, transdiagnostic, vulnerability factor for psychopathology; whereby individual differences in interoception give rise to the 'P-Factor' - representing general vulnerability to develop and maintain psychiatric symptoms (Barrett & Simmons, 2015; Brewer et al., 2021; Caspi et al., 2014; Khalsa et al., 2018; Murphy et al., 2017; Quattrocki & Friston, 2014).

* Corresponding authors at: Department of Experimental Psychology, University of Oxford Woodstock Rd, Oxford, OX2 6GG, United Kingdom (D. Plans); Huma Therapeutics Ltd, London, United Kingdom (S. Ponzo).
E-mail addresses: david.plans@psy.ox.ac.uk (D. Plans), soni.ponzo@gmail.com (S. Ponzo).
[1] These authors contributed equally.

https://doi.org/10.1016/j.biopsycho.2021.108171
Received 10 February 2021; Received in revised form 10 August 2021; Accepted 12 August 2021
Available online 17 August 2021
0301-0511/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Despite great interest in interoception and its relevance for mental health, measuring interoception remains challenging. Most studies of interoception use one of two tasks; the heartbeat counting task (HCT; Dale & Anderson, 1978; Schandry, 1981) or a variant of the heartbeat detection task (HDT; Katkin et al., 1983; Whitehead et al., 1977). In the HCT, participants are asked to count the number of heartbeats they perceive in a brief period, and accuracy is determined by the correspondence between their count and the actual number of heartbeats. Despite widespread use, the HCT has been criticised on the basis that good performance may be achieved simply by knowing one's resting heart rate and using this knowledge to estimate the number of heartbeats (Windmann et al., 1999).

Concerns over the validity of the HCT have prompted increased focus on the HDT. In all HDT variants, participants are required to detect whether an external stimulus (typically a tone) occurs synchronously or asynchronously with respect to their heartbeat (Brener & Kluvitse, 1988; Whitehead et al., 1977). In the most common variant, the 2-alternative--forced-choice (2AFC) HDT, participants are presented with a tone after a delay following the heart's R-wave of either approximately 200 ms (thought to be perceived as synchronous with the heartbeat) or approximately 550 ms (thought to be perceived as asynchronous; Hickman et al., 2020). Accuracy is determined by comparing the number of trials correctly identified as synchronous or asynchronous (Whitehead et al., 1977). The advantage of the HDT over the HCT is that knowledge of one's heart rate is of no use when completing the HDT as both synchronous and asynchronous tones are presented at the same frequency as the participant's heartbeat. The problematic feature of all 2AFC HDT variants, however, is that the use of predefined delays for all participants relies on there being no individual differences in the delay following cardiac contraction with which individuals perceive an external stimulus to be synchronous with their heartbeat. This is problematic as large individual differences exist (Brener et al., 1993; Brener & Ring, 2016; Brener & Kluvitse, 1988; Clemens, 1984); in part because individuals perceive their heartbeat in different bodily locations (Ring & Brener, 1992), and more peripheral locations are thought to give rise to later perception than central locations (Brener & Kluvitse, 1988). Thus, some individuals may perceive neither delay as synchronous, and falsely be judged to be non-interoceptive (Jasper Brener et al., 1993).

Other HDT variants such as the Method of Constant Stimuli (MCS) and 6-alternative-forced-choice (6AFC) designs (Brener et al., 1993; Brener & Kluvitse, 1988; Clemens, 1984; Yates et al., 1985) do not present stimuli at delays presumed to be perceived as synchronous and asynchronous. Instead, stimuli are presented at several delays and participants indicate the delay they perceive to be synchronous. Accurate heartbeat perception is inferred from the consistency of responses (i.e., the extent to which the same delay is selected as synchronous); as a non-interoceptive participant would select delays at random across trials.

Multi-delay versions of the HDT are unaffected by heart rate knowledge and do not assume that every participant perceives their heartbeat at a predetermined delay, but they do have some limitations of their own, particularly for clinical studies. First, both versions typically involve a large number of stimulus presentations in order to compute reliable consistency scores. This can be problematic, particularly in populations characterized by low motivation, poor attention, fatigue, or demand avoidance (features that characterise several clinical populations). Second, even the longest tasks employ only six delays covering 500 ms after the onset of the R-wave, and it is possible that some participants can resolve intervals smaller than 100 ms. This is compounded when testing populations who require a high degree of precision in their responses, such as those with Autism Spectrum Disorder, who may become frustrated at the inability to select the exact delay appropriate for them. Third, to our knowledge no previous research has examined whether participants perceive their heartbeat at different bodily locations throughout the task, which may affect the delay judged to be synchronous; if an interoceptive participant perceives her heartbeat in several locations, and varies the location used to judge synchronicity across trials, her consistency would likely be reduced and she would falsely be judged non-interoceptive (i.e., even though she can feel her heartbeat at the point of testing, she would be judged to be unable to feel her heartbeat). Finally, these measures can only be administered under laboratory conditions with specialised equipment (e.g. electrocardiograms), limiting testing to populations able to travel to laboratories, and reducing the scope to test large samples.

This brief survey of methods to test cardiac interoception reveals several problems with the existing measures. Some problems make tests difficult to administer, without necessarily affecting their validity (e.g., for the MCS and 6AFC versions of the HDT), while the validity of other tests is likely compromised (HCT and 2AFC HDT). Such problems mean that the common assertion of links between interoception and socio-affective processes such as empathy and mental health, largely made on the basis of studies using the HCT and 2AFC HDT, require confirmation using novel measures. Indeed, the use of potentially problematic tests of interoception may explain inconsistent findings in the literature (alongside inconsistent measurement of key mental health variables; Desmedt et al., 2020). For example, with respect to anxiety, some studies suggest that anxiety (or certain anxiety subtypes) are associated with better interoceptive accuracy as assessed via heartbeat counting (e.g., Dunn, Stefanovitch, et al., 2010; Pollatos et al., 2009; Van der Does et al., 1997), and that heartbeat counting and panic/anxiety symptoms share genetic overlap (Eley et al., 2007). However, others have failed to observe a relationship between anxiety and HCT performance (e.g., Garfinkel, Tiley, et al., 2016; Shah et al., 2016) or have observed a negative association (e.g., Ewing et al., 2017; Lackner & Fresco, 2016). Indeed, when strict instructions are used (e.g., "count only those heartbeats about which you are sure" rather than "count all heartbeats you feel in your body") people with anxiety do not perform better on the HCT, an observation that may account for some of these discrepant findings (Ehlers et al., 1995). Studies using other methods (e.g., heartbeat detection), and across domains (e.g., respiratory), have also produced mixed results; some studies report positive associations (e.g., Critchley et al., 2004), others report no relationship (e.g., Asmundson et al., 1993; Palser et al., 2018), and some recent studies suggest that better interoceptive accuracy may be associated with less anxiety (e.g., Bogaerts et al., 2005; Ewing et al., 2017; Garfinkel, Manassei, et al., 2016; Garfinkel, Tiley, et al., 2016).

It is clear then, that in order to test theories linking interoception, empathy and mental health, and to determine the utility of atypical interoception as a biomarker for psychopathology, valid, reliable and accessible methods for assessing interoception are needed. In this paper, we present a novel measure of interoception - the Phase Adjustment Task (PAT) - designed to overcome the limitations of existing tests. Importantly, in terms of accessibility and scalability, the task requires nothing more than a smartphone, meaning participants can complete it anywhere. The smartphone camera is used to detect heartbeats (Bacciu et al., 2018; Morelli et al., 2018; Murphy, Brewer, et al., 2019), and tones are presented at the same frequency as the participant's heartbeat but out of phase with the heartbeats. Participants use a dial to advance or delay the tones until they perceive the tones to be synchronous (in-phase) with their heartbeat. Thus, the task cannot be completed using knowledge of one's own heart rate, and participants are able to adjust the phase of the tones as they like, rather than having to select a predetermined delay. This feature, plus a novel analysis method, means that multiple trials at multiple different delays are not required, and so the task is much quicker than the MCS and 6AFC approaches. The speed of the task further enhances the accessibility and inclusivity of this method, and increases its suitability for clinical studies.

The PAT classifies participants as interoceptive or not interoceptive at each of three statistical thresholds based on Bayes Factors (the relative likelihood of participants belonging to two populations, one interoceptive and one not). Here 'interoceptive' is a shorthand, used to refer to a participant who can perceive their heartbeat at the particular time of

being tested, and who can complete the task. Conversely, 'non-interoceptive' is a shorthand used to refer to a participant who cannot feel their heartbeat at the time of testing, and/or cannot complete the task successfully. The practice of adopting a categorical classification of participants is discussed further in the General Discussion, but is consistent with research using the MCS approach (Jasper Brener et al., 1993; Brener and Ring, 2016).

Thus, the aim of the first study was to develop and test the Phase Adjustment Task. We also explored the impact of physiological measures (BMI and heart rate variability) on task performance, and the relationship between PAT performance and measures of depression, anxiety, stress, empathy, and self-reported interoceptive accuracy.

In a second, exploratory, study, participants were asked to complete the PAT task on two occasions, separated by one week. Such designs are typically used to assess the reliability of measurement for a particular task, however, such a design only allows measurement reliability to be established if that which is measured (in this case the ability to perceive one's heartbeat) does not itself vary. If the ability to perceive heartbeats varies, then the relationship between scores at Time 1 (T_1) and Time 2 (T_2) are a product of the stability of heartbeat perception and the reliability of the measure, with the relative contribution of each unable to be determined using such a design. Previous evidence of the stability of the ability to perceive heartbeats would be useful in disentangling the relative contribution of each factor, however, given the validity of existing tests has been questioned, existing data on the stability of heartbeat perception is itself in question. Given this caveat, it is interesting that existing studies suggest that there is a moderate 'state' component with respect to heartbeat perception, i.e., that the ability to perceive heartbeats varies in the same individual across time to a moderate-to-large degree (Murphy, Catmur, et al., 2019; see Wittkamp et al. (2018) for data and a review of existing studies). Given the existing evidence in this area, and the fact that most of it is obtained using problematic tests, it is difficult to make predictions relating to the degree of consistency in PAT scores one would expect across T_1 and T_2. However, only a moderate degree of consistency would be expected if the existing research is taken at face value.

2. Study 1

2.1. Materials and methods

2.1.1. Participants

124 healthy participants (excluding those pregnant, or with psychiatric or neurological diagnoses) were recruited via Prolific (prolific.co) and successfully completed the task (70 females, age range: 18–63, M = 26.5, *SD* = 8.02). Ethical approval was obtained from the local ethics committee and informed consent obtained from participants.

2.1.2. Test delivery

A smartphone application developed in swiftUI by BioBeats Ltd (now Huma Therapeutics) was used to administer the PAT task. The application uses a camera-driven photoplethysmogram sensor that detects heartbeats when participants place their finger over the camera (for details and validation see Cropley et al., 2017; Morelli et al., 2018). In addition to technical validation (Cropley et al., 2017; Morelli et al., 2018), this method has been demonstrated to be of equivalent accuracy when compared to lab-based methods (electrocardiogram and pulse oximeter) when recording heartbeats and does not produce a heartbeat sensation in the finger due to pressure as certain pulse oximeters may do (Murphy, Brewer, et al., 2019).

Questionnaire measures were completed via an online platform (Qualtrics) and consisted of the Depression, Anxiety and Stress Scale - 21 items (DASS-21; Lovibond & Lovibond, 1996), Body Perception Questionnaire - Awareness subscale (BPQ-A; Porges, 1993), Interoceptive Accuracy Scale (IAS; Murphy et al., 2020) and Toronto Empathy Questionnaire (TEQ; Spreng et al., 2009). The DASS-21 has three subscales assessing feelings of depression, anxiety and stress over the previous week and the total score ranges from 0 to 21. Each subscale contains 7 items scored on a 4-point scale, ranging from 0 ("Did not apply to me at all") to 3 ("Applied to me very much or most of the time"). The DASS-21 has been shown to have acceptable reliability in a non-clinical US sample (Sinclair et al., 2012) and good reliability, construct and structure validity in an international cohort (Bibi et al., 2020). The Body Perception questionnaire - Awareness subscale is part of a longer tool assessing several components of body perception. The awareness subscale contains 26 items, scored on a 5-point Likert scale ranging from 1 ("Never") to 5 ("Always"). The Awareness subscale is part of the BPQ-Short Form, containing the Awareness and Autonomic nervous system reactivity subscales, which has strong reliability and validity (Cabrera et al., 2018). The Interoceptive Accuracy Scale aims to capture self-reported accuracy of the perception of interoceptive signals. It contains 21 items rated on a scale from 5 ("Strongly agree") to 1 ("Strongly disagree"), with scores ranging from 21 to 105 (where higher scores indicate better interoceptive accuracy). The IAS has good test-retest reliability and excellent internal consistency (Murphy et al., 2020). The Toronto Empathy Questionnaire, assessing individuals' self-reported empathic abilities, contains 16-items scored on a 5-point scale ranging from 0 ("Never") to 5 ("Always"). The TEQ has been shown to have good validity, internal consistency and test-retest reliability (Spreng et al., 2009). Demographic data, including age, gender and height and weight (used to calculate BMI) were also collected.

2.1.3. PAT task: design and procedure

Participants completed the entirety of the task remotely. On each trial participants listen to a series of tones (duration = 156 ms) at the same frequency as their heart rate. They are asked to rotate a virtual dial on their smartphone to advance or retard the tones until they perceive the tones to be synchronous with their heartbeats (see Supplementary Materials [SM] for full instructions). This is equivalent to allowing the participant to continuously adjust the delay between their R-wave and the tone in order to indicate the delay perceived as synchronous, but does not require the R-wave to be measured (just the pulse at the finger using the smartphone camera (Bacciu et al., 2018; Morelli et al., 2018). Interoceptive participants (those able to perceive their heartbeats) are identified by consistency in the delay selected across trials, as those who cannot perceive their heartbeat will choose delays at random.

Importantly, heartbeats were not used to directly trigger each tone (individuals may perceive their heartbeat before it is recorded at the finger due to the delay between heart contraction and arrival of the pulse at the finger). Instead, participants were asked to sit still with their finger on the camera while their heart rate was continuously estimated from beat-to-beat intervals every 3 s throughout the whole of the task, and this estimate was used to predict when the next heartbeat would occur. If participants moved their finger, an on-screen prompt required them to reposition their finger before recording resumed.

The tones were presented at the heart rate frequency but out of phase with heartbeats, so that a single tone was played between each heartbeat and the next, and the delay between each heartbeat and the tone was adjusted by the participant. The starting delay (or phase asynchrony between heartbeats and tones) was randomly determined at the beginning of each trial. Adjustments to the dial resulted in a change of delay that was customized for each participant based on their heart rate - each turn of the dial advanced or delayed the tones by a set proportion of the time between heartbeats (so that those with slower heart rates would not have to work harder to make the tones synchronous with their perceived heartbeats). Participants were free to make as many adjustments to the dial as they wished, and pressed a button to confirm when they perceived the tones as synchronous with their heartbeat (Fig. 1A).

Simulated data were used to assess the impact of the number of trials on expected consistency of selected delays. 20 trials constituted a good compromise between task duration and the expected consistency profile, such that the risk of false negatives was minimised (see SM for details).

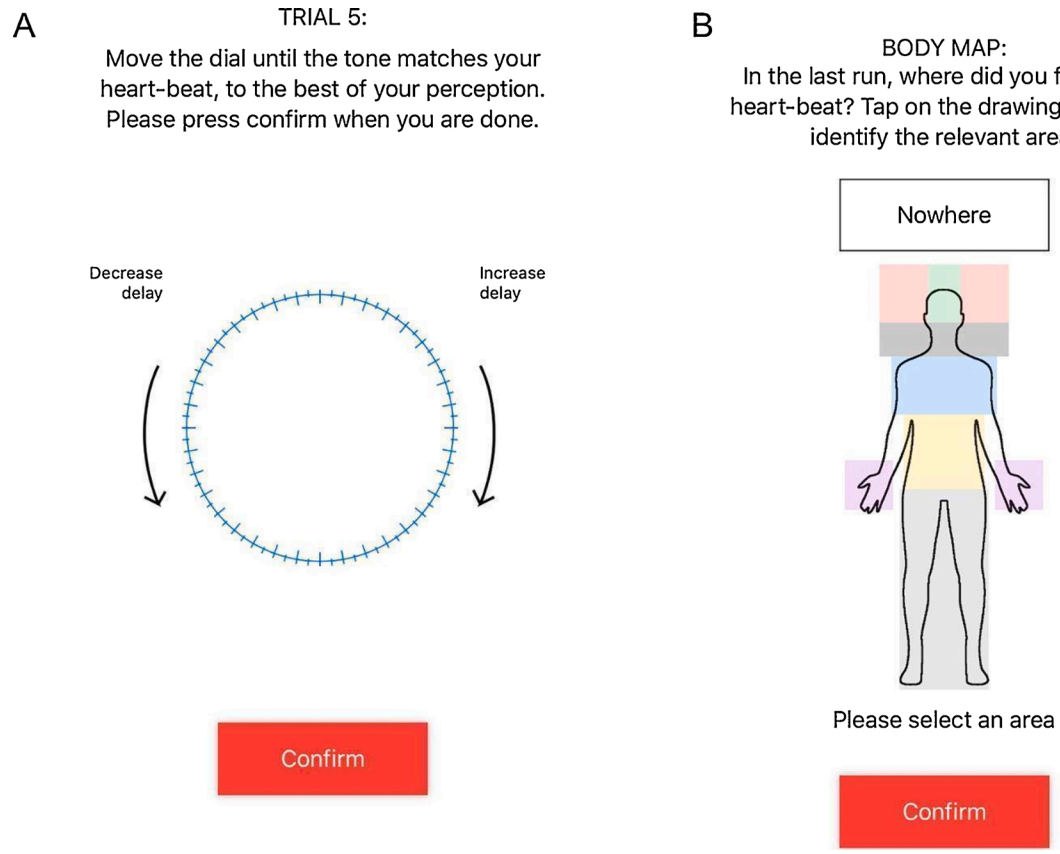

Fig. 1. Screenshots of the PAT task mobile application. A) Screenshot of one of the 20 trials and B) body map screen.

These 20 trials were preceded by 2 practice trials. After selecting their preferred delay on each trial, participants rated their confidence in having successfully completed the trial on a 10-point Visual-Analogue Scale (1="Not at all confident", 10="Extremely confident"). In addition, every 5 trials, participants indicated the location from which they felt their heartbeat on the previous trial using a body map (Fig. 1B). In total, the task took approximately 15−20 min.

2.1.4. PAT task: scoring

Note that, as in the 6AFC and MCS tasks, there is not a 'correct' response in the PAT task. Participants may perceive their heartbeat at any one of a range of delays. The measure used to discriminate between interoceptive and non-interoceptive participants is therefore the consistency of the selected phase relationship or 'delay' across trials.

Given the periodic nature of heartbeats, analysis of the consistency of selected delays should not be conducted using the standard deviation of selected delays. If, for example, a participant completes two trials at 60bpm (interbeat interval 1000 ms) and selects delays of −100 ms and 900 ms, a linear analysis of the mean and standard deviation would yield an average of 500 ms and high dispersion, with the participant classified as inconsistent (non-interoceptive). However, −100 ms and 900 ms are the same value if the period is 1000 ms (e.g. 100 ms before heartbeats), indicating high consistency. This problem is solved if the selected delay (d) is expressed as a periodic function of the period (p) so that a delay of 0 ms is equivalent to a delay that is equal to the period ($d = p$). This is implemented by expressing the delay as an arc of length $2\pi d/p$ on a unitary circumference. Mapping each participant to a standardised circumference allows individual variabilities in heart rate to be accounted for, as the space that delays are mapped onto is not influenced by differences in period length. This formalisation allows the similarity of angles to be used to define cross-trial consistency. In the previous example, delays of −100 ms and 900 ms at 60bpm would result in the same point on the circumference. In the PAT task, each movement of the dial corresponds to $\alpha = d + 2\pi d/p$, and a measure of angular similarity can therefore be used to estimate the consistency of selected delays. By expressing angles as complex numbers of modulus = 1 and argument = $\frac{2\pi d}{p}$, consistency can be computed as the modulus of the sum of the angles divided by the number of considered angles, $consistency(d,\ p) = \frac{1}{n} mod(\sum_{j=1}^{n} e^{i2\pi\frac{d_j}{p_j}})$. If angles are close to one another, the modulus of their average will be close to 1; if they are randomly positioned the modulus will approach 0. This consistency formula is equivalent to the Phase Locking Value (PLV; (Lachaux et al., 1999), a technique used to measure similarity between phases of periodic signals.

2.1.5. Tooling and data analysis

Acquisition of all data was performed using a smartphone application written in Objective-C and Swift programming languages within the XCode (version 12.5) environment and deployed using Microsoft App Centre for distribution to participants. The operating system target was iOS 10 and above, and all participant phones were iPhone 7 (and above) models by Apple (Cupertino, USA). The algorithm used for PPG detection is an implementation of the open source WABP algorithm, available on physionet (Soria, 2015). As documented by Zong et al. (2003) WABP can detect 99.31 % of the 368,364 beats annotated in an ECG, and 96.41 % of the 39,848 beats in the reference database, whereby the difference between the manually-edited and algorithm-determined ABP pulse onset was less than or equal to 20 ms.

Data were analyzed and plotted using the tydiverse, ggplot2 (Wickham, 2016), RHRV (Rodríguez-Liñares et al., 2008), AdaptGauss (Thrun et al., 2019), rstatix, ggpubr and purrr packages for R within the RStudio environment. T-tests with Welch's correction for degrees of freedom were employed to compare interoceptive and non-interoceptive participants. P-values are reported uncorrected but whether comparisons

survive Bonferroni correction for multiple comparisons is noted.

2.1.6. Comparison against random responding

To test the null hypothesis that participants' responses reflect random behaviour, 5000 participants responding at random were simulated and the resultant distribution compared to the distribution of real responses via a Wilcoxon-signed rank test.

2.1.7. Classification of participants as interoceptive or non-interoceptive

In order to classify participants as either interoceptive or non-interoceptive, a gaussian mixture model with 2 mixtures was applied to the consistency values following the assumption that the population is made of two subpopulations; interoceptive and non-interoceptive participants. The gaussian mixture model returned two distributions, one for non-interoceptive and one for interoceptive participants by means of an expectation-maximisation algorithm (Fig. 3). Z-scores were calculated for each participant for non-interoceptive and interoceptive distributions separately ($z = x - \underline{x}/s$). The obtained z-scores were used to calculate the probability of each participant belonging to the non-interoceptive and interoceptive distributions. Specifically, the probability of being non-interoceptive or interoceptive was calculated as $P_{NI}(z) = 1 - F_{NI}(|z|) + F_{NI}(-|z|)$ and $P_I(z) = 1 - F_I(|z|) + F_I(-|z|)$ respectively, with F_{NI} and F_I being the cumulative distribution functions with a mean of 0 and unitary variance. $P_{NI}(z)$ and $P_I(z)$ provide the probability of obtaining a z-score at least as extreme as the value z that was observed for each participant, $F_{NI}(z)$ and $F_I(z)$ provide the probability lower than z.

Comparison of the probability that each participant is part of the interoceptive and non-interoceptive distributions allows a Bayes Factor (BF) to be calculated as the ratio of the probability of belonging to one of the two distributions over the probability of belonging to the other distribution. Each participant can therefore be classified as more likely to be interoceptive or not interoceptive, with the BF reflecting the strength of evidence in favour of a classification. In order to allow for participants to be categorised one can apply a threshold to BFs such that participants can be classified as interoceptive, not interoceptive, or unclassified (i.e. where there is not sufficient evidence to classify a participant as either interoceptive or not). By convention, there are three thresholds that might be used for classification; BFs greater than 3 (providing moderate evidence that a participant is interoceptive or not interoceptive), greater than 10 (providing strong evidence), and greater than 30 (providing very strong evidence). We therefore present three analyses when assessing whether interoception predicts individual difference variables relating to mental health and empathy, using each of the thresholds identified above (BFs >3, >10, >30).

2.1.8. Bodily locations, average delays, and task engagement

The modal location of the four bodily location reports by each participant was used to derive a summary statistic for each participant. The mode of the most selected location was used to check whether the average delay chosen by interoceptive participants corresponded to previously reported delays between tones and perceived heartbeats. Mean delays were calculated using selected delays and heartbeat periods to obtain a complex number signalling the mean delay (calculating the modulus as period/2π). To examine the impact of participant response profiles on consistency, consistency scores were correlated with the mean time spent on trials and the number of dial movements.

2.1.9. Physiological variables

After the heartbeat occurrences from the software photoplethysmograph have been detected, the HRV time series is calculated using the RHRV package in R. Following extraction of beat-to-beat intervals (RR), core time-domain methods are deployed to calculate time analysis statistics (HRV, SDNN, RMSSD and pNN50) as measures of heart rate variability (HRV). These are calculated from a pre-testing baseline period heart rate recording of 120 s in duration.

In order to check whether physiological variables such as heart rate, HRV and BMI differed between interoceptors and non-interoceptors at the different Bayes Factor thresholds, t-tests were conducted.

2.1.10. Task confidence and self-reported measures

T-tests between interoceptors and non-interoceptors at the different BF threshold were conducted to explore differences in task confidence, self-reported interoceptive accuracy, anxiety, stress, depression and empathy.

3. Results

3.1. Comparison against random responding

Whilst the simulated data were normally distributed (Fig. 2, red line), the distribution of real consistency values was binomial (Fig. 2, black line) and only partially overlapped the simulated distribution. A Wilcoxon-signed rank test revealed the simulated and real response distributions differed significantly ($Z = 10.40$, $p < 0.0001$, $r = 0.15$), confirming that, as a group, real participants were not responding randomly. This result was confirmed using a subsequent simulation of 100,000 random responders which produced identical results.

3.2. Classification of participants as interoceptive or non-interoceptive

At the different Bayes Factor thresholds, 60 participants were classified as interoceptive and 44 participants classified as non-interoceptive (BF > 3), 51 participants classified as interoceptive and 30 participants classified as non-interoceptive (BF > 10), and 42 participants classified as interoceptive and 0 participants classified as non-interoceptive (BF > 30), respectively. Due to the lack of participants meeting the BF threshold of 30 to be considered non-interoceptive, the third analysis compared participants for whom there was very strong evidence they were interoceptive with all other participants. The ratio of participants classified as interoceptive with Bayes Factor thresholds of 10 and 30 is consistent with previous estimates that approximately $^1/_3$ of healthy participants are interoceptive on multi-delay HDT tasks (Brener & Ring, 2016).

3.3. Bodily locations, average delays, and task engagement

The means and standard deviations of consistency values in the 3 most commonly selected bodily locations did not differ (chest: N = 64, M = 0.30, *SD* = 0.16; neck: N = 18, M = 0.32, *SD* = 0.14; fingers: N = 170, M = 0.28, *SD* = 0.13). Additionally, no difference was found in consistency values between participants declaring one bodily location and

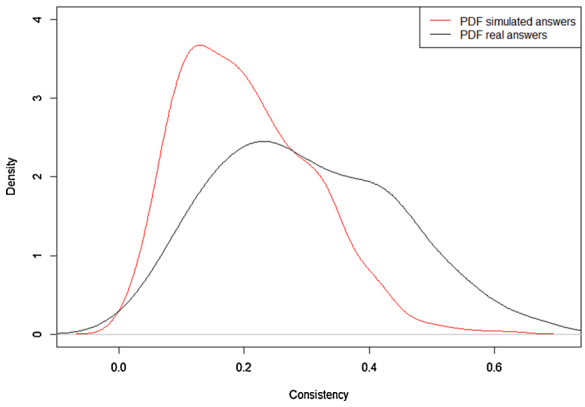

Fig. 2. Probability density function of data from simulated participants responding at random (red line) and real participants' data (black line) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

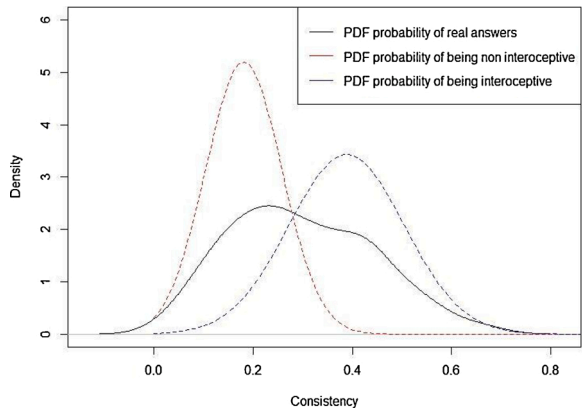

Fig. 3. Probability density function of real participants' data (black line), estimated distribution of non-interoceptive participants (red line) and interoceptive one (blue line) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

those declaring multiple locations ($t_{105.81} = 0.75$, $p = 0.45$, $d = 0.14$).

The mean delay for the most selected location by interoceptive participants (chest) was 13 ms, 23 ms, and 11 ms, for Bayes Factor thresholds of 3, 10 and 30, respectively. When added to the R-wave to finger pulse interval of 200 ms at seated rest (Obata et al., 2017), these delays fit with previous findings from the MCS that tones 200−300 ms after the R-wave are perceived as synchronous (Brener et al., 1993; Brener & Kluvitse, 1988).

Neither the mean time spent per trial, nor the number of dial movements, were significantly correlated with consistency in the overall sample (rs<.08, all ps>0.42), nor in the interoceptive subgroups (rs<+/-0.25, all ps>0.121). Furthermore, t-tests comparing interoceptive and non-interoceptive participants for these variables at each BF threshold were not significant (see Tables 1 and 2).

3.4. Physiological variables

Physiological variables such as heart rate, HRV and BMI did not differ significantly between interoceptive and non-interoceptive participants under any of the BF thresholds (see Table 1 for means and SDs for interoceptive and non-interoceptive participants under each of the three BF thresholds, and Table 2 for the results of t-tests comparing interoceptive and non-interoceptive participants). No difference in consistency was observed between males and females ($t_{111.2} = 0.906$, $p = 0.3668$, $d = 0.17$).

3.5. Task confidence and self-report interoception measures

There was no difference between task confidence reported by interoceptive and non-interoceptive participants, nor was there a difference on the Awareness subscale of the Body Perception Questionnaire or the Interoceptive Accuracy Scale (see Tables 1 and 2).

3.6. Empathy and mental health

There were clear differences, in the expected direction, between interoceptive and non-interoceptive participants in empathy, with significantly higher TEQ scores for interoceptive participants across all Bayes Factor thresholds. Note, however, that these were not seen in study 2 (which it should be noted had a far smaller sample size), and thus require further investigation in subsequent studies. The mental health data were less clear; although interoceptive participants were less stressed, anxious and depressed than non-interoceptive participants, the difference between groups of participants only approached significance for anxiety under the Bayes Factor > 10 (p = .104) and > 30 (p = .061) thresholds, and stress under the Bayes Factor > 30 threshold (p = .108;

Table 1
Mean and standard deviations of the physiological, task engagement, self-reported interoception, and mental health variables for interoceptive and non-interoceptive participants, when participants are classified as interoceptive and non-interoceptive at Bayes Factor thresholds of 3, 10, and 30 in Study 1.

Variable	BF3	BF10	BF30
HR	Interoceptive: M = 78.33, *SD* = 9.30 Non-interoceptive: M = 75.63, *SD*=7.97	Interoceptive: M = 78.48, *SD* = 9.5 Non-interoceptive: M = 76.33, *SD*=8.61	Interoceptive: M = 78.33, *SD* = 9.97 Non-interoceptive: M = 76.83, *SD*=8.35
SDNN	Interoceptive: M = 149.19, *SD* = 87.36 Non-interoceptive: M = 175.53, *SD*=108.97	Interoceptive: M = 150.17, *SD* = 89.4 Non-interoceptive: M = 174.37, *SD*=110.81	Interoceptive: M = 152.18, *SD* = 62.62 Non-interoceptive: M = 172.26, *SD*=101.38
pNN50	Interoceptive: M = 48.11, *SD* = 19.42 Non-interoceptive: M = 44.09, *SD*=17.67	Interoceptive: M = 48.14, *SD* = 19.53 Non-interoceptive: M = 41.92, *SD*=17.78	Interoceptive: M = 47.87, *SD* = 19.45 Non-interoceptive: M = 44.66, *SD*=18.89
RMSSD	Interoceptive: M = 125.18, *SD* = 66.93 Non-interoceptive: M = 114.61, *SD*=65.33	Interoceptive: M = 124.87, *SD* = 66.85 Non-interoceptive: M = 107.9, *SD*=58.52	Interoceptive: M = 125.28, *SD* = 67.72 Non-interoceptive: M = 120.21, *SD*=63.61
BMI	Interoceptive: M = 23.28, *SD* = 3.44 Non-interoceptive: M = 23.66, *SD*=5.25	Interoceptive: M = 23.46, *SD* = 3.54 Non-interoceptive: M = 23.43, *SD*=5.59	Interoceptive: M = 23.61, *SD* = 3.81 Non-interoceptive: M = 23.68, *SD*=6.26
Dial movements	Interoceptive: M = 5.08, *SD* = 3.67 Non-interoceptive: M = 4.77, *SD*=2.69	Interoceptive: M = 5.27, *SD* = 3.82 Non-interoceptive: M = 5.05, *SD*=2.88	Interoceptive: M = 5.65, *SD* = 4.1 Non-interoceptive: M = 4.81, *SD*=3.21
Time	Interoceptive: M = 19.8, *SD* = 12.55 Non-interoceptive: M = 17.69, *SD*=10.71	Interoceptive: M = 20.24, *SD* = 13.05 Non-interoceptive: M = 18.62, *SD*=12.26	Interoceptive: M = 21.15, *SD* = 13.96 Non-interoceptive: M = 17.37, *SD*=9.86
Confidence	Interoceptive: M = 5.67, *SD* = 1.86 Non-interoceptive: M = 5.6, *SD*=1.76	Interoceptive: M = 5.52, *SD* = 1.92 Non-interoceptive: M = 5.39, *SD*=2.03	Interoceptive: M = 5.4, *SD* = 2.04 Non-interoceptive: M = 5.81, *SD*=1.6
IAS	Interoceptive: M = 82.68, *SD* = 8.33 Non-interoceptive: M = 84.57, *SD*=9.87	Interoceptive: M = 82.84, *SD* = 8.05 Non-interoceptive: M = 85.87, *SD*=8.02	Interoceptive: M = 83.26, *SD* = 7.53 Non-interoceptive: M = 83.21, *SD*=10.21
BPQ	Interoceptive: M = 131.4, *SD* = 40.37 Non-interoceptive: M = 127.16, *SD*=37.72	Interoceptive: M = 131.35, *SD* = 41.79 Non-interoceptive: M = 133.43, *SD*=37.5	Interoceptive: M = 130.74, *SD* = 40.17 Non-interoceptive: M = 129.17, *SD*=38.17
DASS-21 Anxiety	Interoceptive: M = 3.17, *SD* = 3.32 Non-interoceptive: M = 3.77, *SD*=4.05	Interoceptive: M = 2.71, *SD* = 2.69 Non-interoceptive: M = 4.2, *SD*=4.48	Interoceptive: M = 2.71, *SD* = 2.72 Non-interoceptive: M = 3.84, *SD*=3.83
DASS-21 Stress	Interoceptive: M = 6.12, *SD* = 3.95 Non-interoceptive: M = 6.16, *SD*=4.4	Interoceptive: M = 5.55, *SD* = 3.78 Non-interoceptive: M = 6.5, *SD*=4.86	Interoceptive: M = 5.36, *SD* = 3.73 Non-interoceptive: M = 6.56, *SD*=4.23
DASS-21 Depression	Interoceptive: M = 4.68, *SD* = 4.75 Non-interoceptive: M = 5.96, *SD*=4.98	Interoceptive: M = 4.37, *SD* = 4.28 Non-interoceptive: M = 6.13, *SD*=5.26	Interoceptive: M = 4.38, *SD* = 4.62 Non-interoceptive: M = 5.68, *SD*=4.71
TEQ	Interoceptive: M = 42.1, *SD* = 4.35 Non-interoceptive: M = 39.14, *SD*=7.06	Interoceptive: M = 42.25, *SD* = 4.47 Non-interoceptive: M = 38.1, *SD*=7.08	Interoceptive: M = 42.38, *SD* = 4.31 Non-interoceptive: M = 40.51, *SD*=6.53

Heart rate (HR); Heart Rate Variability components: Standard Deviation of NN intervals (SDNN), proportion of NN50 (the number of times successive heartbeat intervals exceed 50 ms) divided by the total number of NN (R-R) intervals (pNN50) and Root Mean Square of Standard Deviations (RMSSD); Questionnaires: Interoceptive Accuracy Scale (IAS); Body Perception Questionnaire (BPQ) - awareness subscale; Depression, Anxiety, Stress Scale - 21 items (DASS-21); Toronto Empathy Questionnaire (TEQ); Dial movements: mean dial movements throughout the task; Time: mean time spent on trials. M: mean; SD: standard deviation.

Table 2
T-tests of comparison between interoceptive and non-interoceptive participants on physiological, task engagement, self-report interoception and mental health variables, when participants are classified as interoceptive and non-interoceptive at Bayes Factor thresholds of 3, 10, and 30 in Study 1.

Variable	BF3	BF10	BF30
HR	$t_{99.467} = 1.592$, p = 0.115, $d = 0.31$	$t_{65.82} = 1.04$, p = 0.30, $d = 0.23$	$t_{71.107} = 1.116$, p = 0.27, $d = 0.22$
SDNN	$t_{80.114}$=-1.322, p = 0.19, $d = 0.27$	$t_{51.11}$=-1.017, p = 0.314, $d = 0.25$	$t_{89.67}$=-1.106, p = 0.272, $d = 0.2$
pNN50	$t_{97.29} = 1.1$, p = 0.27, $d = 0.22$	$t_{65.582} = 1.466$, p = 0.147, $d = 0.33$	$t_{80.651} = 0.878$, p = 0.38, $d = 0.17$
RMSSD	$t_{94.055} = 0.807$, p = 0.422, $d = 0.16$	$t_{67.532} = 1.195$, p = 0.236, $d = 0.27$	$t_{78.334} = 0.403$, p = 0.69, $d = 0.08$
BMI	$t_{69.309}$=-0.426, p = 0.67, $d = 0.09$	$t_{42.937} = 0.029$, p = 0.98, $d = 0.007$	$t_{118.36}$=-0.074, p = 0.94, $d = 0.01$
Dial movements	$t_{102} = 0.487$, p = 0.627, $d = 0.09$	$t_{74.023} = 0.303$, p = 0.763, $d = 0.07$	$t_{67.431} = 1.161$, p = 0.25, $d = 0.24$
Time	$t_{99.617} = 0.925$, p = 0.358, $d = 0.18$	$t_{64.072} = 0.55$, p = 0.58, $d = 0.12$	$t_{62.554} = 1.566$, p = 0.122, $d = 0.33$
Confidence	$t_{95.537} = 0.20$, p = 0.842, $d = 0.04$	$t_{58.19} = 0.287$, p = 0.775, $d = 0.067$	$t_{67.501}$=-1.136, p = 0.26, $d = 0.23$
IAS	$t_{83.161}$=-1.027, p = 0.31, $d = 0.21$	$t_{61.154}$=-1.637, p = 0.107, $d = 0.98$	$t_{106.68} = 0.034$, p = 0.97, $d = 0.006$
BPQ	$t_{96.146} = 0.554$, p = 0.58, $d = 0.11$	$t_{66.312}$=-0.231, p = 0.82, $d = 0.05$	$t_{79.138} = 0.209$, p = 0.835, $d = 0.04$
DASS-21 Anxiety	$t_{81.48}$=-0.81, p = 0.42, $d = 0.17$	$t_{41.575}$=-1.66, p = 0.104, $d = 0.43$	$t_{109.54}$=-1.893, p = 0.061, $d = 0.32$
DASS-21 Stress	$t_{86.76}$=-0.051, p = 0.96, $d = 0.01$	$t_{49.654}$=-0.92, p = 0.36, $d = 0.23$	$t_{92.626}$=-1.624, p = 0.108, $d = 0.30$
DASS-21 Depression	$t_{90.156}$=-1.311, p = 0.19, $d = 0.26$	$t_{51.483}$=-1.556, p = 0.126, $d = 0.38$	$t_{84.347}$=-1.475, p = 0.144, $d = 0.28$
TEQ	**$t_{66.519} = 2.464$, p = 0.016, $d = 0.52$**	**$t_{42.846} = 2.89$, p = 0.006, $d = 0.75$**	$t_{114.29} = 1.905$, p = 0.059, $d = 0.32$

Heart rate (HR); Heart Rate Variability components: Standard Deviation of NN intervals (SDNN), proportion of NN50 (the number of times successive heartbeat intervals exceed 50 ms) divided by the total number of NN (R-R) intervals (pNN50) and Root Mean Square of Standard Deviations (RMSSD); Questionnaires: Interoceptive Accuracy Scale (IAS); Body Perception Questionnaire (BPQ) - awareness subscale; Depression, Anxiety, Stress Scale - 21 items (DASS-21); Toronto Empathy Questionnaire (TEQ); Dial movements: mean dial movements throughout the task; Time: mean time spent on trials. T(df); t values and degrees of freedom. p; uncorrected p values, note none of these comparisons were significant after multiple comparisons correction. d; Cohen's d value.

see Tables 1 and 2 for details). It should be noted, however, that these are 2-tailed p values, and it could be argued that 1-tailed p values are more appropriate. While this argument could be made for stress, the mixed findings in the literature for anxiety (see Introduction) make this questionable for anxiety.

4. Study 2: consistency of PAT scores after a one week delay

4.1. Materials and methods

4.1.1. Participants and procedure

90 Prolific users were invited to participate in the Study. Prior to completing the PAT, participants completed a screening task that was identical to the PAT, except that each participant's heartbeats were sonified, and participants were required to synchronize the two auditory signals as in the PAT task (by changing the delay, or phase asynchrony, between the heartbeat tone and the delayed tone). This task was included in order to ensure that individual differences on the PAT could not be attributed to poor understanding of the task instructions, ability to synchronise signals, or poor motivation and/or attention. The screening task was analysed in the same way as the PAT task. 73 participants completed the screening task and were invited through Prolific to take part in two measurement occasions (T_1, T_2) of the PAT task. 60 accepted the invitation. Of these 60, 59 completed the PAT task at T_1, and 43 at T_2. After accounting for missing data and skipped trials, two further participants' datasets were discarded, resulting in 41 usable datasets (30 female, 11 male, ages 19–58, mean age = 29.7). After filtering for users who passed the screening task successfully, the population was reduced to 23 usable datasets (scr + population, 17 female, 6 male, ages 19–53, mean age = 30.6) with data at T_1 and T_2. The mean delay between T_1 and T_2 was 7.1 days. Participants were given identical instructions as in Study 1. For all participants, usage of the app was checked to ensure that the task was not completed between testing occasions.

4.2. Results

4.2.1. Bayes factor classification

Table 3 shows numbers of participants across Bayes Factor classifications (interoceptive, non-interoceptive, unknown) for BF > 3, BF > 10 and BF > 30 across the 23 participants in the scr + population.

4.2.2. Consistency in PAT similarity scores from T_1 to T_2

Calculation of the Intraclass Correlation Coefficient was performed using the icc functions in the irr package in R (Gamer et al., 2019) across the two measurement points in those that successfully passed the screening task suggested moderate consistency in PAT scores of 0.508 (note that in the complete sample of those with usable data in T_1 and T_2 – i.e., including those who did NOT pass the screening task – ICC was 0.618). Paired *t*-test results show a significant difference between T_1 and T_2, suggesting a potential practice effect ($T_{(22)} = -2.44$, $p = .022$, $d = 0.51$), with a mean similarity score of 0.33 at T_1 (SD = 0.12) and 0.42 at T_2 (SD = 0.18).

4.2.3. Classification SAcross measurement points

This moderate consistency in PAT similarity scores was reflected in consistency in classification, for example, at BF > 3, across T_1 and T_2 five participants were consistently interoceptive, four were consistently non-interoceptive, five switched from interoceptive to non-interoceptive, five switched from non-interoceptive to interoceptive, one switched from non-interoceptive to unknown, three from unknown to interoceptive, for a total of nine consistent participants, and fourteen inconsistent participants. Tables S1 and S2 present physiological and task-engagement variables separately for participants who remain consistent in their classification as interoceptive or not across T_1 and T_2, vs those who are inconsistent (Table S1), and the same variables separately for those who are consistently interoceptive (n = 4), consistently non-interoceptive (n = 4), who change from interoceptive to non-interoceptive (n = 4), from non-interoceptive to interoceptive (n = 9), from non-interoceptive to unknown (n = 1), and from unknown to non-interoceptive (n = 1) (Table S2) at BF > 3. Small sample sizes render these data unsuitable for statistical analysis, but data is reported in the Supplementary Materials for information.

Table 3
Bayes factor classification of participants successfully passing the screening test for Study 2.

	T1	T2
Bayes Factor >3		
Interoceptive participants	8 (34.8 %)	13 (56.5 %)
Non-interoceptive participants	14 (60.9 %)	9 (39.1 %)
Unknown	1 (4.3 %)	1 (4.3 %)
Bayes Factor >10		
Interoceptive participants	7 (30.4 %)	13 (56.5 %)
Non-interoceptive participants	13 (56.5 %)	8 (34.8 %)
Unknown	3 (13.0 %)	2 (8.7 %)
Bayes Factor >30		
Interoceptive participants	6 (26.1 %)	13 (56.5 %)
Non-interoceptive participants	13 (56.5 %)	2 (8.7 %)
Unknown	4 (17.4 %)	8 (34.8 %)

5. General discussion

Study 1 outlined the development of a novel measure of interoceptive ability - the PAT task. Participants' consistency scores differed significantly from random responses, with between 34 % and 48 % of participants classified as interoceptive depending on the threshold adopted. Consistency scores did not differ according to the bodily location from which the heartbeat was perceived, nor the number of locations from which it was perceived. Neither physiological characteristics of the participants (including HRV and heart rate), nor their response profiles, related to interoceptive ability. Study 2 explored the consistency in PAT scores across a one-week delay. The Intraclass Correlation Coefficient indicated moderate consistency in scores, and this was reflected in moderate consistency in classification of participants as interoceptive or not.

No relationship between performance on the PAT task and confidence was observed, but it is worth noting that the confidence question was poorly specified; future versions of the PAT task will employ separate questions concerning confidence in heartbeat perception and confidence in task performance. Notably, previous studies of the relationship between self-reported and objective measures of interoception produced mixed results (Murphy, Catmur, et al., 2019), and a lack of a relationship is consistent with proposals that interoception is a multidimensional construct, with dissociations between subjective and objective aspects (Garfinkel et al., 2015; Murphy, Catmur, et al., 2019). This may also explain why Interoceptive Accuracy Scores and Body Perception Questionnaire did not dissociate between interoceptive and non-interoceptive participants, although a recent study has highlighted problems with how participants interpret questions on the Body Perception Questionnaire (Gabriele et al., 2020), which may also explain a lack of association with PAT scores on this measure.

Currently, the PAT task classifies participants as interoceptive, not interoceptive, or unclassifiable. Individual differences in consistency scores are not treated as meaningful, except for the evidence they provide for participant classification. This is because if there is no statistical evidence that a participant can perceive their heartbeat (i.e., if the evidence that they are interoceptive is not sufficiently greater than that they are not-interoceptive), then whether that participant's consistency score is greater than another participant's consistency score (for whom there is also insufficient evidence that they are either interoceptive or not-interoceptive) is not useful information. Conversely, if two participants achieve a consistency score indicating that they are interoceptive, there is currently no way of knowing whether the difference in their scores reflects anything other than motivation, attention to the task, or timing ability (note that these factors cannot produce interoceptive performance in the absence of heartbeat perception). The validity of classifying participants as interoceptive or not - instead of using continuous consistency scores - is supported by the proportion of participants who cannot be classified under each threshold. These participants vary in their consistency scores but not to a degree that enables them to be classified as either interoceptive or not with a sufficient degree of certainty. The classification of participants as either interoceptive or not is motivated by the fact that, when tested at rest, participants can either feel their heartbeat or they cannot. Unless the heartbeat is at the very limit of being perceptible, it is unclear why some heartbeats should be perceived and some not (the basis of individual differences in HCT scores). Instead, we suggest that individual differences in interoception should be considered as the percentage of time, or variety of circumstances, in which one can perceive interoceptive signals in everyday life. Such a suggestion is consistent with evidence of state effects on interoception (Murphy, Catmur, et al., 2019; Wittkamp et al., 2018), and the variance in PAT classification observed in Study 2. Future research should therefore seek to examine performance on the PAT across situations and time, coupled with tests of interoception in domains other than cardiac (e.g., respiratory, physiological arousal). It should be clear then, that our use of the terms interoceptive and non-interoceptive relate to a participant's performance at a particular time, and for the cardiac signal alone. It should also be noted that interoceptive participants could perceive their heartbeats when completing the PAT task which provides an external signal of their heartbeats, whether they can perceive their heartbeats without such a signal is unknown. Indeed, the consistency of an individual's interoceptive ability across time, situations, and domains is likely to be more useful than exactly how many heartbeats they count in a given period when testing theorised relationships between interoception, health and cognition. It is of interest therefore, that a significant correlation was observed between the change in PAT similarity scores from T_1 to T_2 in Study 2 and change in HRV(RMSSD) in those participants who were inconsistently classified as interoceptive or not ($r(12) = .573$ $p = 0.034$). The change in RMSSD suggests a state change between T_1 and T_2 occasions, as RMSSD is the primary time-domain measure used to estimate the vagally-mediated changes reflected in HRV, and more accurately represents parasympathetic nervous system activity than other HRV variables. This potential state change and its relationship to cardiac interoception warrants further investigation with much larger sample sizes. It is especially important to determine how much of the inconsistency in PAT scores is a product of the test itself, or reflective of true variation in the ability to perceive one's heartbeat across time (of course, ideally this would require an additional valid test of cardioception). Of interest also is testing the stability of PAT scores when tested under laboratory conditions using a smartphone as instantiated here, and using an ECG to trigger tones. Furthermore, it would be interesting to compare stability when administered remotely via smartphone to participants who are more naive than the somewhat 'professional' participants available for recruitment via Prolific.co. Once the test-retest reliability of the PAT task is established, this will inform future studies of links between cardioception and individual differences variables so that they are appropriately powered.

With respect to task validation, the absence of an effect of physiological factors is encouraging. This is particularly true for heart rate variability which, if especially extreme, could have been problematic as it would have made heartbeats difficult to predict. Previous studies of interoception have controlled for physiological factors statistically when they correlate with interoceptive ability (Brener & Ring, 2016; Murphy et al., 2018) - but the absence of physiological effects on the PAT task makes this unnecessary. The smartphone data collection has been extensively validated (Cropley et al., 2017; Morelli et al., 2018; Murphy, Brewer, et al., 2019), however the PAT task can be conducted using any objective method of heartbeat detection (for example an electrocardiogram) in situations in which smartphones cannot be used.

The absence of robust correlations with mental health variables, or with self-reported interoception, could be considered evidence that the PAT task has low validity. However, as noted above, current evidence suggests that self-report measures of interoception do not typically predict interoceptive accuracy (see Murphy et al., 2020 for discussion), and studies of the link between interoception and mental health have produced variable results (see Introduction for review of studies concerning anxiety, and also (Desmedt et al., 2020), and these studies have typically used interoception tests of questionable validity. For example, when it comes to the link between interoception and depression, although depression is often associated with poor cardiac interoception (e.g., Furman et al., 2013; Lackner & Fresco, 2016; Pollatos et al., 2009; Terhaar et al., 2012); (for a review see Eggart et al., 2019), this relationship may be non-linear (Dunn et al., 2007), and there have been some failures to replicate this association (e.g., Dunn, Stefanovitch, et al., 2010; Mussgay et al., 1999; Shah et al., 2016; Van der Does et al., 1997). It must also be acknowledged that the lack of a relationship between PAT scores and mental health variables may reflect the fact that a community sample was tested, and none of the participants had scores in the clinical range on anxiety and depression measures - a limitation of this study.

With respect to task validity, it is unlikely that the PAT incorrectly

classifies non-interoceptive participants as interoceptive, but classifying interoceptive participants as non-interoceptive is possible on any task if a participant has low motivation or attention. Furthermore, individuals with poor timing ability may also perform poorly on the PAT task even if they can perceive their heartbeat (although timing ability would have to be very poor indeed to be problematic). The development of the screening task for Study 2 guards against these possibilities, however, by requiring participants to successfully complete a task that is structurally identical to the PAT task, but which does not require any ability to perceive heartbeats. Any problems with attention or motivation, or poor timing ability, would mean that participants would not be able to pass the screening task. Given that the PAT task is designed to enable remote testing, any experimenters with concerns even with participants who successfully pass the screening task may wish to observe task completion via video chat software.

Problematically, given the questionable validity of the HCT and HDT tasks, they cannot be used to validate the PAT task. It is therefore encouraging that the delays at which participants judge heartbeats as synchronous with the tones on the PAT task matched those previously observed with the MCS and 6AFC tasks (Brener et al., 1993; Brener & Kluvitse, 1988; Yates et al., 1985). The task is also far more scalable and accessible for clinical populations than currently available measures. We therefore suggest that the PAT task represents a significant advance in methodology available for research on interoception.

Funding statement

This work was supported by Huma Therapeutics.

Declaration of Competing Interest

DP, DM, MC and AC are employees of Huma Therapeutics. GB has a financial interest in Huma Therapeutics. CR, CTK, CC and JM have no conflicts of interest to declare. Huma Therapeutics is a digital health company, specifically interested in identifying optimal remote patient monitoring solutions as well as novel biomarkers. Huma covered participants' expenses as well as part of the research team's salaries (SP, DM, DP, MC).

Acknowledgements

We are grateful to Leonardo Bartoloni for helpful discussion about analyses.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.biopsycho.2021.108171.

References

Ardizzi, M., Ambrosecchia, M., Buratta, L., Ferri, F., Peciccia, M., Donnari, S., … Gallese, V. (2016). Interoception and positive symptoms in schizophrenia. *Frontiers in Human Neuroscience, 10*, 379. https://doi.org/10.3389/fnhum.2016.00379

Asmundson, G. J., Sandler, L. S., Wilson, K. G., & Norton, G. R. (1993). Panic attacks and interoceptive acuity for cardiac sensations. *Behaviour Research and Therapy, 31*(2), 193–197. https://doi.org/10.1016/0005-7967(93)90071-2

Bacciu, D., Colombo, M., Morelli, D., & Plans, D. (2018). Randomized neural networks for preference learning with physiological data. *Neurocomputing*. https://www.sciencedirect.com/science/article/pii/S092523121830211X.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16*(7), 419–429. https://doi.org/10.1038/nrn3950

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0011

Bibi, A., Lin, M., Zhang, X. C., & Margraf, J. (2020). Psychometric properties and measurement invariance of Depression, Anxiety and Stress Scales (DASS-21) across cultures. *International Journal of Psychology: Journal International de Psychologie, 55*(6), 916–925. https://doi.org/10.1002/ijop.12671

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews, 47*, 520–532. https://doi.org/10.1016/j.neubiorev.2014.09.021

Bogaerts, K., Notebaert, K., Van Diest, I., Devriese, S., De Peuter, S., & Van den Bergh, O. (2005). Accuracy of respiratory symptom perception in different affective contexts. *Journal of Psychosomatic Research, 58*(6), 537–543. https://doi.org/10.1016/j.jpsychores.2004.12.005

Brener, J., & Kluvitse, C. (1988). Heartbeat detection: judgments of the simultaneity of external stimuli and heartbeats. *Psychophysiology, 25*(5), 554–561. https://doi.org/10.1111/j.1469-8986.1988.tb01891.x

Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: The measurement of heartbeat detection. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0015

Brener, J., Liu, X., & Ring, C. (1993). A method of constant stimuli for examining heartbeat detection: Comparison with the Brener-Kluvitse and Whitehead methods. *Psychophysiology, 30*(6), 657–665. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.1993.tb02091.x?casa_token=y3xUYHXn8BwAAAAA:yjMbngP0Sb8d3kwxuHyFXmkBVjjjh2X7iSLadpTWuLOFbWn3v2gqt1EIkjHGwvpvz52z1LeB4bJ3lKY.

Brewer, R., Murphy, J., & Bird, G. (2021). Atypical interoception as a common risk factor for psychopathology: A review. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2021.07.036

Cabrera, A., Kolacz, J., Pailhez, G., Bulbena-Cabre, A., Bulbena, A., & Porges, S. W. (2018). Assessing body awareness and autonomic reactivity: Factor structure and psychometric properties of the Body Perception Questionnaire-Short Form (BPQ-SF). *International Journal of Methods in Psychiatric Research, 27*(2), e1596. https://doi.org/10.1002/mpr.1596

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Clemens, W. J. (1984). Temporal arrangement of signals in heartbeat discrimination procedures. *Psychophysiology, 21*(2), 187–190. https://doi.org/10.1111/j.1469-8986.1984.tb00202.x

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience, 3*(8), 655–666. https://doi.org/10.1038/nrn894

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7*(2), 189–195. https://doi.org/10.1038/nn1176

Cropley, M., Plans, D., Morelli, D., Sütterlin, S., Inceoglu, I., Thomas, G., & Chu, C. (2017). The association between work-related rumination and heart rate variability: A field study. *Frontiers in Human Neuroscience, 11*, 27. https://doi.org/10.3389/fnhum.2017.00027

Dale, A., & Anderson, D. (1978). Information variables in voluntary control and classical conditioning of heart rate: Field dependence and heart-rate perception. *Perceptual and Motor Skills, 47*(1), 79–85. https://doi.org/10.2466/pms.1978.47.1.79

Desmedt, O., Van Den Houte, M., Walentynowicz, M., Dekeyser, S., & Corneille, O. (2020). *A systematic review and meta-analysis on the association between heartbeat counting task performance and mental disorders and their risk factors among adults.* https://doi.org/10.31219/osf.io/h3by9

Dunn, B. D., Dalgleish, T., Ogilvie, A. D., & Lawrence, A. D. (2007). Heartbeat perception in depression. *Behaviour Research and Therapy, 45*(8), 1921–1930. https://doi.org/10.1016/j.brat.2006.09.008

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D., & Dalgleish, T. (2010). Listening to your heart. How interoception shapes emotion experience and intuitive decision making. *Psychological Science, 21*(12), 1835–1844. https://doi.org/10.1177/0956797610389191

Dunn, B. D., Stefanovitch, I., Evans, D., Oliver, C., Hawkins, A., & Dalgleish, T. (2010). Can you feel the beat? Interoceptive awareness is an interactive function of anxiety- and depression-specific symptom dimensions. *Behaviour Research and Therapy, 48*(11), 1133–1138. https://doi.org/10.1016/j.brat.2010.07.006

Eggart, M., Lange, A., Binser, M. J., Queri, S., & Müller-Oerlinghausen, B. (2019). Major depressive disorder is associated with impaired interoceptive accuracy: A systematic review. *Brain Sciences, 9*(6). https://doi.org/10.3390/brainsci9060131

Ehlers, A., Breuer, P., Dohn, D., & Fiegenbaum, W. (1995). Heartbeat perception and panic disorder: Possible explanations for discrepant findings. *Behaviour Research and Therapy, 33*(1), 69–76. https://doi.org/10.1016/0005-7967(94)e0002-z

Eley, T. C., Gregory, A. M., Clark, D. M., & Ehlers, A. (2007). Feeling anxious: A twin study of panic/somatic ratings, anxiety sensitivity and heartbeat perception in children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 48*(12), 1184–1191. https://doi.org/10.1111/j.1469-7610.2007.01838.x

Ewing, D. L., Manassei, M., Gould van Praag, C., Philippides, A. O., Critchley, H. D., & Garfinkel, S. N. (2017). Sleep and the heart: Interoceptive differences linked to poor experiential sleep quality in anxiety and depression. *Biological Psychology, 127*, 163–172. https://doi.org/10.1016/j.biopsycho.2017.05.011

Furman, D. J., Waugh, C. E., Bhattacharjee, K., Thompson, R. J., & Gotlib, I. H. (2013). Interoceptive awareness, positive affect, and decision making in major depressive disorder. *Journal of Affective Disorders, 151*(2), 780–785. https://doi.org/10.1016/j.jad.2013.06.044

Füstös, J., Gramann, K., Herbert, B. M., & Pollatos, O. (2013). On the embodiment of emotion regulation: Interoceptive awareness facilitates reappraisal. *Social Cognitive and Affective Neuroscience, 8*(8), 911–917. https://doi.org/10.1093/scan/nss089

Gabriele, E., Spooner, R., Brewer, R., & Murphy, J. (2020). *Dissociations between interoceptive accuracy and attention: Evidence from the interoceptive attention scale*. https://doi.org/10.31234/osf.io/vjgh6

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *Coefficients of interrater reliability and agreement for quantitative, ordinal and nominal data: ICC, Finn-coefficient, Robinson's A, Kendall's W, and Cohen's kappa*.

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology, 104*, 65–74. https://doi.org/10.1016/j.biopsycho.2014.11.004

Garfinkel, S. N., Manassei, M. F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H. D., & Engels, M. (2016). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0014

Garfinkel, S. N., Tiley, C., O'Keeffe, S., Harrison, N. A., Seth, A. K., & Critchley, H. D. (2016). Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychology, 114*, 117–126. https://doi.org/10.1016/j.biopsycho.2015.12.003

Hechler, T. (2021). Altered interoception and its role for the co-occurrence of chronic primary pain and mental health problems in children. *Pain, 162*(3), 665–671. https://doi.org/10.1097/j.pain.0000000000002099

Herbert, B. M., Herbert, C., Pollatos, O., Weimer, K., Enck, P., Sauer, H., & Zipfel, S. (2012). Effects of short-term food deprivation on interoceptive awareness, feelings and autonomic cardiac activity. *Biological Psychology, 89*(1), 71–79. https://doi.org/10.1016/j.biopsycho.2011.09.004

Hickman, L., Seyedsalehi, A., Cook, J. L., Bird, G., & Murphy, J. (2020). The relationship between heartbeat counting and heartbeat discrimination: A meta-analysis. *Biological Psychology, 156*, Article 107949. https://doi.org/10.1016/j.biopsycho.2020.107949

Katkin, E. S., Reed, S. D., & Deroo, C. (1983). A methodological analysis of 3 techniques for the assessment of individual-differences in heartbeat detection. *Psychophysiology, 20*, 452.

Kever, A., Pollatos, O., Vermeulen, N., & Grynberg, D. (2015). Interoceptive sensitivity facilitates both antecedent- and response-focused emotion regulation strategies. *Personality and Individual Differences, 87*, 20–23. https://doi.org/10.1016/j.paid.2015.07.014

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., … Interoception Summit 2016 participants. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry Cognitive Neuroscience and Neuroimaging, 3*(6), 501–513. https://doi.org/10.1016/j.bpsc.2017.12.004

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8*(4), 194–208. https://doi.org/10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c

Lackner, R. J., & Fresco, D. M. (2016). Interaction effect of brooding rumination and interoceptive awareness on depression and anxiety symptoms. *Behaviour Research and Therapy, 85*, 43–52. https://doi.org/10.1016/j.brat.2016.08.007

Lovibond, S. H., & Lovibond, P. F. (1996). Manual for the depression anxiety stress scales. *Psychology Foundation of Australia*. https://play.google.com/store/books/details?id=mXoQHAAACAAJ.

Morelli, D., Bartoloni, L., Colombo, M., Plans, D., & Clifton, D. A. (2018). Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. *Healthcare Technology Letters, 5*(2), 59–64. https://doi.org/10.1049/htl.2017.0039

Murphy, J., Brewer, R., Catmur, C., & Bird, G. (2017). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental Cognitive Neuroscience, 23*, 45–56. https://doi.org/10.1016/j.dcn.2016.12.006

Murphy, J., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018). Is alexithymia characterised by impaired interoception? Further evidence, the importance of control variables, and the problems with the Heartbeat counting Task. *Biological Psychology, 136*, 189–197. https://doi.org/10.1016/j.biopsycho.2018.05.010

Murphy, J., Brewer, R., Plans, D., Khalsa, S. S., Catmur, C., & Bird, G. (2020). Testing the independence of self-reported interoceptive accuracy and attention. *The Quarterly Journal of Experimental Psychology, 73*(1), 115–133. https://doi.org/10.1177/1747021819879826

Murphy, J., Brewer, R., Coll, M.-P., Plans, D., Hall, M., Shiu, S. S., Catmur, C., & Bird, G. (2019). I feel it in my finger: Measurement device affects cardiac interoceptive accuracy. *Biological Psychology, 148*, Article 107765. https://doi.org/10.1016/j.biopsycho.2019.107765

Murphy, J., Catmur, C., & Bird, G. (2019). Classifying individual differences in interoception: Implications for the measurement of interoceptive awareness. *Psychonomic Bulletin & Review, 26*(5), 1467–1471. https://doi.org/10.3758/s13423-019-01632-7

Mussgay, L., Klinkenberg, N., & Rüddel, H. (1999). Heart beat perception in patients with depressive, somatoform, and personality disorders. *Journal of Psychophysiology, 13*(1), 27. https://psycnet.apa.org/journals/jop/13/1/27/.

Naqvi, N. H., & Bechara, A. (2010). The insula and drug addiction: An interoceptive view of pleasure, urges, and decision-making. *Brain Structure & Function, 214*(5–6), 435–450. https://doi.org/10.1007/s00429-010-0268-7

Obata, Y., Ong, Q. J., Magruder, J. T., Grichkevitch, H., Berkowitz, D. E., Nyhan, D., … Barodka, V. (2017). Noninvasive assessment of the effect of position and exercise on pulse arrival to peripheral vascular beds in healthy volunteers. *Frontiers in Physiology, 8*, 47. https://doi.org/10.3389/fphys.2017.00047

Palser, E. R., Fotopoulou, A., Pellicano, E., & Kilner, J. M. (2018). The link between interoceptive processing and anxiety in children diagnosed with autism spectrum disorder: Extending adult findings into a developmental sample. *Biological Psychology, 136*, 13–21. https://doi.org/10.1016/j.biopsycho.2018.05.003

Paulus, M. P. (2007). Neural basis of reward and craving-a homeostatic point of view. *Dialogues in Clinical Neuroscience, 9*(4), 379. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3202500/.

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure & Function, 214*(5-6), 451–463. https://doi.org/10.1007/s00429-010-0258-9

Paulus, M. P., & Stewart, J. L. (2014). Interoception and drug addiction. *Neuropharmacology, 76*, 342–350. https://doi.org/10.1016/j.neuropharm.2013.07.002

Paulus, M. P., Tapert, S. F., & Schulteis, G. (2009). The role of interoception and alliesthesia in addiction. *Pharmacology, Biochemistry, and Behavior, 94*(1), 1–7. https://doi.org/10.1016/j.pbb.2009.08.005

Pollatos, O., Füstös, J., & Critchley, H. D. (2012). On the generalised embodiment of pain: How interoceptive sensitivity modulates cutaneous pain perception. *Pain, 153*(8), 1680–1686. https://doi.org/10.1016/j.pain.2012.04.030

Pollatos, O., Traut-Mattausch, E., & Schandry, R. (2009). Differential effects of anxiety and depression on interoceptive accuracy. *Depression and Anxiety, 26*(2), 167–173. https://doi.org/10.1002/da.20504

Porges, S. (1993). *Body perception questionnaire*. Laboratory of Developmental Assessment, University of Maryland. https://www.stephenporges.com/s/Czech-SF-updated-7-17.pdf.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience and Biobehavioral Reviews, 47*, 410–430. https://doi.org/10.1016/j.neubiorev.2014.09.012

Ring, C., & Brener, J. (1992). The temporal locations of heartbeat sensations. *Psychophysiology, 29*(5), 535–545. https://doi.org/10.1111/j.1469-8986.1992.tb02027.x

Rodríguez-Liñares, L., Vila, X., Méndez, A. J., Lado, M., & Olivieri, D. (2008). RHRV: An R-based software package for heart rate variability analysis of ECG recordings. *3rd Iberian Conference in Systems and Information Technologies (CISTI 2008)*, 565–574. https://www.academia.edu/download/30962624/cisti2008.pdf.

Schaefer, M., Egloff, B., & Witthöft, M. (2012). Is interoceptive awareness really altered in somatoform disorders? Testing competing theories with two paradigms of heartbeat perception. *Journal of Abnormal Psychology, 121*(3), 719–724. https://doi.org/10.1037/a0028509

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology, 18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Schmidt, A. F., Eulenbruch, T., Langer, C., & Banger, M. (2013). Interoceptive awareness, tension reduction expectancies and self-reported drinking behavior. *Alcohol and Alcoholism, 48*(4), 472–477. https://doi.org/10.1093/alcalc/agt024

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*, 220–223. https://doi.org/10.1016/j.cortex.2017.02.010

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 81*, 215–220. https://doi.org/10.1016/j.cortex.2016.03.021

Sinclair, S. J., Siefert, C. J., Slavin-Mulford, J. M., Stein, M. B., Renna, M., & Blais, M. A. (2012). Psychometric evaluation and normative data for the depression, anxiety, and stress scales-21 (DASS-21) in a nonclinical sample of U.S. adults. *Evaluation & the Health Professions, 35*(3), 259–279. https://doi.org/10.1177/0163278711424282

Soria, M. L. (2015). *ABP/PPG peak detection — ECGkit 0.1.1 documentation*. Physionet. https://archive.physionet.org/physiotools/ecg-kit/help/sphinx/build/html/ABP_PPG_peak_detection.html.

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Stevens, S., Gerlach, A. L., Cludius, B., Silkens, A., Craske, M. G., & Hermann, C. (2011). Heartbeat perception in social anxiety before and during speech anticipation. *Behaviour Research and Therapy, 49*(2), 138–143. https://doi.org/10.1016/j.brat.2010.11.009

Terhaar, J., Viola, F. C., Bär, K.-J., & Debener, S. (2012). Heartbeat evoked potentials mirror altered body perception in depressed patients. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 123*(10), 1950–1957. https://doi.org/10.1016/j.clinph.2012.02.086

Thrun, M. C., Hansen-Goos, O., Griese, R., Lippmann, C., & Ultsch, A. (2019). *R Package Gaussian Mixture Models (GMM)*. https://www.researchgate.net/publication/337151568_R_Package_Gaussian_Mixture_Models_GMM.

Van der Does, A. J., Van Dyck, R., & Spinhoven, P. (1997). Accurate heartbeat perception in panic disorder: Fact and artefact. *Journal of Affective Disorders, 43*(2), 121–130. https://doi.org/10.1016/s0165-0327(96)01414-0

Verdejo-Garcia, A., Clark, L., & Dunn, B. D. (2012). The role of interoception in addiction: A critical review. *Neuroscience and Biobehavioral Reviews, 36*(8), 1857–1869. https://doi.org/10.1016/j.neubiorev.2012.05.007

Werner, N. S., Jung, K., Duschek, S., & Schandry, R. (2009). Enhanced cardiac perception is associated with benefits in decision-making. *Psychophysiology, 46*(6), 1123–1129. https://doi.org/10.1111/j.1469-8986.2009.00855.x

Whitehead, W. E., Drescher, V. M., Heiman, P., & Blackwell, B. (1977). Relation of heart rate control to heartbeat perception. *Biofeedback and Self-Regulation, 2*(4), 371–392. https://doi.org/10.1007/BF00998623

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://play.google.com/store/books/details?id=XgFkDAAAQBAJ.

Windmann, S., Schonecke, O. W., Fröhlig, G., & Maldener, G. (1999). Dissociating beliefs about heart rates and actual heart rates in patients with cardiac pacemakers. *Psychophysiology, 36*(3), 339–342. https://doi.org/10.1017/s0048577299980381

Wittkamp, M. F., Bertsch, K., Vögele, C., & Schulz, A. (2018). A latent state-trait analysis of interoceptive accuracy. *Psychophysiology, 55*(6), Article e13055. https://doi.org/10.1111/psyp.13055

Yates, A. J., Jones, K. E., Marie, G. V., & Hogben, J. H. (1985). Detection of the heartbeat and events in the cardiac cycle. *Psychophysiology, 22*(5), 561–567. https://doi.org/10.1111/j.1469-8986.1985.tb01651.x

Zong, W., Heldt, T., Moody, G. B., & Mark, R. G. (2003). An open-source algorithm to detect onset of arterial blood pressure pulses. *Computers in Cardiology, 2003*, 259–262. https://doi.org/10.1109/CIC.2003.1291140